



# VAN DE KAT EN DE BEL

*Tellen en vertellen  
met de eindtoets basisonderwijs*

Karen Heij



# **VAN DE KAT EN DE BEL**

*Tellen en vertellen*

*met de eindtoets basisonderwijs*

Karen Heij

ISBN:978-90-831133-3-3  
Printing: Bijzonderdruk  
Omslag afbeelding: Carlijn Kingma

Copyright © 2021 K. Heij  
All rights reserved. No part of this thesis may be reproduced in any form without written permission of the author or copyright-owning journals for previously published chapters.

# Van de kat en de bel

*Tellen en vertellen met de eindtoets basisonderwijs*

The cat and the bell

*Telling and counting with the primary school end test*

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University  
op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te  
verdedigen ten overstaan van een door het college voor promoties aangewezen  
commissie in de Aula van  
de Universiteit op woensdag 21 april 2021 om 14.00 uur

door

Karen Heij,  
geboren op 25 oktober 1963  
te Leiden

**Promotor:**

Prof. dr. P.H.A. Frissen, Tilburg University

**Copromotor:**

Dr. F.J. de Vijlder, HAN University of Applied Sciences

**Promotiecommissie:**

Prof. dr. G.C.G. Dehue, Emeritus Rijksuniversiteit Groningen

Prof. dr. E.H. Hooge, Tilburg University

Prof. dr. R. Standaert, Emeritus Universiteit Gent

Prof. dr. M.J.M. Vermeulen, Tilburg University

Prof. mr. dr. S. Zouridis, Tilburg University

*Voor Josephine*





# INHOUD

<b>Woord vooraf</b> .....	11
<b>Hoofdstuk 1 'De Cito'</b> .....	15
1.1 De werking van de eindtoets .....	18
1.2 Vraagstelling van het onderzoek.....	23
1.3 Beschrijving van de hoofdstukken.....	26
<b>Hoofdstuk 2 Beoordelen en beleid</b> .....	29
2.1 Ordenen en classificeren .....	32
2.1.1 Tellen.....	34
2.1.2 Vergelijken in het sociale domein .....	36
2.1.3 Onderwijs en classificaties .....	38
2.2 Testen en ordenen .....	40
2.2.1 Test als institutie.....	42
2.2.2 Test als beleidsinstrument.....	44
2.2.3 Test of toets?.....	46
2.3 Samengevat .....	48
<b>Hoofdstuk 3 Waarde en waardering van onderwijs</b> .....	51
3.1 Onderwijs en samenleving.....	54
3.2 Onderwijs: organisatie van een stelsel .....	57
3.2.1 Sturing .....	58
3.2.2 Structuur.....	59
3.2.3 Kwaliteitszorg/kwaliteitsbewaking .....	60
3.3 Het primair onderwijs in Nederland.....	62
3.3.1 Sturing .....	62
3.3.2 Structuur funderend onderwijs.....	64
3.3.3 Kwaliteitszorg en kwaliteitsbewaking .....	65
3.4 Eindtoets als instrument van sturing, structuur en kwaliteit .....	69
3.4.1 De eindtoets nader beschouwd.....	70
3.4.2 Leerlingvolgsysteem en eindtoets.....	80
3.4.3 Beleidsambities voor eindtoets en LVS .....	82
3.5 Geldigheid van de eindtoets.....	85
3.5.1 Validiteit en selectie.....	89
3.5.2 Validiteit en kwaliteit .....	93
3.6 Samengevat .....	94
<b>Hoofdstuk 4 Onderwijs en selectie</b> .....	97
4.1 Ontwikkeling van onderwijs en toelating.....	100
4.2 Toelating en uitval.....	102
4.2.1 Verbeteren van de aansluiting.....	103
4.2.2 Tests en selectie .....	103

4.2.3	Rangorde of cesuur.....	105
4.3	Intelligentie en tests.....	108
4.3.1	Intelligentietests en Nederland.....	110
4.3.2	Kennen of kunnen als selectieprincipe.....	113
4.4	Selectie en groei van onderwijsdeelname .....	114
4.4.1	Zoeken naar talent.....	115
4.4.2	Onderwijs als kans.....	117
4.4.3	Zorg over onderwijsdeelname.....	118
4.5	Stelselveranderingen.....	120
4.5.1	Mammoetwet.....	120
4.5.2	De Middenschool.....	122
4.6	Selectie en tegenstrijdige beleidsambities.....	123
4.6.1	Uitstel van selectie.....	123
4.6.2	Onderwijs en arbeidsmarkt.....	124
4.6.3	Onderwijs en onderwijsachterstanden .....	127
4.7	Selectie en efficiency.....	129
4.8	De eerste vorm van eindtoetsing.....	133
4.8.1	Het Koninklijk Besluit van 1965 en de Amsterdamse Schooltoetsen .....	133
4.8.2	Opzet van de Amsterdamse Schooltoetsen .....	138
4.9	Samengevat.....	140
<b>Hoofdstuk 5 Onderwijskwaliteit en toetsing .....</b>		<b>143</b>
5.1	Veranderende rol van de overheid.....	146
5.1.1	Overheid op afstand.....	146
5.1.2	Prestaties meten .....	149
5.2	Basisonderwijs en kwaliteit.....	150
5.2.1	Onderwijsbeleid en kwaliteit.....	153
5.2.2	Resultaten volgen.....	158
5.2.3	Rol van de Onderwijsraad .....	160
5.3	Kwaliteit en testen.....	163
5.3.1	Kwaliteit van onderwijs vanuit internationaal perspectief .....	164
5.3.2	Ontwikkelingen in Nederland.....	169
5.3.3	Inspectie van het Onderwijs en toezicht op kwaliteit.....	172
5.3.4	Effect van toezicht gebaseerd op testen .....	175
5.4	Eindtoets en kwaliteit van onderwijs.....	175
5.4.1	Relatief meten en kwaliteit .....	176
5.4.2	Referentieniveaus en kwaliteit .....	177
5.5	Samengevat.....	179
<b>Hoofdstuk 6 Maatschappelijke ordening en testen.....</b>		<b>183</b>
6.1	Maatschappelijke ordening en verdienste .....	186
6.1.1	Stand en structuur.....	187
6.1.2	Intelligentie als dominant concept .....	188
6.1.3	Meritocratie en samenleving .....	189
6.2	Kansengelijkheid en onderwijs.....	192

6.2.1	Kansengelijkheid en onderwijs in Nederland.....	195
6.2.2	Individualisering.....	196
6.2.3	Nieuwe doelgroepen.....	198
6.3	Meritocratie en onderwijs in de praktijk.....	199
6.3.1	Meritocratie als wedstrijd.....	200
6.3.2	Meritocratie en testen.....	201
6.3.3	Eindtoets en LVS in meritocratisch perspectief.....	203
6.4	Samengevat.....	205
<b>Hoofdstuk 7 Rechtvaardigheid en testen.....</b>		<b>209</b>
7.1	Rechtvaardigheid en verdeling.....	212
7.1.1	Schaarste en verdelingen.....	214
7.1.2	Politiek van verdelen.....	217
7.1.3	Arbeidsmarkt en schaarste.....	220
7.2	Verdelingen en selectie in onderwijs.....	222
7.2.1	Vroege selectie.....	223
7.2.2	Selectie en afkomst.....	224
7.2.3	Toegankelijkheid: twee modellen.....	225
7.2.4	Toedelende en remediërende selectie.....	227
7.3	Verdelingen en testen in de praktijk.....	229
7.3.1	Toedeling en intelligentie.....	229
7.3.2	Eindtoets, LVS en verdelingen.....	231
7.3.3	Eindtoets, LVS en effecten voor sociale groepen.....	233
7.3.4	Kansengelijkheid en toetsing.....	235
7.4	Selectie, toelating of plaatsing?.....	236
7.5	Samengevat.....	237
<b>Hoofdstuk 8 Wetenschap en testen.....</b>		<b>241</b>
8.1	Objectiviteit in onderzoek en tests.....	244
8.2	Internationale ontwikkelingen en testen.....	245
8.2.1	Intelligentie en IQ.....	246
8.2.2	Het Coleman-rapport.....	251
8.2.3	No Child Left Behind.....	253
8.3	Empirisch-analytische wetenschapontwikkeling in Nederland.....	254
8.3.1	Intelligentietests en De Groot.....	258
8.3.2	Intelligentietests en onderwijs.....	259
8.3.3	Intelligentietests en onderzoek.....	260
8.4	Testwetenschap.....	261
8.4.1	Betrouwbaarheid.....	261
8.4.2	Equivaleren en standaardiseren van scores.....	263
8.4.3	Validiteit.....	264
8.4.4	Test- en itembias.....	267
8.4.5	Belang van testen.....	272
8.5	Samengevat.....	273

<b>Hoofdstuk 9 Balans en debat</b> .....	277
9.1 De balans opmakend.....	280
9.2 Institutioneel 'oververtrouwen'.....	285
9.3 Eindtoets onder druk.....	286
9.4 Democratisch debat.....	287
<b>Nawoord</b> .....	290
<b>Literatuur</b> .....	293
<b>Summary</b> .....	307
<b>Over de auteur</b> .....	309
<b>Het weefsel der mensheid</b> .....	310

## Woord vooraf

In 1987 kwam ik in het kader van mijn gecombineerde stage en scriptie voor de studie toegepaste taalwetenschap terecht bij Cito, bij het project 'Instaptoets Anderstalige Volwassenen'. Dit project was gericht op de ontwikkeling van een nieuw instrument, een toets om de fase van ontwikkeling van taalvaardigheid van anderstalige volwassenen mee te kunnen vaststellen. De inhoudelijke beschrijving van de meetlat waaraan het instrument gerelateerd zou worden, was de eerste stap die gezet moest worden. Want, zo leerde ik toen, een toets kan pas iets meten als de meetlat gedefinieerd is. Pas later leerde ik dat dit een vorm van criteriumgerelateerd meten was, in tegenstelling tot de dan toe gangbare vormen van meten, namelijk relatieve metingen of normgerelateerde vormen van meten waarbij de vergelijking en ordening van scores centraal staat. De nieuwigheid van het project waar ik begon, bleek meer betrekking te hebben op de vorm van meten dan op de doelgroep, zoals ik aanvankelijk dacht. Zo rolde ik als vanzelf de wereld van het criteriumgerelateerde meten in.

Het lezen van het boekje 'Vijven en Zessen' uit 1966 van de man die gezien wordt als grondlegger van het Cito, A.D. de Groot, was verplichte kost voor iedereen die bij Cito begon. Het onrecht waar volgens De Groot sprake van was in het voortgezet onderwijs, sprak mij, net uit de middelbare schoolbanken, destijds bijzonder aan. En de laatste zin van dat boek 'Wie bindt de kat de bel aan?' bleef hangen. Het is een zin die me in mijn latere professionele leven bij het ontwikkelen van toetsen regelmatig te binnen schoot. Toen ik me, in mijn rol als algemeen directeur bij Bureau ICE, ging bezighouden met de ontwikkeling van een 'alternatieve' eindtoets in 2012, was het een vraag die me niet meer losliet. Ik legde mijn functie neer en besloot me te verdiepen in wat er de afgelopen vijftig jaar op het gebied van eindtoetsing was gebeurd. De uitkomsten van mijn onderzoek leiden als vanzelf terug naar de zin van A.D. de Groot en met dit onderzoek bind ik de kat de bel aan. Met als doel de eindtoets het maatschappelijk debat terug te geven dat deze na zo lange tijd verdient en nodig heeft.

Hoewel het schrijven van een proefschrift te zien is als een individuele proeve van bekwaamheid, is ook dit een test waarbij kansen op succes in hoge mate samenhangen met de context waarin je je bevindt. Het denken over en schrijven van een proefschrift als dit, is de uitkomst van een langdurig proces van lezen, discussiëren, spiegelen, omgaan met weerstand maar vooral van ontwikkelen en groeien. Dat kun je niet alleen. In die zin gaat dit proefschrift niet alleen óver onderwijs maar is het schrijven van een proefschrift onderwijs. Onderwijs dat in de relatie met anderen tot stand komt. Ik wil dan ook allereerst iedereen bedanken die de afgelopen jaren de tijd nam met mij over mijn proefschrift te willen spreken.

Maar ik realiseer me dat ook leermeesters van lang geleden tijdens het schrijven voorbij kwamen met hun wijze lessen van weleer. Zoals 'mijn' meester Kiers van de basisschool. Zijn historische woorden 'Kan niet ligt op het kerkhof en wil niet ligt ernaast' hebben me op moeilijke momenten in dit proces altijd weer moed gegeven om door te zetten. Tijdens mijn studie Toegepaste Taalwetenschappen en nog jaren daarna was het Annemieke Jansen-Van Dielen die met haar wijsheid een onuitwisbare indruk op mij maakte en me voortdurend prikkelde, bleef (be)vragen en nooit genoeg nam met mijn antwoord. Roger Standaert wil ik bedanken voor het vuur dat hij bij mij ontstak en me voorbij de techniek van het testen op een meer politieke manier naar

toetsing leerde kijken. Ferdinand Mertens ben ik grote dank verschuldigd voor de boeken die hij me liet lezen en alle discussies die hij met niet aflatende energie met me bleef voeren en om me meer vertrouwd te laten worden met sociologisch en beleidsmatig denken. Met zijn enorme historische kennis van onderwijsbeleid en toezicht heb ik geprobeerd mijn voordeel te doen. Ben Wilbrink heeft mij met zijn psychometrische kennis en krachtige, eigenzinnige manier van denken vanaf het eerste begin vertrouwen gegeven dat dit onderwerp het meer dan waard was om grondig te onderzoeken en hij was steeds weer bereid mij bij het schrijven met raad en daad terzijde te staan.

Bij de Nederlandse School voor Openbaar Bestuur heb ik mij als buitenpromovendus uitstekend op mijn plek gevoeld. De combinatie van intervisie en onderwijs zoals de NSOB die in het kader van een promotie biedt, hebben mij en mijn onderzoek naar een hoger plan getild. Speciale dank gaat uit naar mijn beide promotoren Paul Frissen en Frans de Vijlder. Gezamenlijk en eensgezind hebben ze mij met een goede mix van vertrouwen, kritische reflectie en uitdagende vragen de begeleiding gegeven die ik en het verhaal nodig hadden.

Ik dank Carlijn Kingma voor de ruimte die ze me heeft geboden om vrijelijk, naar eigen inzicht, gebruik te maken van haar inspirerende kunstwerk 'Het weefsel der mensheid' en daarmee het proefschrift letterlijk een passend gezicht gaf.

De tijd en aandacht die een proefschrift nodig heeft, trekt een wissel op je omgeving. Alleen als daar ruimte en begrip is voor het belang en de waarde die je hecht aan het realiseren daarvan, kan zo'n project slagen. Ik prijs me gelukkig met een meer dan warm vangnet dat me heeft gestimuleerd om deze stap te zetten en vertrouwen heeft uitgestraald in een goede afronding ervan.

Allereerst zijn daar mijn ouders. Ik dank hen voor de liefde waarmee ze mij al meer dan 57 jaar omringen. Zij hebben mij altijd warm en vol vertrouwen gestimuleerd mijn weg te zoeken en mijn grenzen te verleggen. Met het zelf ouder worden neemt mijn dankbaarheid en bewondering voor hun rol in mijn leven alleen maar toe en realiseer ik me wat een voorrecht het is om zulke ouders te hebben. Zij staan dan ook aan de basis van dit proefschrift.

Familie en vrienden vormen een belangrijk fundament van je bestaan. Dat realiseer je je vooral als elkaar zien en spreken niet meer zo vanzelfsprekend is zoals het afgelopen jaar. Ik dank hen voor de belangstelling die ze steeds maar weer hadden voor mij en dit langdurige proces waarvan uitkomst noch tijdstip van afronding helder was. Hoewel ik niet altijd veel voortgang kon melden, hielp hun betrokkenheid me wel met het willen doorgaan met de uitdaging.

Hier, in dit voorwoord, wil ik speciaal Elizabeth, Renske en Robinet danken voor de manier waarop ze mij in mijn leven en ook weer tijdens dit traject terzijde hebben gestaan. Wat fijn om vriendinnen te hebben die vanuit liefde en betrokkenheid je zowel privé als professioneel weten te spiegelen en te corrigeren waar het moet en kan, en tegelijkertijd te weten dat ze er altijd voor je zijn als je ze nodig hebt. Dank voor alle gesprekken, het 'tegedenken' en het vertrouwen dat jullie altijd hebben gehad in het slagen van dit onderzoek.

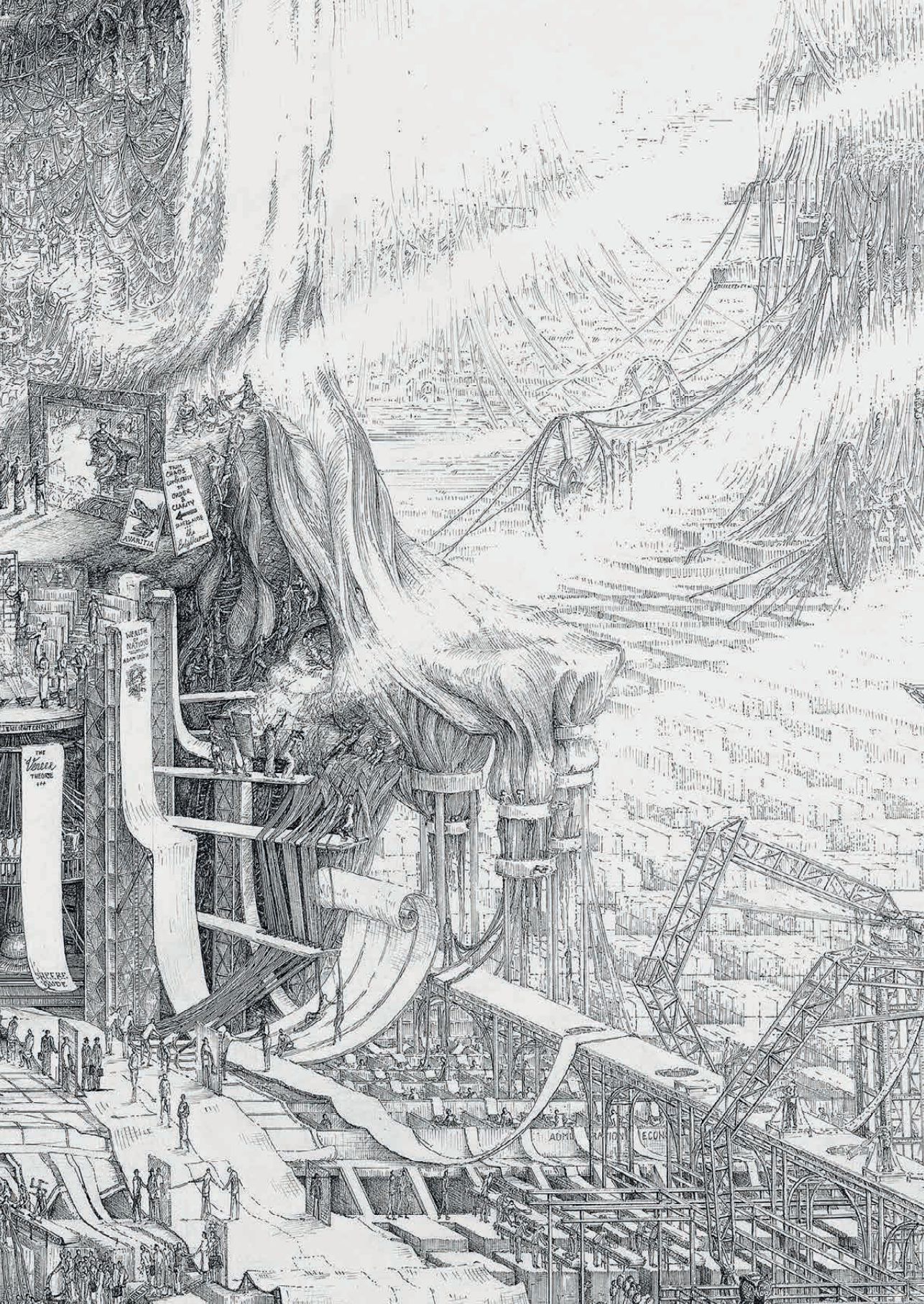
Het stevige fundament van liefde en vertrouwen waarop ik sta, heb ik geprobeerd mee te geven

aan mijn kinderen, Suzanne en Branko. Wat een voorrecht om samen met hen op te mogen groeien en samen met hen mee te bewegen in hun zoektocht naar hun plek in het leven. Hun ervaringen en verhalen over onderwijs en werk hebben mij ook professioneel altijd gevoed. Onze liefde, steun en betrokkenheid op elkaar maken me blij en sterk. Iedere dag. Dank dat jullie zijn wie jullie zijn!

Met de liefde kwamen er ook nieuwe kinderen in mijn leven. Samen met hun vader kwamen eerst Lara, Sander en Stefan. Nog weer later voegden Imre, Kelvin, Hella en Jente zich bij het groter groeiende gezin. Behalve veel gezelligheid hebben zij met hun verhalen over onderwijs ook nieuwe bouwsteentjes voor dit verhaal gebracht. Dank jullie voor de plek die jullie in mijn leven zijn gaan vervullen.

Tenslotte wil ik Alex bedanken, mijn man. Ik ga geen woorden kunnen en willen vinden om mijn liefde voor hem hier uit te drukken en volsta met te zeggen dat hij dit proefschrift mogelijk heeft gemaakt door te zijn wie hij is. Masha danki, dushi!

Dit proefschrift draag ik op aan mijn kleindochter Josephine van wie ik onmetelijk veel ben gaan houden. Haar geboorte heeft mij gestimuleerd mijn bakens te verzetten en dit proefschrift te schrijven. Voor haar, voor haar broertje Boris, voor alle kleinkinderen die nog zullen volgen, en voor alle kinderen die met hen opgroeien, hoop ik van harte dat het onderwijs een plek kan zijn waar samen opgroeien centraal staat, niet in competitie maar in verbondenheid.



THE CHAIR OF TRUTH  
ON WHICH  
CLARITY  
SHINES  
IN THE  
HEAVENS

THE HOUSE OF WISDOM  
AND KNOWLEDGE

THE HOUSE OF WISDOM  
AND KNOWLEDGE

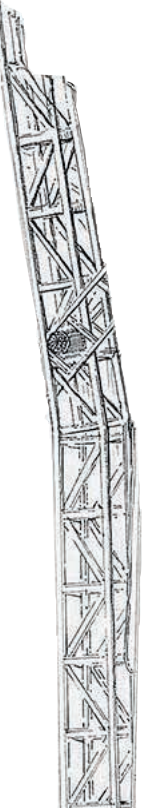
ADMINISTRATION  
ECONOMY





# HOOFDSTUK 1

'De Cito'



## Hoofdstuk 1 'De Cito'

'Ik wist dat er veel zou afhangen van de [...] Cito-toets. We hadden het maandenlang nergens anders over gehad in de klas. Op momenten dat ik wakker lag van de spanning, probeerde ik mezelf in te praten door aan de woorden van meester Gregory te denken, die me meermaals openlijk had geprezen om mijn schrijf- en leesvaardigheid.[...] Ik wist dat een slecht resultaat bij de Cito-toets mijn ouders zou teleurstellen.[...] Dat stond er dus voor mij op het spel, samen met de angst om voor altijd in middelmatigheid te leven. Ik wist dat ik boven mezelf moest uitstijgen en vooral de schade bij het onderdeel rekenen moest beperken om niet door het ijs te zakken. Sinds de derde klas had ik een moeizame relatie met rekenen. Ondanks de bijlessen van meneer Rolf wist ik me nog steeds geen raad met staartdelingen en breuken, en meester Ronald, mijn leraar in de derde en vierde klas, miste het geduld en de empathie om voor verlichting te zorgen.[...] Meester Ronald weet mijn onvermogen om de rekensommen op te lossen aan luiheid en desinteresse. Regelmatig liet hij me nablijven, net zo lang tot ik huilend willekeurige antwoorden invulde in mijn schrift en met een waarschuwing op zak naar huis werd gestuurd. Zo kwam het dat ik sindsdien al sidderde wanneer een overhoring voor rekenen werd aangekondigd. Daarom vreesde ik de Cito-toets. Ik wist dat ik op het onderdeel taal meer dan voortreffelijk moest scoren om de rampspoed die mij met rekenen wachtte niet fataal te laten worden voor mijn eindresultaat. Op de eerste grote dag van mijn leven zat ik ontspannen aan het tafeltje waar ik al veel complimenten van meester Gregory had ontvangen. Ik concentreerde me hard toen ik me richtte op de rekenvragen, gokte hier en daar wanneer ik een antwoord niet wist. De taalvragen kwamen me zo eenvoudig voor dat ik ervan overtuigd was dat ik iets over het hoofdstuk zag. Daarom las ik sommige vragen wel drie keer, twijfelde ik even aan het antwoord dat ik had gegeven, maar ging ik daarna verder omdat de tijd begon te dringen. Na afloop wist ik zeker dat ik de Cito-toets uitstekend had gemaakt en zag ik mijn ouders in gedachten al trots glimlachen. Maar in de dagen die volgden, brokkelde dat goede gevoel langzaam af toen ik me afvroeg of ik niet te veel had gegokt bij het onderdeel rekenen en of ik de taalvragen niet had onderschat. Zo kwam het dat ik de ochtend van de uitslag wakker werd met een steek in mijn maag die zich langzaam verspreidde naar de rest van mijn lichaam. Het was misschien wel de belangrijkste dag van mijn leven tot dan toe, want vandaag zou duidelijk worden of ik in de voetsporen van mijn zus zou treden, vandaag zou ik te horen krijgen of ik net als zij naar het vwo mocht.[...] Er werden al snel scores voorgelezen door mijn klasgenoten. '520,' riep Ishana als eerste. Orlando volgde: 'Shit, ik heb 517.' Niet veel later riep Esmeralda dat ze 533 punten had. Toen Saleem zijn briefje kreeg, keek hij er een tijdje in stilte naar. Net als een paar anderen weigerde hij zijn score hardop met de klas te delen. Fenuku begon heen en weer te wippen op zijn stoel nadat hij zijn briefje in ontvangst had genomen. Hij keek alsof hij zojuist te horen had gekregen dat hij terug moest naar Ghana. Toen meester Gregory eindelijk aan mijn tafel stond, zei hij: 'Goed gedaan, Metin.' Hoewel ik voorvoelde dat ik rekenen had verprutst, kreeg ik weer vertrouwen door zijn woorden en hoopte ik heimelijk op een score van ver boven de 540. Maar wat ik had gevreesd, was uitkomen: rekenen had mij genekt. Ik had weliswaar uitmuntend gescoord op het onderdeel taal en hoorde met mijn score bij de drie besten van de klas, maar toch was ik teleurgesteld. 'Wat heb je gehaald?' vroeg Saleem. '531,' antwoordde ik. 'Jezus, waarom kijk je dan zo teleurgesteld?' Saleem stompte me op mijn bovenarm. 'Wees blij, man.' Op gedempte toon voegde hij eraan toe: 'Mijn score is 520.' Ik kon Saleem niet vertellen dat hij mijn referentiekader niet was. Ik had me al die tijd gericht op de score van mijn zus. Zij had 540 punten gehaald en ik zat daar ver onder en wist dat het nu nog maar de vraag was of ik naar het vwo kon.

*Meester Gregory weigerde te vertellen op welk niveau hij ons zou laten instromen op de middelbare school. 'Dat ga ik op de ouderavond met jullie ouders bespreken, maar voor de meesten van jullie zal het geen verrassing zijn, gelet op jullie score.' Ik kon aan hem zien dat hij teleurgesteld was over de algehele uitslag in onze klas.'*

*Uit: Murat Isik, 'Wees Onzichtbaar' (2017)*

Dit is het relaas van Murat Isik over zijn ervaringen met 'de Cito' zoals de eindtoets basisonderwijs door veel leerlingen en hun ouders wordt genoemd. Dit is een verhaal zoals veel kinderen dat kunnen vertellen want bij een toets met zo'n lange historie heeft vrijwel iedereen een verhaal over 'zijn cito'. Wat de verhalen gemeen hebben is de impact die de afname en uitkomst van de toets voor hen heeft gehad, voor hun levensloop, hun persoonlijke ontwikkeling en hun zelfbeeld. Hoe een toets die in een eerste versie in 1966 is verschenen, heeft kunnen uitgroeien tot een zo bepalend instrument voor het leven van veel kinderen en hoe terecht dat is, staat centraal in dit proefschrift.

### 1.1 De werking van de eindtoets

In Nederland sorteren wij met behulp van een toets al ruim vijftig jaar aan het eind van de basisschool onze kinderen voor het vervolgonderwijs op een vergelijkbare manier: op basis van hun score op een combinatie van taal- en rekentoetsen worden leerlingen gerangschikt van hoog naar laag scorend. Die ordening heeft grote gevolgen voor de levens van de kinderen. Want deze ordening wordt gebruikt voor een sortering voor vervolgonderwijs. Er zijn verschillende 'bakjes' in die sortering en die bakjes worden elk jaar volgens een vaste systematiek gevuld zodat die verdeling, over het gehele stelsel gezien, elk jaar ongeveer hetzelfde is, of het onderwijsniveau nu daalt of niet. En die bakjes hangen in hoge mate samen met posities op de maatschappelijke ladder. Kinderen wedijveren met elkaar voor een plek in de verdeling. De kans om in een bepaald bakje terecht te komen, hangt samen met de sociaal-economische positie van de ouders. Gezien de betekenis van de sortering voor de levenskansen van de kinderen zitten alle betrokkenen op het vinkentouw; ouders, kinderen, leerkrachten en scholen. Want op basis van de scores van de leerlingen worden vervolgens ook de scholen gerangschikt. Dus ook voor hen geldt dat er veel gemoeid is met het behalen van hoge resultaten. Zo is de eindtoets een jaarlijks terugkerend ritueel met juichende winnaars en beschaamde verliezers.

De selectie in ons onderwijs aan het eind van het basisonderwijs lijkt op het eerste gezicht bedoeld om kinderen vervolgonderwijs op maat van hun kunnen aan te bieden ten einde 'recht te doen aan de talenten, ontwikkeling en capaciteiten van leerlingen' (OCW, 2020: 2): eerlijk, objectief en op maat van het kind een passend vervolgadvis. Vanuit dat perspectief kan ouders die meer willen dan hun kind aankan verweten worden dat ze te hard pushen. Maar zo eenvoudig ligt het niet. Het gaat om iets anders: de positie op de maatschappelijke ladder en de bijbehorende status. 'Stop met 'omhoog drukken' van kinderen naar de hoogste schoolniveaus' waarschuwt minister Slob op 9 februari 2018 in het AD, daaraan toevoegend dat het vmbo 'niet een soort vergaarbak is van kinderen die niet naar havo of vwo kunnen, en dus maar naar het vmbo moeten. Dit is geen restonderwijs.' Het is inderdaad lastig te spreken over 'rest'onderwijs als ongeveer de helft van de leerlingen aan het eind van het basisonderwijs een advies krijgt voor het vmbo. Met het 'omhoog drukken' verwijst de minister naar een proces dat gaande is en door menig ouder als zodanig wordt herkend: alles doen wat binnen het vermogen ligt om een

zo hoog mogelijk advies te krijgen voor je kind. Het wijzen met de vinger naar ouders die zich inspinnen om hun kinderen van een havo of nog liever een vwo-schooladvies te voorzien, is snel gedaan, maar levert weinig inzicht op in wat er aan de hand is. De realiteit is dat ouders meer en meer bereid zijn te investeren in 'schaduwonderwijs', zoals het geheel van bijlessen, huiswerkbegeleiding en toets-trainingen genoemd wordt, om een zo hoog mogelijk plaatsingsadvies binnen te halen. De selectie op 12-jarige leeftijd heeft namelijk grote gevolgen voor de toelating tot vervolgonderwijs en de daar later aan gekoppelde positie op de arbeidsmarkt en in de samenleving. Binnen een 'relatief vroeg selecterend onderwijsstelsel' - zoals de OECD het Nederlandse stelsel typeert - wordt al op jonge leeftijd bepaald wat de route wordt die een kind mag of moet gaan lopen (OECD, 2016). De opwaartse druk waar minister Slob zich druk om maakt is begrijpelijk maar het is te eenvoudig om van ouders te verlangen hun ambities bij te stellen. Waar het in de kern om draait is dat onderwijs, en dan vooral de behaalde diploma's, meer en meer bepalend zijn geworden voor de positie die op termijn in de samenleving kan worden ingenomen. Het advies aan het eind van het basisonderwijs doet er steeds meer toe voor de kansen die een kind heeft op rechtstreekse doorstroming naar de hoogste vormen van vervolgonderwijs en daarmee naar een hogere sport op de maatschappelijke ladder. Onderwijs is een competitie waarbij het aantal kinderen dat bovenaan de piramide terecht kan komen beperkt is. Het is dus vechten om een plek op de ladder naar de top. Kinderen kunnen op de leeftijd van twaalf jaar niet overzien hoe groot de gevolgen zijn van het op dat moment belanden in het maatschappelijk meest kansrijke bakje. En ouders, met name hoogopgeleide ouders, weten dat wel. En werken daaraan. Dit heeft vooral te maken met de maatschappelijke positie die diplomering op de hogere niveaus van ons onderwijs met zich meebrengt. In het boek 'De Bijlesgeneratie' stelt Elffers de treffende vraag: 'Mag onze koning naar het mbo? Zouden we in Nederland minder geneigd zijn onze koning als staatshoofd te accepteren als hij een mbo-opleiding had gevolgd?' (Elffers, 2018:17).

De selectie op basis van de eindtoets in samenhang met het gegeven schooladvies heeft grote gevolgen. Voor de leerling en zijn ouders maar ook voor de leerkracht en de school. Want de schooladviezen, gebaseerd op testcores, worden gebruikt als maat voor de bepaling van de kwaliteit van het gegeven onderwijs. Daarmee zijn de belangen die samenhangen met de toetsing aan het eind van groep 8 op basis van de eindtoets in combinatie met de uitkomsten van de toetsen van een leerlingvolgsysteem (LVS) ook voor leerkrachten en scholen groot. Met de uitkomsten van de eindtoets wordt niet alleen een oordeel gegeven over de leerling maar ook over zijn genoten onderwijs. Dit maakt het vraagstuk complex. De eindtoets en de werking ervan is onderdeel van een krachtenveld met belangen en belangenbehartigers. Een onderzoek naar de eindtoets vraagt dan ook om onderzoek naar de interactie die ontstaan is en bestaat in het geheel van krachten die van invloed zijn op onderwijs en toetsing.

De afgelopen eeuw zijn de mogelijkheden op het volgen van voortgezet onderwijs voor iedereen, ongeacht afkomst of geslacht, uitgebreid. Tegelijkertijd is het belang van een succesvolle schoolloopbaan de afgelopen eeuw steeds groter geworden. Is de hoogste positie in de samenleving anderhalve eeuw geleden nog vanzelfsprekend voorbehouden aan de aristocratie, vanaf het begin van de vorige eeuw verandert dat. Waarden ten aanzien van de verdeling van goederen veranderen: afkomst mag niet langer bepalend zijn voor het verkrijgen van (schaarse) goederen. Belangrijke posities in de samenleving moeten worden verdeeld op basis van verdiensten ('merites' in het Latijn). Diploma's als bewijs van verdiensten hebben daarmee een steeds belangrijker

functie in het verwerven van hogere maatschappelijke posities. Zo worden tegenwoordig posities aan de bovenkant van de maatschappelijke ladder vaker bekleed door mensen die op basis van gevolgd onderwijs hun posities hebben weten te verwerven. Het feit dat sinds de tweede helft van de vorige eeuw vrijwel iedereen een vorm van voortgezet onderwijs kan volgen, betekent echter niet dat er voor iedereen ook een plek naar keuze gerealiseerd kan worden. Hoewel er sprake is geweest van een expansie van onderwijs en deelname zijn de plekken die recht geven op doorstroom naar de hogere vormen van vervolgonderwijs en hogere maatschappelijke posities relatief nog altijd in de minderheid. Met een dergelijke schaarste gaat een verdelingsvraagstuk gepaard; hoe dit schaarse goed te verdelen? Al vanaf de jaren zestig is het percentage leerlingen dat doorstroomt naar eerst v.h.m.o. en later vwo stabiel: ongeveer een vijfde van alle leerlingen krijgt het advies voor de hoogste vorm van vervolgonderwijs. Na de komst van het vmbo heeft de verdeling, ook voor de andere vormen van voortgezet onderwijs, zich gestabiliseerd; 20% van de leerlingen aan het eind van het basisonderwijs krijgt op basis van de score op de eindtoets een vwo-advies, 30% een havo-advies en de overige 50% krijgt een verwijzing richting een variant van het vmbo. Hoe hoger de score, hoe groter de kans op een (toets)advies voor een plek in de hogere vormen van voortgezet onderwijs. Dit advies, gebaseerd op de uitkomsten van een eindtoets, wordt gecombineerd met een advies van de school. De volgorde van advies (eerst de toets en dan het schooladvies of andersom) staat politiek ter discussie, en niet het feit dat scholen geacht worden hun advies meer te baseren op uitkomsten van toetsen van een leerlingvolgsysteem (LVS). Terwijl dit ertoe heeft geleid dat feitelijk de verdeling van onderwijskansen steeds meer afhangt van een beperkte set parameters waarop het onderwijs selecteert, namelijk taalvaardigheid, en dan vooral begrijpend lezen, in combinatie met rekenvaardigheid. En waarmee steeds vroeger wordt geselecteerd. Want verschillen tussen kinderen worden door het verplichte gebruik steeds eerder geïdentificeerd. Het verplicht laten meewegen van de uitkomsten van gestandaardiseerde toetsen van een leerlingvolgsysteem (LVS, verplicht vanaf groep 3) in het schooladvies zorgt er daarbij voor dat verschillen in de vroegste ontwikkeling een steeds groter effect hebben op de latere levenskansen.

De huidige cijfers voor plaatsing in het voortgezet onderwijs laten zien dat zich in de afgelopen eeuw een proces voltrokken heeft dat consolidatie van maatschappelijke posities weliswaar in de loop van de tijd anders van karakter heeft gemaakt, maar niet minder krachtig. De inspectie van het Onderwijs wijst in haar jaarlijkse rapportage over de staat van het onderwijs op het feit dat er een nieuwe kansenongelijkheid aan het ontstaan is waarbij het opleidingsniveau van ouders de nieuwe scheidslijn vormt: 'Net als voorgaande jaren krijgen kinderen van ouders met een opleidingsniveau van maximaal mbo-2 voornamelijk vmbo-adviezen terwijl kinderen van ouders met een hbo-master of wo-opleiding vooral havo- en vwo-adviezen krijgen.' (Inspectie van het Onderwijs, 2020: 15). Leerlingen waarvan 'de ouders wetenschappelijk zijn opgeleid blijven minder vaak zitten, worden hoger geplaatst en stromen na het diploma vaker door binnen het voortgezet onderwijs' (Inspectie van het Onderwijs, 2020: 17). Voor leerlingen met een niet-westerse migratieachtergrond stelt de Staat van het Onderwijs dat zij 'nu vaker dan vijf jaar geleden hogere onderwijsniveaus volgen maar dat het aantal vwo'ers onder hen nog wel beperkt is' (Inspectie van het Onderwijs, 2020: 18).

Als we kijken hoe nu selectie en de daarmee samenhangende allocatie verloopt, dan is feitelijk vast te stellen dat de verwijzing naar het voortgezet onderwijs in het voordeel uitvalt van de

kinderen van hoger opgeleiden: zij hebben nog altijd meer kans een havo/vwo-advies te krijgen en komen vaker hoger op de ladders terecht, zelfs bij gelijke geschiktheid. De 'witte raven' die ondanks de lagere vooropleiding van hun ouders via het selectiesysteem wel de hoogste niveaus bereiken, hebben de afgelopen decennia de indruk gewekt dat het systeem 'eerlijk' werkt en dat de eindtoets hieraan een belangrijke bijdrage levert. Maar vijftig jaar eindtoets en inspanningen om kansengelijkheid met behulp van onderwijs te bevorderen hebben niet kunnen voorkomen dat onderwijsminister Slob op 13 maart 2019 een zorgelijke brief schreef aan de Tweede Kamer over de noodzaak om kansengelijkheid in het onderwijs aan te pakken. 'De Inspectie van het Onderwijs (hierna: inspectie) constateert de laatste jaren dat kinderen van ouders met een lager inkomen en minder diploma's lang niet altijd in staat worden gesteld het niveau te bereiken dat zij cognitief aankunnen. Daar willen we ons niet bij neerleggen. Daarom is het een van de voornaamste ambities van dit kabinet om kansengelijkheid in het onderwijs te bevorderen. [...] Met het onderwijs geven we mede vorm aan onze toekomstige samenleving. Dit kabinet staat voor een inclusieve samenleving waarin iedereen mee kan doen en iedereen een eerlijke kans heeft om iets van haar of zijn leven te maken.' (Ministerie OCW, 2019a). In datzelfde voorjaar komt het Centraal Plan Bureau met een rapport waarin geadviseerd wordt de eindtoets eerder af te nemen zodat de eindtoets leidend wordt bij het advies dat de basisscholen geven zodat leerlingen meer gelijke kans hebben op een passend schooladvies (CPB, 2019). In tweede helft van 2020 doet de Onderwijsraad een oproep via sociale media<sup>1</sup> om mee te denken over differentiatie en selectie in het onderwijsstelsel in relatie tot kansengelijkheid als voorbereiding op een advies waar onderwijsminister Slob de Onderwijsraad om heeft gevraagd. In het advies zal het gaan om de vraag hoe vergaande differentiatie en selectie kansengelijkheid kan vergroten in plaats van deze te beperken. In de praktijk blijkt dat het als kind van lager opgeleide ouders nog altijd lastig is om binnen het huidige selectiesysteem een even grote kans te hebben op de hoogste vormen van vervolgonderwijs in vergelijking met kinderen van hoog opgeleide ouders. De eindtoets speelt daarin een belangrijke rol. Deze toets vergelijkt de scores van kinderen en geeft inzicht in wie er meer of minder scoort op wat er in de eindtoets aan bod komt. Deze uitkomsten verworden tot scores die in de beleving van betrokkenen een beeld geven van potentiële geschiktheid in termen van wie slim is en wie niet en wie in aanmerking mag komen voor de meest kansrijke posities in het vervolgonderwijs. Opleidingsniveau is vooral een relatief gegeven, een positioneel goed. Het gaat er niet om hoeveel onderwijs iemand heeft gevolgd maar om hoeveel meer of minder onderwijs iemand heeft genoten in vergelijking met zijn concurrenten op de arbeidsmarkt: hoe meer je ervan hebt, hoe groter de kans op nog meer en uiteindelijk op een goede positie in de samenleving. Voor een individu is het van groot belang als voorste in de wachtrij te belanden. Zeker omdat er niet van alle vormen van onderwijs evenveel beschikbaar is. Zouden de meest begeerde onderwijsplekken breder beschikbaar zijn, dan verliezen ze hun waarde als bepaler van een hoge status. Vandaar dat in de toenemende onderwijsdeelname een tragiek schuilt: hoe meer mensen eraan deelnemen, hoe minder de waarde ervan zal zijn. Het onderwijs kan daarop reageren door het onderwijsaanbod te differentiëren en de toegankelijkheid ervan te beperken of juist door het aanbod uit te breiden zodat er nieuwe mogelijkheden komen om onderscheid te maken. Zo ontstaat nieuwe schaarste waar vooral de sterkeren, goed op de hoogte van wat er gaande is, van weten te profiteren. Zij bereiden zich goed voor en doen hun best zo hoog mogelijk te eindigen. Het vergelijken zoals ondersteund wordt met toetsen zoals de eindtoets, creëert een voortdurende competitie in het voordeel van de sterkeren die zoveel mogelijk proberen hun posities te behouden.

1] <https://www.onderwijsraad.nl/adviezen/adviezen-in-voorbereiding/differentiatie-en-selectie>

Naast hulpmiddel voor de verdeling van onderwijsposities is de eindtoets in de laatste decennia uitgegroeid tot een instrument waaraan de kwaliteit van onderwijs wordt afgemeten. De Inspectie van het Onderwijs is vanaf het begin van deze eeuw scores op de eindtoets en scores op de gestandaardiseerde voortgangstoetsen als prestatie-indicatoren voor scholen en hun leerkrachten gaan hanteren. Niet alleen leerlingen worden op basis van toetsscores met elkaar vergeleken. Op grond van dezelfde data worden ook scholen, en in hun kielzog de leerkrachten, aan vergelijkingen onderworpen. De aldus verkregen informatie wordt in het kader van transparantie openbaar gemaakt wat gevolgen heeft voor de reputatie. Een 'reputatie-wedloop' wordt hiermee gecreëerd en aangewakkerd met de gedachte dat competitie en kwaliteit positief op elkaar inwerken. Daarmee hebben zowel scholen als leerkrachten als leerlingen (en hun ouders) belang bij hoge scores.

De eindtoets in combinatie met gestandaardiseerde voortgangstoetsen van het LVS gedurende het onderwijs is uitgegroeid tot meer dan zomaar een toets. Het is een steeds langer durende vorm van selectie geworden. De combinatie van metingen is de bril waarmee we naar de werkelijkheid zijn gaan kijken en vervolgens is die manier van kijken de werkelijkheid gaan vormen.

Dit onderzoek is te beschouwen als een zoektocht naar de eindtoets zoals die in de afgelopen eeuw is ontstaan en nu de werkelijkheid mede vorm geeft. Met behulp van het onderzoek wordt (weer) zichtbaar gemaakt wat er schuilgaat achter een instrument dat in de loop van de tijd ingebed is geraakt in een beleidsproces gericht op het verdelen van (schaarse) onderwijsposities. En welke keuzes zijn gemaakt bij de totstandkoming van het instrument zelf maar

*'Toen ik in een van mijn eerste jaren voor de klas zogenoemde midden- en eindtoetsen moest afnemen, kleurde driekwart van de scores oranje/rood en moest ik op het matje komen bij de intern begeleider. Daar sta je dan met tranen in je ogen, want ik heb toch gewoon goed lesgegeven? De kinderen hebben de doelen toch gehaald? Wist ik veel dat er zo zwaar getild werd aan dit soort toetsen. Dus ga je in het vervolg 'oefenen'. Net als die enthousiaste juf en al die andere leerkrachten. Want afgerekend worden door directie, bestuur en ouders, dat wil je niet meer. Ik praat erover met collega's uit het hele land. Een juf vertelt: 'Persoonlijk vind ik het grootste probleem dat het gemiddelde van die toetsen als norm wordt gehanteerd. En daarom gaan ouders en leerkrachten er alles aan doen om boven dat gemiddelde uit te komen. Dat oefenen resulteert in nog hogere scores en vervolgens wordt de normering weer aangepast. Zo blijven we in die vicieuze cirkel ronddwalen.' 'We toetsen niet de doelen die de kinderen zouden moeten beheersen!', aldus een andere leerkracht. 'De kinderen presteren altijd anders dan verwacht en het geeft geen realistisch beeld. Het wrange is ook dat scholen ambities opstellen aan de hand van die resultaten.' 'Moeten we allemaal hactivisten en vwo'ers afleveren? Wat is de wereld zonder tegelzeters en kapsters? Een kind help je door een passende score te geven en niet een zo hoog mogelijke score.' Inderdaad. Dat 'oefenen' voor rekenen en spelling heb ik trouwens maar even volgehouden omdat het resulteerde in niveaus waarin mijn leerlingen op hun tenen moesten lopen. Sindsdien besteed ik enkel aandacht aan het werken in een toetsboekje, want het is verplicht dat we die toetsen afnemen. Het toepassen van rekenvaardigheden in een context en correct spellen heeft mijn aandacht. Maar dat is onderdeel van ons lesprogramma.*

*Om die toetsen maak ik me dus niet meer druk. Ik vertrouw op de kracht van mijn leerlingen en de kennis en kunde die ik heb als leerkracht. Als we blijvend worden afgerekend op slechte resultaten en ons mee laten voeren in die idiote toetsgekte, is het onderwijs echt gezakt naar een dieptepunt.'*

**Trouw, Naomi Smits, 5 juni 2019**



ook bij de inzet ervan in de beleidspraktijk. De toedeling van leerlingen aan voortgezet onderwijs, de daarbij behorende processen in relatie tot kansengelijkheid en de beoordeling van de kwaliteit van het gevolgde onderwijs zijn dermate grote, en voor alle betrokkenen belangrijke thema's dat democratische beschouwing en besluitvorming erover van belang zijn. Daarvoor is allereerst een gefundeerd beeld nodig van wat de toets inhoudt, wat de toets doet en wat de maatschappelijke gevolgen zijn. De deugdelijkheid van de eindtoets kan dan vervolgens bevestigd worden. Die deugdelijkheid kan beschouwd worden vanuit het perspectief van doelen en ambities van onderwijs maar ook breder vanuit perspectieven van rechtvaardigheid.

## 1.2 Vraagstelling van het onderzoek

De eindtoets<sup>2</sup> vervult volgens de huidige wetgeving (WPO, artikel 9b) nu twee functies: de eindtoets fungeert als selectie-instrument bij de overgang van basis- naar voortgezet onderwijs en fungeert als maat voor de bepaling van de 'opbrengsten' van onderwijs en is daarmee een belangrijke indicator voor het bepalen van de kwaliteit van onderwijs. Die tweede functie werkt twee kanten op. De eindtoets levert informatie op waarmee scholen zelf hun eigen functies kunnen evalueren en de uitkomsten van de eindtoets worden door de Inspectie van het Onderwijs gebruikt voor toezicht op en beoordeling van de kwaliteit van onderwijs (Ministerie OCW, 2011a). Scores op de eindtoets hebben daarmee gevolgen voor allen die bij het onderwijs betrokken zijn. Voor leerlingen en hun ouders is de toets in hoge mate (mede-)bepalend voor de mogelijkheden voor vervolgonderwijs. Voor docenten, scholen en schoolbesturen ligt op basis van de scores een oordeel in het verschieft over de kwaliteit van het geboden onderwijs, hetgeen door de openbaarheid ervan gevolgen kan hebben voor nieuwe aanmeldingen. En op landelijk niveau kunnen de onder regie van de Inspectie van het Onderwijs tot stand gekomen beoordelingen aanleiding vormen voor een oordeel over de kwaliteit van het stelsel als geheel. Daarmee rust een veelheid van belangrijke, te vervullen taken op de schouders van de eindtoets.

De eindtoets is in vijftig jaar uitgegroeid tot een onder regie van de overheid ontwikkeld en door de overheid verplicht gesteld instrument. Daarmee is de eindtoets uitgegroeid tot enerzijds een belangrijk beleidsinstrument waarmee de overheid in belangrijke mate inhoud en vorm geeft aan haar huidige invulling van een deel van artikel 23 van de Grondwet, het toezicht houden op de kwaliteit van onderwijs. Anderzijds speelt de eindtoets een belangrijke rol in de verdeling van (onderwijs)kansen. Daarmee gaat het belang van de toets verder dan het onderwijs zelf en speelt deze een rol in de maatschappelijke ordening zoals die in Nederland tot stand komt. De eindtoets geeft daarmee uitdrukking aan normatieve doelstellingen die de maatschappelijke orde

2] Bij de indiening van het wetsontwerp waarin de verplichte afname van de eindtoets is opgenomen (2011), komt verzet tegen het feit dat alle scholen verplicht worden dezelfde toets te gebruiken. De als verplicht voorgestelde eindtoets, is de basiseindtoets van Cito die vanaf de aanname van de wet over zal gaan in handen van het ministerie en daarmee formeel een 'overheidstoets' zal worden met de naam 'Centrale Eindtoets'. Deze toets wordt onder eindverantwoordelijkheid van het College voor Toetsen en Examens (CvTE) ontwikkeld waarbij Stichting Cito zal worden belast met de daadwerkelijke ontwikkeling van de eindtoets. Op basis van een motie moet de wet worden aangepast zodat het mogelijk wordt dat naast de overheid ook ruimte wordt geboden aan andere toetsontwikkelaars om te komen met een eindtoets zoals blijkt uit het Besluit van 3 juni 2014, houdende regels over de centrale eindtoets of andere eindtoetsen in het primair onderwijs. Op dit moment zijn er naast de Centrale Eindtoets nog vier andere eindtoetsen waaruit scholen kunnen kiezen.

nastreeft ten aanzien van de verdeling van kansen. Is de verdeling van kansen zoals die nu tot stand komt op basis van de eindtoets de verdeling van kansen zoals beoogd wordt? Doen we wat we zeggen dat we doen? De eindtoets is meer dan een technisch hulpmiddel; de eindtoets laat zien wat relevant wordt geacht en wat niet in het kader van die verdeling. Daarmee normeert en beïnvloedt de eindtoets dit proces. De eindtoets is zo beschouwd behalve een instrument ook een institutie in sociologische zin waarmee bedoeld wordt dat de eindtoets een instrument is dat mede bepalend is voor de manier waarop de samenleving geordend wordt. Politicoloog Lehning wijst op het feit dat instituties in iedere sociaalpolitieke orde een centrale rol spelen en niet louter een instrumentele rol. Instituties moeten niet beschouwd worden als zuiver technische hulpmiddelen om bepaalde maatschappelijke doelstellingen te bereiken. Instituties geven uitdrukking aan de normatieve doelstellingen die de maatschappelijke orde nastreeft waarmee elke institutie naast een empirisch ook een normatief aspect kent. En dat normatieve keuze-element hangt samen met het mensbeeld dat gehanteerd wordt en de rationaliteitsopvatting die als uitgangspunt wordt genomen (Lehning, 1991: 213). Het is van belang dat er debat is over de vraag hoe institutionele arrangementen er uiteindelijk uit moeten zien. Dit debat ontbreekt veelal volgens Lehning: 'Wie in de bestuurskunde over instituties die ons samenleven vormgeven wil spreken, kan over de normatieve implicaties van die instituties niet zwijgen.' (Lehning, 1991: 221). De lange ontstaansgeschiedenis van de eindtoets, ingebed in verschillende periodes van denken en handelen, heeft het aanvankelijke debat verstomd, heeft de keuzes die eraan vooraf zijn gegaan aan het zicht en daarmee aan de discussie onttrokken en de normatieve implicaties ervan voor een belangrijk deel onbesproken gelaten.

Dit onderzoek richt zich op een reconstructie van de werkelijkheid zoals die met behulp van de eindtoets gestalte heeft gekregen. Daarbij wordt de eindtoets beschouwd als een institutie met een technische, een instrumentele maar ook een normatieve doelstelling. Dit betekent dat de eindtoets wordt onderzocht zowel op de meer instrumentele rol die de toets vervult maar ook op het normatieve keuze-element dat erin besloten ligt, dat samenhangt met het mensbeeld dat gehanteerd wordt en de rationaliteitsopvatting die daarbij tot uitgangspunt is genomen. Behalve een instrument is de eindtoets ook een verhaal, of beter, een geheel van verhalen die stuk voor stuk van invloed zijn of zijn geweest op de economische, sociale en culturele werkelijkheid die nog altijd mede op basis van de eindtoets vormgegeven wordt en die daarmee weer het belang van het instrument als zodanig onderstrepen. Wat telt en vertelt de eindtoets, welke keuzes gaan daarachter schuil, welke aannames zijn er (gedaan) en wat is de boodschap die wordt overgebracht? Tellen en vertellen liggen niet zo ver uit elkaar als op het eerste gezicht wellicht lijkt. De totstandkoming van besluitvorming over en de uitvoering van beleid kunnen juist worden gezien als een 'voortdurend samenspel waarbij cijfers en feiten en verhalen en frames met elkaar verweven raken' (De Jong et al., 2018: 24). Soms staan tellingen aan de wieg van verhalen en soms is het omgekeerde het geval en geven verhalen aanleiding tot tellingen en produceren getallen nieuwe feiten en daarmee nieuwe werkelijkheden. Soms gaat dit onbewust en soms gaat dit bewust en worden tellingen ingezet om beleids- en besluitvorming te sturen. Voor het framen van beleids- en besluitvorming zijn verschillende stappen te onderscheiden: het in woorden benoemen van een bepaalde kwestie of probleem (naming), het aanbrengen van een ordening (classificeren) en het inbedden in een verhaal waar impliciet vaak al een richting van een oplossing in besloten ligt (narrating) (De Jong et al., 2018: 28). Deze stappen zijn evengoed van toepassing op cijfers. Cijfers kunnen een middel zijn om een kwestie te duiden en deze in een groter verhaal

te plaatsen (De Jong et al., 2018: 28). Met eindtoetsing in Nederland is een verhaal van tellen en vertellen met elkaar verknoot geraakt. Voor het opnieuw kunnen beschouwen van de rol van de eindtoets en de deugdelijkheid ervan in relatie tot doel en functie van onderwijs, de rol van onderwijs in het bieden van kansen en de maatschappelijke gevolgen van de huidige manieren van verdelen in termen van rechtvaardigheid, is het van belang de ontwikkeling van de huidige eindtoets in het licht van deze vraagstukken te reconstrueren.

Belangrijke vraag in dit onderzoek is hoe de eindtoets heeft kunnen uitgroeien van een test naar een beleidsinstrument en zich heeft weten te positioneren als instituut en als institutie. De eindtoets als instituut verwijst naar de betekenis van het woord 'instituut' in de betekenis van 'icoon' of 'legende' zoals in de context van 'het huwelijk als instituut'. Hiermee wordt aangegeven dat de eindtoets niet alleen staat voor een instrument, een test, maar voor een begrip dat weinig woorden nodig heeft om geduid te worden en gezien kan worden als iets van belang en statuut zoals blijkt uit het verhaal van Isik over 'De Cito'. De eindtoets is uitgroeid tot een 'instituut' in die zin dat de uitkomsten ervan een belangrijke waarde zijn gaan vertegenwoordigen in de perceptie van betrokkenen. Omdat de eindtoets in hoge mate bepalend is voor de verdeling van (onderwijs)kansen is de eindtoets ook een bepaler van maatschappelijke spelregels, wat maakt dat de eindtoets te beschouwen is als een institutie meer in de sociologische betekenis van het woord.

Gegeven de invalshoeken van de eindtoets als instituut en als institutie zijn de volgende deelvragen van belang in het kader van dit onderzoek. Ten aanzien van de rol van 'instituut' is het relevant te bevragen waar het vertrouwen dat aan de toets en de uitkomsten ervan wordt toegekend, vandaan komt en op gebaseerd is. Met daaraan gekoppeld de vraag of de eindtoets in de huidige vorm dit vertrouwen kan waarmaken. Deze vraag overlapt deels met het onderzoek naar de eindtoets als institutie. Voor een beleidsinstrument dat een dermate belangrijke positie heeft weten te verwerven als verdeler van onderwijskansen is het relevant te onderzoeken hoe met behulp van de eindtoets aan de verschillende functies inhoud en vorm wordt gegeven met daaraan gekoppeld de vraag of en in hoeverre de eindtoets die functies waar kan maken. En welke opvattingen over mens en maatschappij en in het verlengde hiervan over onderwijs hieraan ten grondslag liggen en hoe zich dat verhoudt tot rechtvaardigheid van verdelingen. Daartoe zal de eindtoets in de komende hoofdstukken worden onderzocht met verschillende invalshoeken als normatief uitgangspunt.

Een belangrijk deel van het onderzoek is uitgevoerd aan de hand van primaire bronnen zoals beleidsnota's, rapporten, toezichtkaders, kamerstukken, brieven en nota's. Uitgangspunt voor de selectie vormden de trefwoorden 'onderwijs/doelen van onderwijs/functies van onderwijs', 'selectie/toelating/allocatie/plaatsing', 'kwaliteit/rendement/opbrengsten', 'talent/intelligentie/verdienste/meritocratie', 'schaarste/verdelen/rechtvaardigheid' en 'toetsing/eindtoets/LVS'. Daarnaast is literatuur geraadpleegd, gerelateerd aan de onderwerpen zoals relevant gebleken voor een goed begrip van de tijd zelf en de te onderzoeken ontwikkelingen. In het kader daarvan hebben interviews plaatsgevonden met toetsexperts, onderwijshistorici, ex-ministers van onderwijs, beleidsambtenaren en onderzoekers.

Dit alles leidt tot een beeld over de ontwikkeling van de eindtoets, de maatschappelijke betekenis ervan, de beleidsachtergrond, het onderwijsbeleid, nationale en internationale wetenschapsont-

wikkelingen, ontwikkelingen in de testwetenschap in het algemeen en die voor de eindtoets in het bijzonder. Daarbij is gebleken dat een samenspel van ontwikkelingen zijn stempel heeft gezet op de rol en positie van de eindtoets nu. Deels bewust, deels onbewust, maar nooit het geheel overziend heeft de eindtoets een bepalende rol in de verdeling van onderwijsposities en onderwijskansen weten te verkrijgen. Om de normatieve vraag naar de deugdelijkheid van de eindtoets te kunnen beantwoorden, is de eindtoets gepositioneerd in relatie tot vraagstukken met betrekking tot doel en functie van onderwijs in het algemeen, rechtvaardigheid van verdelingen, kansengelijkheid en vraagstukken van verdelingen in de praktijk en psychometrische deugdelijkheid. Vanuit verschillende invalshoeken en bijpassende normatieve kaders is de eindtoets nader onderzocht.

Hoe heeft de eindtoets kunnen uitgroeien tot een instrument dat zo bepalend is geworden voor kansen van kinderen zoals blijkt uit het verhaal van Murat Isik. Tegen welke historische, maatschappelijke en economische achtergrond is de eindtoets tot stand gekomen? Wat heeft de eindtoets aantrekkelijk gemaakt om in te zetten als beleidsinstrument door de overheid? Is de eindtoets sterk genoeg om alle last die op de eindtoets drukt, te kunnen dragen? Wat zijn de gevolgen van de belangrijke rol die de eindtoets nu vervult voor de samenleving, het onderwijs en alle betrokkenen bij het onderwijs? Brengt de eindtoets beleidsambities die samenhangen met kansengelijkheid en kwaliteit van onderwijs dichterbij?

### 1.3 Beschrijving van de hoofdstukken

De eindtoets basisonderwijs is uitgegroeid tot veel meer dan een test. De opzet van dit proefschrift is zodanig gekozen dat aan de veelheid van verhalen, vertellingen en tellingen in de volle breedte zo veel mogelijk recht kan worden gedaan. Het geheel vormt een raamwerk van verklaringen waaruit de vormende kracht van de eindtoets als instituut en institutie naar voren komt. Omdat elke invalshoek eigen historische wortels heeft, bevat elk hoofdstuk een deel historie. Soms overlappen de historische beschrijvingen die met de thema's verbonden zijn elkaar voor een deel. In het belang van het goede historische begrip van ieder thema is ervoor gekozen die overlap te accepteren en niet te kiezen voor een chronologische reconstructie.

De eindtoets is meer dan een toets, gebruikt in en voor het onderwijs. De reikwijdte ervan gaat verder. In hoofdstuk 2 wordt een benadering van tests uiteengezet waarbij tests niet worden beschouwd als neutrale representanten van een werkelijkheid maar als vormers van de werkelijkheid. Tests zijn te zien als bijzondere vormen van beleidsinstrumenten die daarmee te beschouwen zijn als instituties en als zodanig onderzocht kunnen worden.

Een eerste, belangrijke verklaring voor de ontwikkeling en groei van de eindtoets houdt verband met de mogelijkheid die een eindtoets in combinatie met een LVS de overheid biedt voor het hebben van grip op en controle van een onderwijsstelsel dat grondwettelijk veel ruimte geeft aan scholen, ouders en leerlingen om eigen keuzes te maken. Dit roept de vraag op hoe de eindtoets die grip en controle realiseert, wat de waarde daarvan is en hoe dat te waarderen is. Hierbij gaat het om de validiteit van de eindtoets als test maar ook als beleidsinstrument. Dit staat centraal in hoofdstuk 3.

Een tweede verklaring voor het belang van de eindtoets wordt gevormd door de historische achtergrond waartegen de voorloper van de eindtoets tot stand komt en zich weet te positioneren

als een belangrijke speler bij het verdelen van onderwijsposities. Er is aanvankelijk vooral behoefte aan een instrument om de uitval uit het middelbaar onderwijs tegen te gaan. Dit verbindt zich met de behoefte aan een instrument waarmee geschiktheid voor de hogere vorm van voortgezet onderwijs bepaald kan worden omdat de druk op het benutten van alle talent groot is om mee te kunnen in de vaart der volkeren in de na-oorlogse periode. In hoofdstuk 4 staan historische ontwikkelingen ten aanzien van onderwijs en onderwijsbeleid centraal. Het vormt de achtergrond waartegen vormen van selectie met als uiteindelijk instrument de eindtoets, de voorloper en de ontwikkeling tot de huidige toets, gepositioneerd kunnen worden.

Een meer recente, derde verklarende factor wordt gevormd door de tweede belangrijke functie van de eindtoets, namelijk het inzichtelijk maken van de opbrengsten van onderwijs. Het denken in termen van rendement van onderwijs gaat al ver terug in de vorige eeuw. De ontwikkelingen in de rol van de overheid ten aanzien van het sturen van publieke organisaties in het algemeen en onderwijs in het bijzonder hebben ruimte gemaakt voor beleidsinstrumenten zoals tests als indicatoren van opbrengsten en daarmee van kwaliteit. De nadruk hierop heeft ertoe geleid dat de eindtoets een belangrijke rol heeft gekregen in toezicht op kwaliteit van onderwijs, zoals blijkt uit hoofdstuk 5. Dit heeft de positie van de eindtoets verder versterkt.

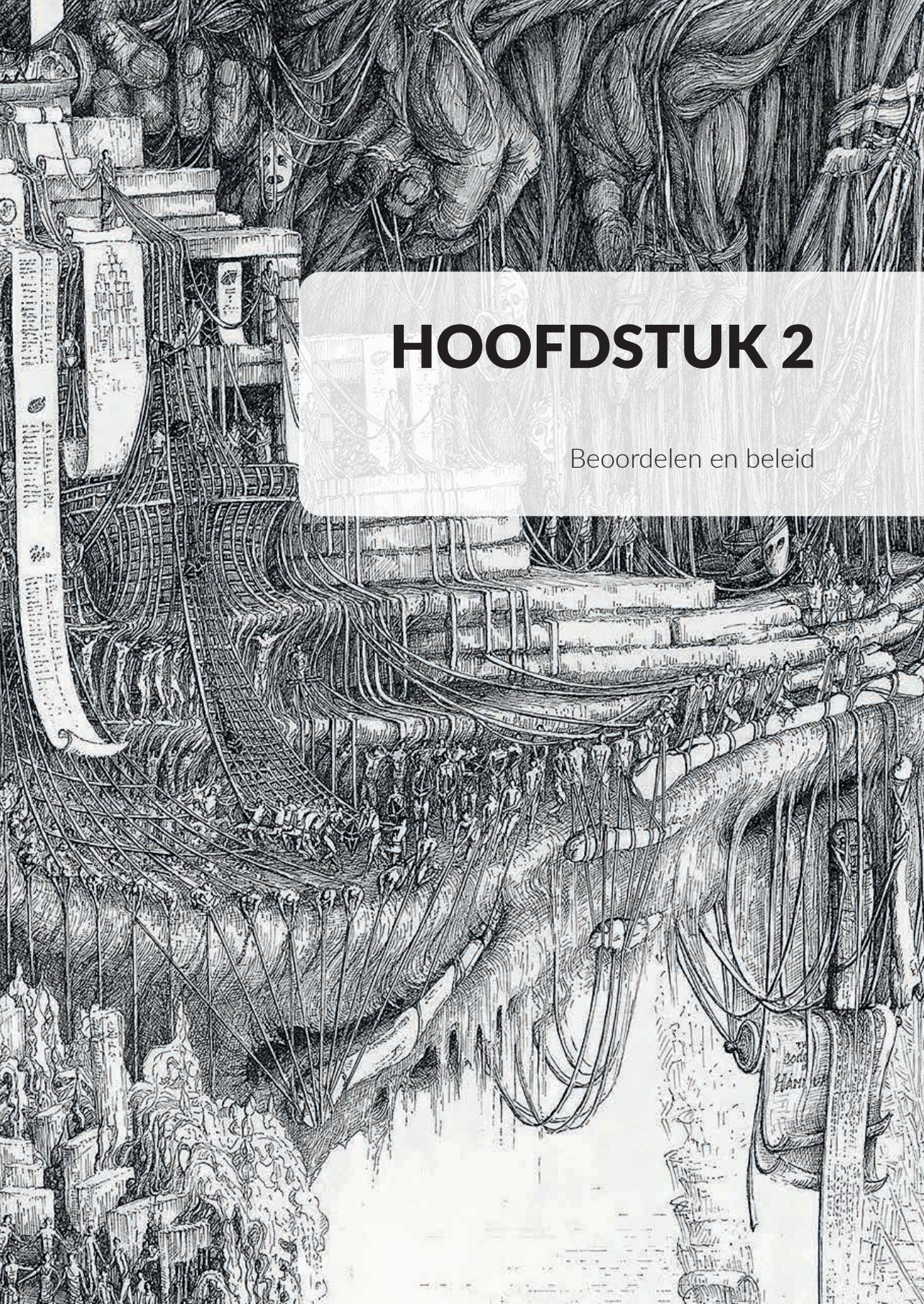
Een vierde verklaring is het feit dat zowel selectie als kwaliteit baat hebben bij een competitieve ordening. Verdienste, geoperationaliseerd als een mix van intelligentie en inzet, is een belangrijk uitgangspunt geworden van een samenleving die nadrukkelijk verkiest zich te ordenen op basis van het meritocratische gedachtegoed. Dit heeft gevolgen voor leerlingen en hun scholen. In hoofdstuk 6 wordt de inpassing van de eindtoets in deze meritocratische ordeningsprincipes voor samenleving en onderwijs geschetst. Dit heeft de positie van de eindtoets verder kunnen versterken.

Aan deze vierde verklaring is een vijfde nauw verbonden. Uitgaan van verdienste als ordeningsprincipe werpt een bepaald licht op maatschappelijke verdelingsvraagstukken en de rechtvaardigheid ervan. Schaarste is te beschouwen als een noodzaak voor meritocratische ordeningsprocessen die worden gezien als voorwaarde voor kansgelijkheid. De rol van de eindtoets in het licht van rechtvaardigheid van verdelingen en meer op praktisch niveau als instrument voor de praktische operationalisatie van verdelingen van schaarste staat centraal in hoofdstuk 7.

Ontwikkelingen in de wetenschap hebben de eindtoets in de lift gezet. Enerzijds heeft de testwetenschap zichzelf binnen de menswetenschappen, maar ook breder binnen de wetenschappen een serieuze positie weten te verwerven met een empirisch analytische aanpak en anderzijds is de eindtoets (samen met andere tests) gebruikt als instrument in tal van onderzoek in de menswetenschappen. De vervlechting van testen als wetenschap met de rol van tests in de wetenschap heeft de waarde van tests als de eindtoets en met name het gebruik van de uitkomsten ervan in onderzoek versterkt en vormt een zesde verklaring zoals blijkt uit hoofdstuk 8.

In het laatste hoofdstuk, hoofdstuk 9, worden de belangrijkste elementen uit de eerste hoofdstukken samengebracht. Het geheel zorgt voor een samenhangend beeld van de werkelijkheid waarin de eindtoets fungeert en die de eindtoets creëert. De thematisch samenvattende beschrijvingen van voorgaande hoofdstukken beantwoorden de vraag naar de deugdelijkheid van de eindtoets.





# HOOFDSTUK 2

Beoordelen en beleid





## Hoofdstuk 2 Beoordelen en beleid

*'Ik denk altijd: zo'n schaap van tien jaar, wat kan je d'r nou helemaal met zekerheid van zeggen; en als ik later in de hoogste klas een flinke jongen of een flinke meid heb die ik indertijd niet op een cursus heb gedaan omdat het toen nog niet zo absoluut zeker voor me was dat-ie flink zou worden – zie je, dan kan ik mezelf wel voor m'n kop slaan van spijt. Ik wil ze allemaal d'r lui kans geven en mijn stelling is: geef ze allemáál op voor 'n vreemde taal, behalve die waarvan je absoluut zeker weet dat het tóch hopeloos is. Nou, en hoeveel zijn er nou absoluut hopeloos? [...] Het is een ellendig wanhopige school, want ik ben gedwongen de kinderen al te vonnissen als ze tien jaar oud zijn, door hele gebieden van het leven voor ze af te sluiten, zonder genade.'*

*Uit: Theo Thijssen, 'De Gelukkige klas' (1926)*

Theo Thijssen laat zien dat oordelen, ordenen en groeperen om kinderen wel of geen kansen te bieden op doorstroom naar de hoogste niveaus van voortgezet onderwijs honderd jaar geleden al een dilemma was waar een onderwijzer in het toenmalige lager onderwijs mee te maken had. De rol van onderwijs in het ordenen, classificeren en toewijzen van onderwijsloopbanen, het afgeven van diploma's en in het verlengde daarvan ook in het bieden van kansen op de maatschappelijke ladder is in de afgelopen eeuw steeds belangrijker geworden. De geschiedenis van het beoordelen reconstruerend stelt Wilbrink dat in onderwijs altijd al gebruik is gemaakt van 'externe prikkels zoals beloning en straf, met naar plaats en tijd verschillende accenten daarin, en in de tijd een ontwikkeling van lijfstraffen naar stelsels van abstracte beloning en straf in de vorm van punten, en later van cijfers om daarmee leerlingen te motiveren om hun best te doen' (Wilbrink, 1996). Het onderwijs heeft zich ontwikkeld van een prijzenstelsel naar een puntenstelsel dat is vervangen door een cijferstelsel. De cijferschalen die in de vroegere vormen van getuigschriften en schoolrapporten voorkomen, verschillen aanvankelijk: van 1-5, van 1-10 en van 1-12. In de loop van de vorige eeuw worden cijferschalen gelijkgeschakeld en wordt de cijferschaal 1-10 de standaard (Dane, 2014). De standaardisering van cijferstelsels is een belangrijke voorwaarde geweest om vergelijkingen tussen leerlingen en tussen scholen mogelijk te maken en zo resultaten van onderwijs in een context van politiek en beleid te kunnen plaatsen. Aanvankelijk ontstaat de druk op het gebruiken van prestaties in het onderwijs zelf. Bij de invoering van de HBS achtte minister Thorbecke toelatingsexamens onder alle omstandigheden 'verwerpelijk' (Bartels, 1963: 94). Maar onder druk van inspecteurs en van leraren die 'vreesden voor verlies aan orde, tucht en vlijt' stelde hij zijn beleid bij (Bartels, 1963: 96). Het beoordelen van leerlingen ofwel relatief door ze te vergelijken met elkaar, ofwel absoluut door ze te vergelijken met een inhoudelijk criterium, al dan niet uitgedrukt op een cijferschaal, is een voor onderwijs geaccepteerde manier van kijken naar leerlingen. Dit geldt ook voor beoordelen om te selecteren. De toegenomen, externe druk, nationaal en na de Tweede Wereldoorlog ook internationaal, op onderwijs om met behulp van tests te oordelen, te ordenen en te selecteren is niet op veel weerstand gestuit in het onderwijs. Het bleek goed te passen bij de vertrouwde, bestaande beoordelingscultuur en heeft daarmee weinig legitimering nodig gehad. Onderwijs heeft zich ontwikkeld van een competitie bedoeld om leerlingen te motiveren om te (willen) leren naar competitie gericht op selectie. Deze ontwikkeling is bepalend geworden voor de manier van studeren, en via de voorgeschreven inhoud van de examens ook voor de inhoud van het onderwijs (Wilbrink, 1996). Meten heeft effecten en is vanwege die effecten, beoogd of niet beoogd, niet neutraal of onschuldig. Het bepaalt het

handelen van hen die de meting (moeten) ondergaan, van hen die de meting opleggen en van hen die de uitkomsten van metingen gaan gebruiken.

In dit hoofdstuk staat beoordelen in relatie tot ordenen en classificeren voor beleidstoepassingen centraal. In paragraaf 2.1 wordt uitgebreid ingegaan op functies die ordeningen en classificaties meer in het algemeen hebben voor samenlevingen. Tests spelen een belangrijke rol bij het beslissen wat of wie wel of niet tot een categorie behoort. Uitkomsten van tests worden gebruikt voor tal van beleidstoepassingen. Daarmee zijn tests te beschouwen als instituties in sociologische zin die als zodanig onderzocht kunnen worden op hun rol en waarde (2.2). Dit geldt ook voor de eindtoets.

## 2.1 Ordenen en classificeren

Ordeningen en classificaties zijn nauw verbonden met het vormgeven van samenlevingen. Ze zorgen voor structuur en overzicht die nodig zijn om een samenleving te kunnen inrichten en richten. Want tal van aspecten die horen bij georganiseerde samenlevingen vereisen duidelijkheid over 'wie wel en wie niet', zoals voor het maken van wetten, het opleggen van straffen, het toekennen van vergoedingen of het betalen van belastingen. Hóe in een staat grenzen worden bepaald ten behoeve van het classificeren, is minder vanzelfsprekend. Die keuzes hangen samen met de richting waarin de staat en de samenleving (willen) bewegen. Iedere standaard en elke categorie drukt een waardering uit voor een bepaald standpunt en leggen een ander gezichtspunt het zwijgen op. Daar is volgens Bowker en Star op zich niks verkeerd aan, het is volgens hen onvermijdelijk. Maar tegelijkertijd is het wel van belang dit te zien als een ethische en ook politieke keuze die als zodanig gevaarlijk is: 'Not bad, but dangerous' (Bowker & Star, 2000: 5). Een classificatie definiëren Bowker en Star als volgt: 'A classification is a spatial, temporal, or spatio-temporal segmentation of the world.' (Bowker & Star, 2000: 10). Een classificatie-systeem beschouwen zij letterlijk en figuurlijk als een serie dozen waarin zaken gestopt kunnen worden. Een classificatie-systeem heeft de volgende eigenschappen:

1. Er zijn consistente, unieke classificatie-principes die gehanteerd worden,
2. de gehanteerde categorieën sluiten elkaar uit, en
3. het systeem is compleet.

Volgens Bowker en Star voldoet geen enkel classificatie-systeem in de werkelijkheid aan deze drie, 'eenvoudige' principes en zal ook geen enkel classificatie-systeem er ooit aan kunnen voldoen (Bowker & Star, 2000: 11). Het is altijd nodig om keuzes te maken omdat indelingen zich niet altijd vanzelf uitwijzen. Zij wijzen erop dat het van groot belang is dat classificatie-systemen gezien en gelezen worden als politieke en culturele producties (Bowker & Star, 2000: 324). Bowker en Star maken onderscheid tussen een classificatie-systeem en een standaard in die zin dat classificaties wel of niet gestandaardiseerd kunnen worden maar dat standaarden altijd een onderliggend classificatie-systeem kennen (Bowker & Star, 2000: 15). Daarmee zijn ook standaarden te herleiden naar classificatie-systemen die imperfecties kennen en ooit als politieke en culturele producties tot stand zijn gekomen. Ondanks deze beperkingen is in moderne samenlevingen sprake van een grote toename van classificaties en daarop gebaseerde standaarden. Omdat overheden regie en controle willen hebben op de inrichting van samenlevingen zorgen zij voor een toename van de rol van gestandaardiseerd meten met objectief uit te voeren

beoordelingen, zodat op basis daarvan 'objectief' geordend kan worden. Dit leidt tot een groei van gestandaardiseerde metingen vooral bedoeld om te ordenen en vervolgens te kunnen classificeren en om op basis van de classificatie weer te kunnen selecteren. Objectieve vormen van beoordelen krijgen daarbij de overhand en professionele vormen van beoordeling – minder van buitenaf en op afstand beheersbaar – schuiven naar de achtergrond. Dit proces om te komen tot gestandaardiseerde feiten verloopt volgens Scott in drie stappen (Scott, 1998: 80):

1. De creatie van gezamenlijke meeteenheden,
2. de categorisatie volgens de nieuwe meeteenheden, en
3. dan vervolgens het tellen, het rekenen en het rapporteren in de aldus tot stand gekomen eenheden.

Dit leidt volgens Scott onvermijdelijk tot simplificatie omdat repliceerbaarheid een belangrijk doel wordt van metingen. Dat brengt met zich mee dat bewust uitzonderingen of eigenaardigheden genegeerd moeten worden om het schematische ordenen mogelijk te maken. Bepaalde vormen van kennis en controle vereisen een versmalling van de blik. Het grote voordeel van zo'n 'tunnelvisie' is volgens Scott dat het een beperkte set aspecten scherp in beeld brengt van een anders veel te complexe en veelomvattende werkelijkheid. De simplificatie, op haar beurt, maakt het fenomeen in het centrum van de focus meer legitiem en meer vatbaar voor zorgvuldig meten en berekenen. Gecombineerd met vergelijkbare observaties ontstaat volgens Scott een overall, geaggregeerd, synoptisch beeld van een selectieve werkelijkheid, die een hoge mate van schematische kennis, controle en manipulatie mogelijk maakt (Scott, 1998: 81). Het gebrek aan context en uitzonderingen of particulariteiten is niet iets dat per ongeluk over het hoofd wordt gezien, het is de noodzakelijke voorwaarde van iedere op grote schaal uit te voeren actie. Hier hoort bij dat er heel weinig vertrouwen is in de skills, intelligentie en ervaring van gewone mensen in de alledaagse praktijk, die Scott aanduidt als 'metis' en die hij stelt tegenover 'techne', wat voor hem staat voor de toepassing van algemene regels (Scott, 1998: 346). Scott beargumenteert dat de systematische en onpersoonlijke regels van 'techne' kennisproductie faciliteren omdat ze snel verzameld en begrijpelijk gedocumenteerd kunnen worden maar dat ze beperkt zijn in het kunnen bieden van verklaringen zoals 'metis' die wel kan bieden (Scott, 1998: 320).

De Amerikaanse politieke wetenschapper Stone stelt dat 'policy arguments' voor het classificeren en differentiëren overtuigend zijn in de mate waarin zij bevredigend rekenschap geven over de rechtmatigheid om gevallen gelijk of juist verschillend te behandelen. Politieke redeneringen, daarom, gaan volgens Stone dan ook primair over het beredeneren van gelijkheid of verschil, over goed en kwaad of over juist en onjuist (Stone, 2012: 382). Het maken van keuzes is onlosmakelijk verbonden met het proces van classificeren omdat categorieën van elkaar te onderscheiden moeten zijn om, al dan niet met behulp van tests, te kunnen indelen of toewijzen. Om te kunnen verdelen, zijn categorieën nodig en om categorieën te kunnen gebruiken, moet er geteld kunnen worden. En tellen kan alleen als duidelijk is wat de categorieën zijn waarvoor geteld wordt: er moet immers steeds weer besloten worden of iets wel of niet tot een bepaalde categorie behoort. Dit proces van classificeren is daarmee per definitie een menselijk proces. Daartoe worden kenmerken beschreven die wel horen bij de ene categorie en niet bij de andere. Na het bepalen van de categorieën en de uit te voeren classificaties, moeten de te classificeren objecten gewaardeerd worden: horen ze wel of niet bij een bepaalde categorie? Daarbij gaat het om de vraag of

iets inclusief of exclusief is op basis van welke karakteristieken. Zo wordt gekomen tot criteria, tot toewijzingen op basis van standaarden. Het maken van fouten hoort bij classificeren, fouten die op de koop toe worden genomen, of, als ze als te onrechtvaardig worden ervaren, reden zijn om een classificatie aan te passen. Begrip van classificaties in het sociale domein vraagt om analyses die zowel de historische context, collectieve actie en politieke keuzes daarin meenemen. Want voor classificaties in het sociale domein zijn tal van keuzes, wellicht een oneindige hoeveelheid keuzes, mogelijk. Voor begrip van bestaande classificaties is begrip van de historische context waarbinnen classificaties tot stand komen van belang. Getallen in het politieke debat kunnen volgens Stone niet begrepen worden zonder te onderzoeken hoe mensen ze maken: wat maakt dat mensen beslissen iets te tellen? Hoe zijn degenen die meten en die gemeten worden aan elkaar verbonden? Welke prikkels hebben mensen om vervolgens waarderingen te koppelen aan wat er gemeten wordt? Dit is volgens Stone bij uitstek een politieke zaak, een zaak van 'de polis' zoals zij dat noemt (Stone, 2012: 203). 'Every number is a political claim about "where to draw the line".' (Stone, 2012: 188). Dit samenspel van keuzes voor te hanteren maten, het uitvoeren van metingen en gemeten worden, maakt het gebruik van getallen en metingen volgens Stone tot een geducht wapen in de politieke arena. De keuzes die met tellen samenhangen, zijn niet zonder gevolgen. Getallen beschrijven de wereld met dezelfde vrijheid en interpretatiemogelijkheden die ook schilders en dichters hanteren, in de optiek van Stone (Stone, 2012:205).

### 2.1.1 Tellen

Tellen is de basis van het proces van het steeds weer nemen van besluiten over wat wel en niet meegeteld kan worden met het risico dat iets ten onrechte buitengesloten wordt of ten onrechte meegeteld wordt. Tellen is ook te bezien als een proces van verlies, verlies van informatie omdat er nu eenmaal keuzes gemaakt moeten worden in wat wel en niet mee kan doen in een categorie. Dat kan niet uitputtend. Van eenmaal geconstrueerde categorisaties en ordeningen gaan grote gevolgen uit als op basis daarvan eenmaal geteld gaat worden. Met tellingen worden volgens de Duitse socioloog Mau fenomenen, karakteristieken, eigenschappen gepresenteerd in algemene, abstracte en universele taal, de taal van de getallen: de wiskunde. Getallen in de wiskunde hebben het imago exact en hard te zijn in die zin dat het 0 of 1 is. Getallen kunnen worden opgeteld, afgetrokken, gedeeld en gemiddeld. Zo ontstaan standaardiseringen die hun eigen leven gaan leiden en een beeld creëren van duidelijk te onderscheiden verschillen, zelfs daar waar de verschillen soms maar klein zijn. Dergelijke vormen van standaardiseringen ziet Mau als de natuurlijke vijand van diversiteit (Mau, 2017:227). Het is echter wel de manier waarop vakgebieden kunnen worden ontsloten en toegankelijk gemaakt voor niet-professionals. Want in de overzichtelijke wereld van getallen heeft iedereen begrip van 'meer en minder' en 'veel en weinig'. Deze begrippen roepen ook een gevoel van objectiviteit op. En gevoelens gaan van objectief vaststellen van veel of weinig, hoog of laag gemakkelijk over naar waardeoordelen in termen van goed en fout, dom en slim. Gevoelens en oordelen van kwantitatieve aard verworden zo heel makkelijk en schijnbaar natuurlijk tot kwalitatieve oordelen, en vice versa. Zijn er eenmaal cijfers dan gaat daar vaak zo'n stelligheid van uit dat het nog maar moeilijk te achterhalen is welke keuzes daarachter schuil gaan, laat staan die ter discussie kunnen stellen. Het feit dat keuzes niet meer zichtbaar zijn, verandert niets aan het feit dat cijfers niet gewoon objectieve, ruwe data zijn maar nagenoeg altijd het resultaat van waardegeladen keuzes zijn (De Jong et al., 2018: 48).

Wilson stelt dat een standaard tegelijkertijd compleet arbitrair en tegelijkertijd compleet accuraat

kan zijn. Een standaard is arbitrair omdat deze gebaseerd is op een mening en een besluit wat er één is van de vele die ook gekozen hadden kunnen worden. En het kan compleet accuraat zijn omdat wat je kiest als een standaard een waarde heeft die je exact kunt vaststellen, want het is een kwestie van een definitie (Wilson, 1998: 9.3). Maar daarmee is een standaard niet iets dat aan een meting ontleend is. En het vormt zeker geen meting op zichzelf. Het is een keuze en kan in die hoedanigheid de basis vormen voor meten. Want als er eenmaal een standaard is gekozen, is het continuüm opgeheven en ontstaan categorieën. En bij de categorieën horen grenzen: dit behoort er (nog) wel toe en dat niet (meer). Het is van belang te blijven zien dat de oplossing die getallen bieden niets meer of minder is dan een menselijk besluit over hoe en wat te 'tellen als'. Stone stelt dat er in die zin een analogie is tussen getallen en metaforen: 'Numbers, in fact, work exactly like metaphors. To categorize in counting or to analogize in metaphors is to select one feature of something, assert a likeness on the basis of that feature, and ignore all the other features.' (Stone, 2012: 185). En zo kunnen vervolgens ook hiërarchieën ontstaan. Want wat aan de meeste criteria voldoet wordt al snel gezien als 'het best passend' in een categorie, beter dan hetgeen dat maar aan een enkel criterium voldoet. Daarmee zijn classificaties niet statisch maar ontwikkelen ze zich in de praktijk van de uitvoering en ontstaat er een patroon van normen over de hoogste vorm van 'passendheid' en de laagste vorm van 'passendheid'.

Aan de hand van instrumenten die ordenen en classificeren kan worden gereconstrueerd wat de normen en waarden zijn die volgens een instrument relevant worden geacht voor het verdelingsvraagstuk waarvoor het wordt ingezet. Daarmee geven instrumenten een beeld van de, al dan niet bewust, beoogde inrichting van de samenleving. Daar waar tests worden gebruikt om (schaarse) goederen te verdelen kan aan een test worden afgelezen wat wordt geteld, welke categorieën worden gehanteerd, wat er nog wel en niet meer in valt en welke hiërarchieën hierin feitelijk worden aangebracht. Hoe krachtig de waarde van de uitkomst van een meting is, hangt samen met de macht van de instantie die de meting oplegt aan en uitvoert bij een individu. Hoe bepalender de rol van die instantie in de samenleving, hoe meer impact de uitkomst heeft op het individu maar ook op de samenleving als geheel. De macht die in handen is van degenen die de indicatoren, data en metingen kunnen gebruiken, kan volgens Mau zover reiken dat ze daarmee maatschappelijke domeinen kunnen herstructureren en nieuwe handelingslogica's kunnen doorvoeren (Mau, 2017: 185). Mau stelt dat aan de hand van beoordelingen zichtbaar wordt aan welke activiteiten, prestaties of eigenschappen een hogere of juist een lagere waarde blijkt te worden toegekend. Data leggen bloot hoe dingen gezien moeten worden en sluiten daarmee andere zienswijzen uit. En zo komt er een nauwe verbinding tot stand tussen het schatten van waarden in het proces van kwantificering en waarde-inschatting in de zin van sociale erkenning (Mau, 2017: 31). Er worden niet alleen getallen in stelling gebracht maar er wordt vooral duidelijk welke normatieve principes relevant zijn. En die hebben gevolgen voor zelferkenning en eigenwaarde. Mau spreekt in dat verband van de ontwikkeling van 'das soziale Wir' dat zich meer en meer richt op dat wat gemeten wordt, belangrijk is en op die manier overgaat in 'das metrische Wir' (Mau, 2017:24). Mau wijst op het feit dat 'statistiek' oorspronkelijk de betekenis had van 'staatswetenschap of staatskunde'. Dat de staat zich de taal van de getallen toe-eigende, opende voor de politiek een nieuwe weg om het sociale leven te beschrijven, te ontsleutelen en te ordenen (Mau, 2017: 35).

Simplificeringen en abstracties zijn onlosmakelijk met classificeren verbonden. Het effect is dan

dat getallen niet alleen dienen om de samenleving te beschrijven of, een stap verder, te reguleren maar dat getallen een eigen leven kunnen gaan leiden en daarmee een samenleving richten en ordenen omdat getallen kunnen gaan fungeren als plaatsbepalers in het sociale domein. Bowker en Star laten zien dat classificaties gezien moeten worden als belangrijke elementen van politiek en ethisch handelen die het verdienen zelf opnieuw geïnclassificeerd te worden. Want, stellen zij, we weten nauwelijks meer wat we gebouwd hebben. Classificaties zijn machtige technologieën geworden die ingebed zijn in infrastructuren waarmee ze relatief onzichtbaar zijn geworden zonder aan kracht te verliezen. En niemand is 'in control' van die infrastructuur en niemand heeft de macht dit centraal te veranderen. Alledaagse categorieën zijn precies die zaken die verdwenen zijn in infrastructuur, in gewoontes, in dat wat we voor waar zijn gaan aannemen (Bowker & Star, 2000: 319).

### 2.1.2 Vergelijken in het sociale domein

Getallen in de vorm van scores lenen zich eenvoudig voor het maken van ordeningen, zodanig dat er een volgorde ontstaat, veelal van hoog naar laag scorend. Dergelijke ordeningen maken van scores een rangvolgorde. Hiermee krijgt alles of iedereen een unieke plek in een gekozen ordening. Het gebruiken van rangordeningen of rankings is in het algemeen niet zonder effecten. Rankings hebben volgens Mau de functie van sociale plaatsaanwijzers. Door eerst scores en vervolgens de individuen die de scores hebben behaald, in volgorde te zetten, worden individuen met elkaar vergeleken. Rankings wijzen niet alleen aan op welke rij of plaats iemand staat of zit, maar bepalen tegelijkertijd op basis van welke criteria en processen de plekken moeten worden bezet (Mau, 2017: 73). De toename van belangen die met classificaties en de beoordelingen die daarbij horen, maakt dat er volgens Porter meer en meer behoefte ontstaat aan onpersoonlijke strategieën voor de uitvoering ervan. De waar te nemen overgang van expert-oordelen naar expliciete besliscriteria plaatst Porter meer in het licht van de druk van buiten die op beslissers wordt gelegd. Dit wakkert de behoefte aan om op meer onpersoonlijke strategieën terug te vallen als het gaat om het nemen van belangrijke beslissingen. Porter ziet in het terugvallen op kwantitatieve regels zwakte en kwetsbaarheid. Vooral beslissingen die genomen moeten worden door beslissers die snel het verwijt kunnen krijgen van het nemen van arbitraire beslissingen en partijdigheid omdat ze autoriteit ontberen, zijn geneigd terug te vallen op besluiten genomen op basis van getallen. Dit geeft tenminste een gevoel van eerlijk en belangeloos beslissen. Dit is volgens Porter een reactie op het wantrouwen dat heerst bij de afwezigheid van een veilige en autonome gemeenschap (Porter, 1995: xi). Wetenschappelijke objectiviteit is volgens Porter dan vooral een antwoord op een morele vraag naar onpartijdigheid en fairheid: 'Quantification is a way of making decisions without seeming to decide.' (Porter, 1995: 8). Want het ordenen en plaats aanwijzen dat met vergelijken gepaard gaat, heeft effect. Het zet wat of diegene die geordend wordt in een concurrentiemodus. Het vergelijken brengt met zich mee dat de eigen positie alleen maar waarde heeft in vergelijking met de posities van anderen. Succes is in een dergelijk systeem altijd relatief en nooit absoluut. En een systeem van voortdurende vergelijkingen brengt met zich mee dat stilstaan of even achterover hangen geen optie is. Het vergelijken functioneert volgens Mau echt goed als het vergelijken met affectieve referenties vergezeld gaat en daadwerkelijk terugslaat op het gevoel van eigenwaarde. En helemaal goed lukt het vervolgens als die vergelijkingen door derden worden gemaakt die men zelf voor relevant houdt, zijnde de politiek, peers of interessante openbare instituties (Mau, 2017: 63/64). Vergelijkingen zijn gericht op het laten zien van verschillen in plaats van gemeenschappelijkheden, op het creëren van hiërarchie

in plaats van gelijkheid. Daarmee is vergelijken niet neutraal in de uitwerking. Het zet iedereen die aan de vergelijking onderworpen wordt er voortdurend toe aan om niet terug te vallen, maar aan de bal te blijven, beter te worden, het beste, beter dan onze concurrenten. Het sociale vergelijken creëert een instabiele samenhang met anderen, omdat de eigen waarde afgemeten wordt aan wat anderen doen, hebben of zijn (Mau: 2017: 68). Een test is niet gericht op neutrale toewijzing van een score maar de score heeft een waarde die gezien kan worden als een waardering die vervolgens kan leiden tot een erkenning: waardetoekenning, waardering en erkenning hangen volgens Mau dan ook nauw met elkaar samen (Mau, 2017: 68). Soms gaat het zover dat tests creëren wat ze meten en is de werking feitelijk omgekeerd: door tests te gebruiken kan worden aangetoond dat het construct dat gemeten wordt daadwerkelijk bestaat. Daarmee stellen tests niet alleen een construct vast, ze creëren en legitimeren het daadwerkelijke bestaan van het construct. Want dat wat gemeten kan worden, moet wel bestaan. Volgens Stone is dit een van de verborgen boodschappen die schuilgaan achter manieren van tellen en publiceren van getallen: door iets te tellen laat je zien dat er sprake is van een te identificeren eenheid met heldere grenzen: door een telling te geven laat je het publiek geloven dat iets telbaar is (Stone, 2012:191). Het meten van intelligentie is daarvan een belangrijk voorbeeld. Het construct 'intelligentie' kent geen definitie. Dit heeft er niet voor gezorgd dat gestopt is met het meten van intelligentie. Er zijn bijzonder effectieve strategieën te onderscheiden die maken dat 'intelligentie' nog altijd met veel overtuiging wordt gemeten. Belangrijk voor de overtuigingskracht is dat wat gemeten wordt wel degelijk ervaren wordt als één te meten construct. Bijvoorbeeld door het consequent te benoemen als één ding, 'intelligentie', wordt het acceptabel het uit te drukken in één getal en omgekeerd, als je het in één getal kunt uitdrukken, is het één ding. Door daar een schaal aan toe te voegen gaat er van een intelligentiemeting de suggestie uit dat je er meer of minder van kunt hebben, dat het varieert in hoeveelheid. En wat je van die hoeveelheid hebt, wordt gepresenteerd als een vaststaand en dus te meten gegeven. Zo kan er, zonder dat er een heldere definitie is van een construct, wel degelijk sprake zijn van een bijzonder krachtig geloof in het construct én de meting ervan. Dit maakt dat Hanson over IQ spreekt als over 'kunstmatige' intelligentie (Hanson, 1993: 282). Voor Mau is 'intelligentie' een belangrijk voorbeeld van hoe succesvol een politiek op basis van indicatoren kan zijn. Op het moment dat er in de openbare waarneming een gelijkschakeling gerealiseerd wordt van een theoretisch construct en een indicator op basis waarvan wordt gemeten, is een politiek op basis van gemeten indicatoren zeer krachtig (Mau, 2017: 29).

Bij het vergelijken van scores en daarmee van individuen met elkaar is de gemiddelde score een belangrijk gegeven. Letterlijk heeft het wiskundige concept 'gemiddeld' in het dagelijks taalgebruik een meer specifieke invulling gekregen, zelfs conflicterende betekenissen: aan de ene kant kan gemiddeld de betekenis hebben van 'normaal, voldoende aan de norm en daarmee acceptabel' maar het kan ook betekenen 'middelmatic' waarbij de associatie niet ver weg is dat er ruimte voor verbetering is. Een score 'onder het gemiddelde' wordt dan al gauw een score die 'onvoldoende' is. De interpretatie van het begrip 'gemiddeld' is van groot belang voor de interpretatie van uitkomsten en de effecten ervan op betrokkenen. Gestandaardiseerde vormen van meten, zeker daar waar uitgegaan wordt van vergelijkingen die draaien rond het concept van het gemiddelde, zijn zelden vrijblijvend in hun uitkomsten en effecten. Scores veranderen hoe anderen over een individu denken en hoe het individu denkt over zichzelf: 'Tests act to transform, mold, and even to create what they supposedly measure.' (Hanson, 1993: 52).

Metingen zijn rechtvaardigingen van hoe de zaken te zien en te waarderen zijn. Tellen en vertellen zijn volgens De Jong niet twee gescheiden strategieën maar deze raken elkaar, beïnvloeden elkaar en zijn in een continu samenspel onlosmakelijk met elkaar verbonden (De Jong et al., 2018: 6). Wat aan kwantificeren en classificeren met behulp van metingen voorafgaat, is een palet van keuzes en beslissingen, bewuste en soms ook onbewuste. Van bestaande metingen is niet altijd meer duidelijk wat ooit de keuzes zijn geweest die aan een test vooraf zijn gegaan. In het geval van gestandaardiseerd meten, zeker in de vorm van door de overheid voorgeschreven tests, vallen die keuzes, beslissingen en gevolgen in het politieke domein omdat het hierbij gaat om tests die zijn gebaseerd op waarde-ordeningen die vertaald zijn in beoordelingsmaatstaven die gevolgen hebben voor de individuen die aan tests onderworpen worden. Als degenen die metingen vormgeven en afnemen dezelfde zijn, dan hebben ze veel opties ter beschikking om te bepalen wat ze op welke manier meten. Ze kunnen metingen zo vormgeven dat de uitkomsten passen bij de doelen die ze politiek gezien nastreven of bij de mogelijkheden die daarbij voor hen open staan: ‘Measures, in short, obey the law of the hammer: if all you have is a hammer; every problem looks like a nail.’ (Stone, 2012:204).

### 2.1.3 Onderwijs en classificaties

Ontwikkeling van de natiestaat en de ontwikkeling van nationale onderwijssystemen zijn nauw met elkaar verbonden om te komen tot bredere sociale, culturele en economische ordening, die in elk land zijn specifieke kenmerken heeft. De Vijlder laat zien dat dit niet altijd met de voor de burgers beste bedoelingen is gebeurd: ‘Hoewel er vaak de meest edele doelstellingen aan de overheidsbemoediging zijn toegeschreven, zoals het bevorderen van sociale gelijkheid en sociale cohesie, is de voornaamste rol van scholen en universiteiten voor moderne natiestaten de bevestiging en instandhouding van hun soevereine karakter geweest. Dankzij deze instituties zijn nationale overheden in staat geweest zeggenschap uit te oefenen over de bevolking, de leden van de nationale samenleving een gevoel van veiligheid en groeiende welvaart te bieden en aldus hun gezag te legitimeren’ (De Vijlder, 1996: 4). Socioloog Weber wijst op het feit dat onderwijs enerzijds de kans biedt om op te klimmen op de maatschappelijke ladder, de emanciperende functie van onderwijs, en anderzijds gebruikt de heersende macht onderwijs als vorm van bestendiging van het bestaande machtsevenwicht waarmee onderwijs een reproductiefunctie vervult. Diploma’s worden de nieuwe eenheid waarmee posities in de samenleving worden verdeeld. En waarmee ze worden gelegitimeerd. Weber spreekt in dit verband van ‘social disclosure’: diploma’s geven exclusiviteit bij toegang tot bepaalde groepen of goederen of bij het behoud ervan. Diploma’s fungeren daarmee niet alleen als middel om macht of status te behouden (reproductie), maar ook als middel om macht of status te verkrijgen (emancipatie) (Elffers, 2018: 180). Daarmee wordt onderwijs van belang voor hen die aan hun achterstandspositie willen ontkomen maar wordt het tegelijkertijd belangrijker voor bestaande dominanties van macht, zoals onderwijs om de bestaande balans te bewaken. Per saldo is onderwijs vooral te zien als een bestendiger van een bestaande samenleving, meer dan als een verbeteraar van levensperspectieven (Eidhof et al., 2016: 66). In Nederland, net als in Duitsland en Frankrijk, heeft de staat het monopolie verworven over de afgifte van diploma’s en de regulering van de toegang tot het vervolgonderwijs. Daarmee hebben overheden in feite de macht in handen om de maatschappelijke posities in de samenleving letterlijk zelf actief te verstrekken (De Vijlder, 1996: 110). Foucault geeft inzicht in de gevolgen die daarmee gepaard kunnen gaan. Een examen is volgens Foucault een manier om macht over mensen uit te oefenen en hen te disciplineren. Het examen



is voor Foucault een 'normatieve blik, een vorm van toezicht die kwalificeert, klasseert en straft' (Foucault, 1989: 256). Volgens Foucault is examinering (in de breedste zin van het woord op te vatten) een techniek waarmee macht uitgeoefend kan worden over hen die aan de examinering worden onderworpen. En die macht blijft niet zonder effect, volgens Foucault: 'Door het examen worden de individuen zichtbaar gemaakt, zodat ze gedifferentieerd en gestraft kunnen worden. Daarom is het [examen] in alle disciplinerende voorzieningen zo sterk geritualiseerd. Het verenigt in zich de ceremonie van de macht en de vorm van het experiment, het vertoon van kracht en het vaststellen van de waarheid. In het hart van de disciplinerende procedures vormt het examen de subjectiverende onderwerping van degenen die worden beschouwd als objecten en de objectivering van de onderworpen subjecten.' (Foucault, 1989: 257). Daarmee heeft zich middels examinering een relatie van bewaking, definiëring en regulering genesteld in het hart van het onderwijs. Daar worden oordelen geveld die disciplinerend werken: examinering is te beschouwen als een techniek waardoor een moderne persoon tegelijkertijd geconstrueerd en gecontroleerd wordt. Schaarste is dan juist nodig om die disciplinerende procedures optimaal te laten werken. Beoordelingen gekoppeld aan examinering verminderen in de optiek van Foucault namelijk niet alleen afwijkend gedrag van een persoon maar zij produceren ook de persoon. Wat een kind (of volwassene) doet, wordt gezien in het licht van wat anderen doen. Dit vergelijken is niet bedoeld om te identificeren maar om te differentiëren, de vergelijking is niet gericht op het vaststellen van gelijkenissen maar juist op het kunnen maken van het verschil. Het effect daarvan staat haaks op een gevoel van horen bij of cohesie maar eerder op vervreemding en afscheiding. Onderwijs wordt daarmee een bedreigende plek en de ergste straf is uitsluiting. En daarmee is examinering volgens Foucault geen vrijblijvende handeling toegepast in de context van onderwijs, maar grijpt diep in op de organisatie van de samenleving als geheel. Examinering is daarmee een belangrijk machtsmiddel om een samenleving te classificeren, te normeren en te disciplineren.

Een test, zeker waar de test in handen is van de controlerende macht, vraagt nader onderzoek naar de rol en de functie van de test in de classificatie en de daarop gebaseerde positionering in het sociale domein. Psycholoog Verhaeghe stelt dat 'classificatiesystemen nooit ideologisch neutraal kunnen zijn. Elke ordening in categorieën verwijst impliciet of expliciet naar een rangorde met een politieke en een morele betekenis' (Verhaeghe, 2012: 200). Het zich richten van mensen op metingen is het meest effectief volgens Danziger als door degenen die de aan de meting meedoen waarde wordt toegekend aan de uitkomsten. Dan geven metingen richting aan wat ze beogen te meten. Dan worden degenen die bij de meting betrokken zijn pionnen op het schaakbord en zijn 'governable' zoals Porter dat noemt (Porter, 1995: 45). Kwantificerende vergelijkingen en evaluaties van betrokken individuen leiden tot een ordening van individuen op basis van een statistische norm. Dit soort procedures ziet Danziger tegelijkertijd als technieken om individuen te disciplineren en ze vormen tevens de basis voor het produceren van een bepaald soort kennis die is vereist om bepaalde vormen van sociale controle in bepaalde institutionele contexten, zoals onderwijs, te rationaliseren (Danziger, 1990: 190). Pedagoog Meirieu spreekt over de 'dodelijke werking van classificaties': 'Ze vervangen het recht om een eigen weg te kiezen door het toewijzen van een specifieke plaats. Ze vervangen het project dat we zelf in vrijheid kunnen kiezen door de idee van een vaststaande, geërfde natuur.' (Meirieu, 2016: 98). Classificaties in het onderwijs staan volgens Meirieu haaks op de funderende principes van de democratische samenleving. Hij zet vraagtekens bij de legitimiteit van de classificaties in de context van onderwijs: 'Elk instituut moet de diversiteit beperken om haar te kunnen behappen. Maar het is

zaak om slechts voorlopige indelingen op te stellen met heterogene groepen, die corresponderen met vastgestelde behoeften van leerlingen en die bovendien constant herzien zouden kunnen worden. Een dergelijke benadering is heel anders dan het scheppen van categorieën die zeer snel als vaststaande labels gaan functioneren en dienst doen als een soort getto's, en zo individuele en collectieve vooruitgang blokkeren.' (Meirieu, 2016: 105). Uitkomsten van tests krijgen een eigen dynamiek en hebben invloed op degene die getest wordt, op de testafnemer en op de interpretaties van de uitkomsten die ermee gepaard gaan. Tests kunnen zich ontwikkelen tot mechanismen waarmee de werkelijkheid wordt gevormd, al dan niet bewust. Scores bepalen hoe mensen (en organisaties) worden behandeld: op basis van scores gebaseerd op resultaten uit het verleden, worden voorspellingen over de toekomst gedaan, waarbij wordt gedaan alsof een individu kan worden teruggebracht tot één score en wordt het verleden iets dat bepalend is voor de toekomst. In het licht van de ontwikkeling van kinderen gedurende hun onderwijsperiode acht de Franse pedagoog Meirieu dit niet wenselijk: 'Niemand moet ertoe veroordeeld worden zijn verleden te herhalen.' (Meirieu, 2016: 106).

De invoering van internationale vergelijkingen van onderwijsprestaties gebaseerd op gestandaardiseerde, vergelijkende metingen tussen landen zoals geïnitieerd door de OECD in de vorm van PISA-metingen hebben effecten, niet op de individuen die de tests afleggen maar met name op beleidsniveau, op het niveau van landen. Daar zorgen ze voor een verschuiving en een andere, maar niet minder effectieve, vorm van disciplineren. Ze laten volgens Mau zien hoe groot de invloed van een test in handen van een machtige internationale organisatie kan zijn. Het gebruik van deze tests heeft volgens hem zelfs een verschuiving in de definitiemacht teweeg gebracht: 'Nationaal zeer onderscheiden onderwijsradities zijn vervangen door metingen en vergelijkbaarheid van competenties. Wat als pedagogisch succes geldt, is hiermee door experts van een internationaal consortium vastgelegd waarmee successievelijk alternatieve kwaliteitscriteria zijn verdrongen.' (Mau, 2017: 202). Instrumenten in handen van machtige instanties vormen of vervormen de werkelijkheid. Dit heeft gevolgen voor het handelen van nationale overheden. De uitkomsten van dergelijke internationale metingen hebben soms verstrekkende gevolgen.

## 2.2. Testen en ordenen

Tests zoals die voorkomen in onderwijs en samenleving zijn te beschouwen als een vorm van metingen of meetbaar maken van dat wat moet worden gemeten. In het sociale en psychische domein is informatie over individuen of groepen niet altijd direct beschikbaar of te achterhalen. Daar is vaak nader onderzoek voor nodig. Een test kan worden beschouwd als een speciale vorm van onderzoek om die informatie, die niet rechtstreeks beschikbaar is, te achterhalen. Een test wordt daarmee intermediair tussen werkelijkheid en typering. En die typering wordt vervolgens vaak gebruikt om te ordenen en te classificeren. Daarmee is een test volgens antropoloog Hanson een soort onderzoek waarbij de informatie die verzameld wordt niet per definitie de informatie is die wordt gezocht maar betrekking heeft op de informatie die wordt beschouwd als een acceptabele afspiegeling van waar naar wordt gezocht. Hanson definieert een test als volgt: 'A test is a representational technique applied by an agency to an individual with the intention of gathering information.' (Hanson, 1993:19). Testen gaat daarmee eerst over de vraag wat je te weten wilt komen en vervolgens over wat geacht wordt een goede afspiegeling te zijn van wat we te weten willen komen. Hanson spreekt in dit verband over het eerste als 'the signified', het te meten construct, en over het tweede als 'the signifier', het resultaat van de meting (Hanson,

1993: 42). Hoe de relatie is tussen meting en construct, is in hoge mate cultuurbepaald, tijd- en plaatsgebonden. Die relatie hangt volgens Hanson samen met het 'geloof' dat er daar op dat moment op die plaats is in de relatie tussen dat wat gemeten wordt en dat waar men een uitspraak over wil kunnen doen. Zo werd een te laag gewicht in de middeleeuwen beschouwd als bewijs van heks zijn of stond het voelen van een erwt door tien matrassen heen ooit voor bewijs van hoge afkomst. Wat we als afspiegeling kiezen en voor waar aannemen, is een vorm van sociale en culturele conventie, van afspraken die we met elkaar maken, omdat we geloven dat juist die afspiegeling een goed beeld zou geven van de werkelijkheid en het vertrouwen dat we hebben in de kennis die dat oplevert. Het gevaar van 'vertekening' vindt daarmee plaats op het niveau van het construct als ook op het niveau van de gekozen afspiegeling.

Politiek wetenschapper en filosoof Elster spreekt over ordeningscriteria in het kader van verdelingsvraagstukken als 'proxies'. Dat zijn maten die worden geacht een goede afspiegeling te zijn voor het maken van rechtvaardige keuzes en daarmee samenhangende verdelingen. Het gaat hierbij om een benadering van de werkelijkheid, een proxy, die acceptabel is voor de categorisatie die aan verdelingen voorafgaat. Proxies zijn volgens Elster niet altijd even duidelijk in de normen en waarden die eraan onderliggend zijn. Omdat instituten zich gebonden weten aan wettelijke beperkingen en de publieke opinie er veel aan gelegen is te spreken over gehanteerde selectiecriteria in termen van 'facially neutral criteria of selection' (Elster, 1992:119), is het nog lastiger te achterhalen wat de gekozen principes zijn die met verdelingsvraagstukken samenhangen. En het maakt het lastiger om ze ter discussie te kunnen stellen. Daarvoor moeten ze eerst (weer) naar de oppervlakte gebracht en zichtbaar gemaakt worden. De omgekeerde weg ligt vaak meer voor de hand: door te onderzoeken wat er aan de oppervlakte gebeurt, kan afgeleid worden om welke principes het bij de verdeling van (schaarse) goederen feitelijk gaat.

De uitkomst van een test is per definitie aan subjectiviteit, aan geloof in de gehanteerde proxy, onderhevig. Als tests ingezet worden als instrument voor beleid dan hebben tests ook een politieke functie. Tests kunnen zo bezien geen metingen of indicatoren zijn van een objectieve, onafhankelijk bestaande stand van zaken of werkelijkheid. Tests zijn gericht op het spiegelen van de werkelijkheid zoals die cultureel is bepaald en wordt geloofd of politiek relevant is. De waarde van een test hangt samen met (het geloof in) de waarde die wordt toegekend aan hoe datgene dat zou moeten worden gemeten, geoperationaliseerd is in een test, of met het belang dat een test heeft in het kader van een specifieke beleidscontext. Alle beslissingen die voorafgaan aan het feitelijke testen en samenhangen met de interpretaties van de uitkomsten, zijn per definitie subjectief. Huisman bepleit het maken van een onderscheid tussen kennis die objectief verkregen is, 'objectieve kennis' en 'ware kennis'. Het voordeel van het maken van dit onderscheid is volgens haar dat dan eenvoudiger inzichtelijk gemaakt kan worden dat een objectieve manier van kijken geen garantie is voor het doen van ware uitspraken, en omgekeerd, dat een subjectief tot stand gekomen uitspraak niet per definitie onwaar hoeft te zijn (Huisman, 1997: 37).

De precisie, voorspelbaarheid en controleerbaarheid zijn bij het meten van mensen volgens Wilson moeilijk onder controle te houden. Van een 'ware' meting kan volgens Wilson in de menswetenschappen geen sprake zijn: 'So the very concept of a 'true' measurement resides in the assumption of a stability and permanence in the characteristic being measured, and the boundary conditions of the measurement.' (Wilson, 1998: hfdst 9, p. 5). Een score verkregen op basis

van een instrument moet volgens Wilson dan niet worden gezien als een 'label' dat hoort bij de proefpersoon maar veel meer als een label dat hangt aan de interactie van het geheel van 'theory-experiment-instrument-object' (Wilson, 1998: hfdst 9, p. 6).

Tests kennen veelal een verder gelegen doel, soms openlijk en soms verscholen en aan het zicht onttrokken, maar vinden zelden zomaar plaats voor informatieverzameling zelf. Dat maakt tests bij voorbaat niet onschuldig. Het maatschappelijk belang dat met testen gepaard gaat, is afhankelijk van wat er verdeeld gaat worden, wie verantwoordelijk is voor de verdeling, wat de waarde ervan is in de beleving van deelnemers en wat de gevolgen ervan zijn op de kortere en de langere termijn. Aan de uitkomst van een test wordt veelal een beslissing in de vorm van een classificatie gekoppeld. Testen hangt samen met vraagstukken van inclusie: wie of wat hoort er wel bij en wie of wat niet (meer)? Daarmee zijn tests, ook tests die gebruikt worden in de context van onderwijs, relevant om te onderzoeken op de waarden die samenhangen met het verder gelegen doel.

De eindtoets, gepositioneerd als een test voor en door onderwijs, heeft doelen die verder liggen dan de dagelijkse onderwijspraktijk. Die verder gelegen doelen hebben gevolgen voor leerlingen en hun ouders waar het de toewijzing van mogelijkheden voor vervolgonderwijs betreft, maar ook voor scholen en hun leerkrachten daar waar het de beoordeling van de kwaliteit van onderwijs betreft. De eindtoets ordent, classificeert en vergelijkt en aan de vergelijking zijn verstrekkende gevolgen verbonden. Daarmee dient de eindtoets beschouwd en onderzocht te worden vanuit een sociologisch perspectief: wat doet de eindtoets, hoe werkt het en wat zijn de effecten ervan op onderwijs maar vooral ook de samenleving?

### **2.2.1 Test als institutie**

Institutie, instituut, instelling en organisatie zijn begrippen die in het dagelijks spraakgebruik door elkaar worden gebruikt en voorkomen als synoniemen. In het boek 'Sociale kaart van Nederland; over instituties en organisaties' wordt bewust onderscheid gemaakt tussen de begrippen waarbij de auteurs Duyvendak et al. 'instituties' definiëren als 'de formele en informele regels die ons gedrag in meerdere of mindere mate reguleren, oftewel de spelregels van de samenleving' (Duyvendak et al., 2013: 13). Daarmee zorgen instituties voor een stabiel frame waarbinnen anticipatie onzekerheden reduceert en collectieve actie wordt gestructureerd. Instrumenten – een specifiek type van een institutie – beïnvloeden en structureren beleid. Sociale en politieke actoren hebben volgens Lascoumes en Le Gales de mogelijkheid acties te organiseren die grote verschillen kunnen laten zien, afhankelijk van de instrumenten die worden gekozen. Instrumenten maken, net als elke institutie, collectieve actie mogelijk om te stabiliseren, het gedrag van actoren meer voorspelbaar te maken en meer zichtbaar (Lascoumes & Le Gales, 2007: 9). In het proces van organisatie en institutionalisering worden door Duyvendak et al. drie fases onderscheiden: het ontstaan van instituties en organisaties, het voortbestaan van instituties en organisaties, en het veranderen of zelfs het verdwijnen van instituties en organisaties. Als verklaring voor het feit dat instituties en organisaties nauwelijks veranderen, is volgens Duyvendak et al. het begrip 'padafhankelijkheid' van belang: 'Van padafhankelijkheid is sprake als door een min of meer toevallige gebeurtenis of een toevallig proces in het verleden een dynamiek in gang wordt gezet waarvan vervolgens moeilijk af te wijken is.' (Duyvendak et al., 2013: 17). Institutionele verandering wordt door padafhankelijkheid belemmerd omdat het overgaan naar een alternatief institutioneel arrangement hoge kosten met zich mee kan brengen maar ook 'omdat mensen zo

gesocialiseerd worden in het bestaande arrangement dat zij alternatieven nauwelijks overwegen of serieus nemen' (Duyvendak et al., 2013: 18). Padafhankelijkheid is te zien als 'een conservatieve kracht' die er mede voor zorgt dat organisaties en instituties 'niet van het ene op het andere moment van karakter kunnen veranderen, maar vaak veel continuïteit kennen. Een andere reden voor hun continuïteit kan zijn dat in de loop der jaren de betrokkenheid bij de instituties en organisaties toegenomen is waardoor ze legitiemer zijn geworden. Op een gegeven moment wordt het naleven van de betreffende instituties en het voortbestaan van organisaties als vanzelfsprekend gezien.' (Duyvendak et al., 2013: 17). Institutionele veranderingen kunnen op drie manieren tot stand komen: schoksgewijs, het verliezen van relevantie of door permanent in beweging te blijven waardoor 'een arrangement ogenschijnlijk continuïteit kan vertonen, terwijl het onder de oppervlakte wel degelijk verandert' (Duyvendak et al., 2013: 20).

De eindtoets bestaat meer dan vijftig jaar in een min of meer vergelijkbare vorm. In 1967 verschijnt de voorloper van de huidige eindtoets, 'de Amsterdamse Schooltoetsen', ontwikkeld door een consortium dat zich de Samenwerkende Instituten noemt. Met de oprichting van het Centraal Instituut voor Toetsontwikkeling, in 1968, wordt toetsing letterlijk ondergebracht in een eigen instituut en daarmee institutionaliseert allereerst toetsing zich als een eigen tak van sport in de wetenschap. Door de eindtoets onder te brengen in dit zelfstandig toetsinstituut onder regie van de overheid, groeit de toets uit tot een belangrijk instrument waarmee de overgang van basis- naar voortgezet onderwijs gereguleerd wordt. De eindtoets, in inhoud en werking, bepaalt zo in hoge mate wat relevant geacht wordt voor die overgang. Dit geldt ook voor de bepaling van de kwaliteit van onderwijs. Normen en waarden, gebaseerd op de eindtoets, vormen daarmee het denken en handelen van allen die bij de overgang en kwaliteitsbepaling zijn betrokken. Daarmee groeit de eindtoets uit tot een institutie die past in de definitie zoals Duyvendak et al. die geven.

Instituties zijn nooit zuiver instrumenteel of waardenvrij. Integendeel, instituties geven uitdrukking aan de normatieve doelstellingen die de maatschappelijke orde nastreeft. Elke institutie heeft dan ook, naast een empirisch, een normatief aspect (Lehning, 1991: 213). Een test is enerzijds te zien als een technisch hulpmiddel om bepaalde maatschappelijke doelstellingen te bereiken. Tegelijkertijd geeft een test ook uitdrukking aan normatieve doelstellingen die de maatschappelijke orde nastreeft. Daarmee is een test te beschouwen als een institutie met een constituerende rol. Hoe natiestaten hun onderwijssystemen vormgeven weerspiegelt volgens De Vijlder 'in hoge mate de waarde- en normenpatronen die in die samenleving domineren op het moment dat instituties gevormd worden als sublimaat van die waarden en normen. Instituties en de ontwikkeling van waarde- en normenpatronen houden elkaar voor een belangrijk deel in hun onderlinge relatie gevangen.' (De Vijlder, 1996:47). Lehning stelt dat 'Instituties doel en middel tegelijk zijn' (Lehning, 1991: 205). Bij bestudering van instituties gaat het volgens Lehning dan (ook) om een morele (re)constructie van de werkelijkheid. Lehning wijst op het feit dat instituties in iedere sociaal politieke orde een centrale rol spelen. Daarmee zit volgens Lehning altijd in het empirische ontwerp van een institutie een normatief keuze-element besloten, dat onder andere samenhangt met het mensbeeld dat gehanteerd wordt en de rationaliteitsopvatting die tot uitgangspunt wordt genomen. Van belang is daarbij om te expliciteren welke normatieve doelstellingen nagestreefd worden, welke constituerende rol instituties daarbij moeten spelen.

### 2.2.2 Test als beleidsinstrument

Een beleidsinstrument constitueert een instrument dat zowel technisch is als sociaal, dat speciale sociale relaties organiseert tussen de staat en degenen voor wie het bedoeld is, volgens de afspiegelingen en betekenissen die het draagt. Een beleidsinstrument is daarmee een specifiek soort institutie, een technisch instrument met een generieke bedoeling om een concreet concept van de maatschappij uit te dragen, ondersteund door een vorm van regulering. Beleidsinstrumenten zijn geen tools met een perfecte axiologische neutraliteit, gelijkelijk beschikbaar: integendeel, het zijn dragers van waarden, van brandstof voorzien door één interpretatie van het sociale en door precieze noties van de manier waarop regulatie wordt beoogd. Het betekent dat niet alleen de redenen die leiden tot de inzet van het ene instrument in plaats van een ander moeten worden begrepen maar dat ook de effecten die met de keuzes samenhangen onder ogen moeten worden gezien. Ieder instrument heeft een geschiedenis, waarvan het de drager blijft, en eigenschappen van een instrument zijn niet te scheiden van de doelen die met het instrument beoogd worden. Lascoumes en Le Gales willen niet blijven hangen in eindeloze discussies over de aard van een instrument maar willen vooral een instrument vanuit het oogpunt van de werking ervan met twee elkaar aanvullende invalshoeken bekijken:

1. door de effecten te bezien die door de instrumenten in relatieve autonomie worden gegenereerd, en
2. door die aan te vullen met de politieke effecten van de instrumenten en de machtsrelaties die zij organiseren.

Voor een instrument als de eindtoets betekent dit dat het van belang is de werking van het instrument op zich te onderzoeken, de meer technische kant ervan, maar dat het ook van belang is het instrument in een bredere context te plaatsen en in zijn politieke functies en effecten te onderzoeken. Daarbij zijn ook de gevolgen voor de machtsrelaties relevant die mede op basis van de eindtoets vormgegeven worden. Ook die moeten nader onderzocht worden om de werking van instrument en samenleving te kunnen doorgronden.

Onderzoek naar de rol en aard van beleidsinstrumenten is van belang om overheidsbeleid te begrijpen en te doorgronden. Beleidsinstrumenten maken de relatie tussen de heerser en de gene die beheerst wordt zichtbaar omdat elk instrument een gecondenseerde vorm van kennis is over sociale controle en de manieren waarop die uitgeoefend wordt (Lascoumes & Le Gales, 2007). Een tweede belang dat aan onderzoek naar beleidsinstrumenten kan worden gehecht, is dat instrumenten in de praktijk geen neutrale instrumenten zijn: ze produceren speciale effecten, onafhankelijk van het doel waar ze naar streven (beoogde doelen) waarmee ze overheidsbeleid structureren volgens hun eigen logica. Hoewel instrumenten een technische of functionalistische benadering gebruiken, verbergen ze wat er feitelijk 'at stake' is, politiek gezien. Door de nadruk te leggen op de politieke sociologie van beleidsinstrumenten, kunnen machtsrelaties benadrukt worden die met de instrumenten samenhangen en vraagstukken van legitimiteit, politisering of depolitiseringsdynamieken die met verschillende beleidsinstrumenten geassocieerd worden (Lascoumes & Le Gales, 2007).

Het type instrument dat gebruikt wordt, de eigenschappen ervan en de rechtvaardigingen van de keuzes voor bepaalde instrumenten lijken meer onthullend dan verantwoordingen van motieven

of latere discursieve rationalisaties. Instrumenten die hun werk doen zijn niet puur technisch: ze produceren speciale effecten, onafhankelijk van hun geformuleerde of aan hen toegeschreven doelen, en ze structureren public policy volgens hun eigen logica. Ze hebben hun eigen actie-kracht met eigen effecten: het inertia-effect, een speciale/eigenaardige (particular) representatie van het issue at stake, of een specifieke problematisering van het issue. Het belang van het onderzoeken van beleidsinstrumenten is volgens Lascoumes en Le Gales dat daarmee onzichtbare en gedepolitiseerde dimensies van beleid weer zichtbaar gemaakt kunnen worden (Lascoumes & Le Gales, 2007: 17). Achter de keuze voor de eindtoets als instrument van beleid gaan politieke keuzes schuil en vormen van rechtvaardiging. Door de lange looptijd van de eindtoets hebben onderliggende politieke keuzes en vormen van rechtvaardiging zich ontwikkeld. Voor de eindtoets is het van belang te onderzoeken welk handelen met de toets als instrument wordt beoogd, gereguleerd en zichtbaar wordt gemaakt en welke keuzes daaraan ten grondslag liggen. Lascoumes en Le Gales wijzen op het belang van onderzoek naar de langetermijnpolitieke carrières van beleidsinstrumenten, om de debatten rond de creatie en introductie te analyseren, de manieren waarop de instrumenten zijn aangepast c.q. veranderd te analyseren en de ermee samenhangende controverses. In een politieke context waar ideologische vaagheid de voorkeur heeft – of, ten minste, waar ideologie minder zichtbaar is – en waar het maken van verschil tussen discoursen en programma's steeds moeilijker te bewijzen is, kan het standpunt worden ingenomen dat juist in beleidsinstrumenten gedeelde beelden rond sociale kwesties zich stabiliseren (Lascoumes & Le Gales, 2007: 18). Voor de eindtoets met een historie van meer dan vijftig jaar voor het instrument zelf, maar ook geworteld in ontwikkelingen van de periode daarvoor, geldt dat wisselende politieke en maatschappelijke ontwikkelingen zijn neergedaald en geconsolideerd in de eindtoets. Met de wetwijziging van 2011 is de eindtoets formeel een beleidsinstrument van de overheid die het instrument inzet als verplichte toets voor alle leerlingen van het basisonderwijs aan het eind van groep 8 voor de toewijzing aan het voortgezet onderwijs en voor de beoordeling van kwaliteit van het primair onderwijs. Deze verandering van positie en functie van de eindtoets heeft gevolgen voor de relaties die betrokkenen in het onderwijs met elkaar hebben. De overname van een marktproduct van een commerciële partij door een overheid en het door diezelfde overheid verplichte gebruik ervan, maakt van dat marktproduct een nieuw product, los van de aard van het product. Zelfs als aan de eindtoets zelf niets verandert, is de nieuwe constellatie waarbinnen de eindtoets als beleidsinstrument gepositioneerd wordt, bepalend voor de relatie die er is tussen de actoren, in termen van de 'heerser' en degene die 'beheerst' wordt (governing versus governed). De overheid is nu verantwoordelijk voor het beleidsinstrument en de normatieve uitwerking ervan. Controle op zowel toelating tot het voortgezet onderwijs als op de beoordeling van kwaliteit van het voorgaande onderwijs is in handen van de overheid. Dit vraagt enerzijds een goede doorgronding van en reflectie op het instrument zoals de overheid dat heeft overgenomen en een bezinning op hoe de overheid het instrument inzet in de praktijk van onderwijs, de beoogde beleidsdoelen en gewenste en ongewenste effecten ervan voor onderwijs en samenleving. Door de eindtoets in zijn (historische) context nader te onderzoeken kan een beeld worden verkregen van wat de overheid voor ogen heeft gehad en heeft, bewust of onbewust, met de inzet van dit instrument ten aanzien van de inrichting van onderwijs en samenleving, wat daaraan vooraf ging en welke rol de overheid daar toen in vervulde en waarom juist dit instrument is gekozen en of hiermee de beleidsdoelen die de overheid met het beleidsinstrument beoogt, kunnen worden gerealiseerd.

De eindtoets als technisch instrument met een normatief karakter is uitgegroeid tot een beleids-

instrument met een rol van betekenis voor het denken over en richten van onderwijs en samenleving gebaseerd op de normen en waarden waaraan met behulp van de eindtoets uitdrukking wordt gegeven. Een eerste belangrijk aspect van de onderliggende waarde heeft te maken met de benaming van de eindtoets: is hier sprake van een test of een toets?

### 2.2.3 Test of toets?

Het woord 'test' komt uit het Engels. Met behulp van toevoegingen wordt in het Engelstalige literatuur nader uitgewerkt wat voor soort test wordt bedoeld. In zijn 'Essentials of Psychological Testing' (1960) ordent en benoemt Lee J. Cronbach tests naar de beslissingen die op basis van een test genomen moeten kunnen worden. Een test is volgens Cronbach een 'systematische procedure om het gedrag van een persoon te observeren en te beschrijven op basis van een numerieke schaal of een categorisatie-systeem' (Cronbach, 1960: 34). Hij onderscheidt tests gericht op 'selectie' van 'classificatie' en van 'placement'. Het Engels gebruikt naast het woord 'test' ook het woord 'assessment' maar daarbij treedt de nodig verwarring op door de inconsequente toepassing van termen. In het Nederlands zijn zowel de woorden 'test' als 'toets' in gebruik waarbij de keuze voor ofwel het ene ofwel het andere woord diffuus blijkt. In hun boek 'Studietoetsen' (1969) stellen de auteurs De Groot en Van Naerssen dat de woorden 'test' en 'toets' als equivalent te beschouwen zijn (De Groot & Van Naerssen, 1969: 198). In de verantwoording van de Amsterdamse Schooltoetsen geeft De Groot geen onderbouwing van de reden dat hij kiest voor het woord 'toets' in plaats van 'test'. Hij volstaat met de opmerking dat hij school- of studietoetsen een gelukkiger term vindt dan schoolvorderingentests (Samenwerkende Instituten, 1967: 18). Het Cito gebruikt sinds de oprichting in 1968 niet alleen 'toets' in de naam van het instituut, 'Centraal Instituut voor Toetsontwikkeling', maar ook consequent in publicaties. Drenth en Sijtsma scharen in hun boek 'Testtheorie' (1990) vorderingentests onder de groep 'tests voor prestatieniveau'. In het trefwoordenregister van hun boek komt het woord 'toets' niet voor (Drenth & Sijtsma, 1990: 38). Uiterwijk werkt het onderscheid tussen 'test' en 'toets' nader uit door te stellen dat het begrip 'toets' in het Nederlands kan worden beschouwd als een verbijzondering van het begrip test. Het begrip 'toets' wordt gebruikt voor een meetprocedure van door onderwijs en studie verworven kennis, inzicht en vaardigheid op een of meer vakgebieden. 'Test' wordt gebruikt voor een meetprocedure van niet door intentioneel onderwijs en studie verworven eigenschappen van de persoon' (Uiterwijk, 1994: 7). Daarmee zou een test duidelijk te onderscheiden zijn van een toets. Cito beschikt naast de eindtoets die Cito typeert als een 'leervorderingentoets' ook over een intelligentietest. In de verantwoording van die intelligentietest wordt gesteld dat 'de twee testsoorten een continuüm vormen, waarvan de uitersten duidelijk zijn te onderscheiden, maar waarin geen nauwkeurige grens tussen de soorten te trekken is'. Dit geldt ook voor de inhoud volgens de verantwoording: 'Kortom, volgens deze opvatting bestaat er dus geen fundamenteel onderscheid tussen taken binnen intelligentietests en leervorderingentoetsen, zoals dat wel bestaat tussen bijvoorbeeld intelligentietests en motorische tests, maar alleen een gradueel, door onze onderwijscultuur bepaald verschil.' (Boxtel & Hemker, 2009: 9). Het verschil tussen de beide tests is meer te relateren aan de functie: 'De pretentie van de Intelligentietest is dat deze er beter dan de leervorderingentoets in slaagt om zicht te geven op het intellectuele potentieel dat bij de leerling aanwezig is. De Eindtoets daarentegen geeft beter weer in hoeverre een leerling erin slaagt dit potentieel ook daadwerkelijk tot ontplooiing te brengen.' (Boxtel & Hemker, 2009: 15).

De Nederlandse psychometrici Van den Brink en Mellenbergh sluiten aan bij de definitie die



Cronbach geeft van een test. Cronbach stelt dat een test 'een systematische procedure is om het gedrag van twee of meer personen te vergelijken' (Van den Brink & Mellenbergh, 2006: 19). Zodra een test niet meer bedoeld is voor het vergelijken van personen met elkaar maar met een extern criterium zoals een onderwijsdoelstelling dan geldt de definitie van Cronbach niet meer en is er geen sprake meer van een 'test' en dat heeft volgens hen gevolgen voor de psychometrische theorievorming (Van den Brink & Mellenbergh, 2006: 20). In de 'richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen' (1988) wordt door KOTAN (Kommissie Testaangelegenheden Nederland) gesteld dat zij aan het woord 'test' een wat 'lossere en meer omvattende' definitie geven zodat 'ook studietoetsen en andere niet volledig persoonsgebonden instrumenten' hieronder vallen (KOTAN, 1988: 3). Tegelijkertijd maakt KOTAN een onderscheid tussen 'test' en 'toets' waarbij een toets bedoeld is om te 'meten in welke mate iemand een bepaalde hoeveelheid kennis bezit, een bepaalde mate van inzicht toont, of een bepaalde vaardigheid beheerst, meestal op het gebied waarop men is onderwezen'. Een test is volgens KOTAN gericht op 'het mogelijk maken van vergelijkingen tussen personen' (KOTAN, 1988: 138). De COTAN, Commissie Testaangelegenheden Nederland (eerder KOTAN), een centrale commissie van het Nederlands Instituut voor Psychologen (NIP), hanteert consequent het begrip 'test' en lijkt geen onderscheid te maken (Evers et al., 2009). COTAN beoordeelt zowel tests als toetsen om op basis van die beoordeling tests en toetsen een keurmerk te kunnen toekennen. COTAN heeft ook de toetsen van het Leerlingvolgsysteem (LVS) van Cito beoordeeld. In het beoordelingssysteem dat gehanteerd wordt voor de beoordeling van tests en toetsen blijkt dat de beoordeling op een zelfde manier verloopt voor een test als voor een toets. Psychometricus Wilbrink wijst op het wel degelijk bestaan van een verschil tussen een test en een toets. Hij noemt het onderscheid tussen tests en toetsen 'een klein verschil met grote gevolgen' (Wilbrink, 1986: 278). Tests staan volgens hem voor de categorie meetinstrumenten waarbij verondersteld wordt dat er geen specifieke voorbereiding (in termen van onderwijs en/of zelfstudie) op de test heeft plaatsgevonden. Voor een toets geldt dat men zich er door onderwijs en/of zelfstudie behoorlijk op heeft voorbereid omdat het hierbij gaat om het vaststellen van de beheersing van voor de kandidaat bekende leerstof. Voorbereiding op een test heeft gevolgen voor de validiteit van een test. En andersom, het niet voorbereiden op een toets heeft ook gevolgen voor de bepaling van de validiteit van een toets. Wilbrink acht het van belang oog te hebben voor het verschil in interpretatie van de begrippen test en toets en vraagt 'om argwaan wanneer begrippen en technieken uit de psychometrie [bedoeld voor tests] worden geleend zonder argumentatie waarom deze ook voor toetsen adequaat zouden zijn' (Wilbrink, 1986: 283). Hij wijst op de schadelijkheid van het 'met huid en haar slikken van de onderliggende filosofie van het psychometrisch bouwwerk dat er vooral om gaat verschillen tussen studenten ... te meten' (Wilbrink, 1986: 283). Hij stelt dat er 'voldoende verschillen zijn tussen (het gebruik van) toetsen en tests om te kunnen stellen dat het construeren van een analogie van tests naar toetsen voor het onderwijs een minder gelukkige ontwikkeling is geweest die enige correctie behoeft' (Wilbrink, 1986: 284).

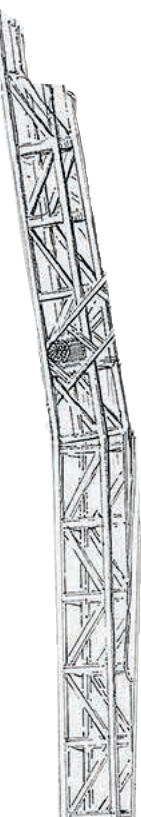
Het al dan niet terecht en of consequent gebruiken van het ene of het andere begrip heeft in Nederland geen vaste grond onder de voeten gekregen. Uit de naamgeving kan momenteel niet met zekerheid worden afgeleid of er sprake is van een test of een toets of om wat voor test of toets het daarbij gaat. Er is verdere analyse van het instrument zelf nodig, van de ontwerpprincipes en de gehanteerde psychometrie om te achterhalen met wat voor soort instrument we te maken hebben. Vanuit het perspectief van de functie van selectie is de eindtoets te typeren als

psychologische test, een test waar leerlingen zich niet op hoeven voor te bereiden. Vanuit het perspectief van kwaliteit is de eindtoets meer te beschouwen als een eindexamen waar voorbereiding wel degelijk voor de hand ligt. De eindtoets heeft dan ook twee gezichten en is te typeren als een hybride vorm van toetsing. Dit zorgt voor spanning waarbij het de vraag is of beide functies even goed vervuld kunnen worden. De onduidelijkheid die bestaat rond de hantering van de begrippen 'test' en 'toets' is, zeker achteraf gezien, belangrijk geweest voor de rol en positie die de eindtoets heeft kunnen innemen, zoals in het vervolg zal blijken.

Hier volstaat het nu om het gebruik van de begrippen 'test' en 'toets' in dit onderzoek toe te lichten. Vanwege het ontbreken van een algemeen gangbare definitie is ervoor gekozen geen systematisch onderscheid te maken in het gebruik van de woorden 'test' en 'toets' en zal voornamelijk het woord 'test' (of instrument) gebruikt worden als het gaat om de meer algemene aanduiding van een meetinstrument waarbij het vergelijken centraal staat en wordt gekozen voor het woord 'toets' als het gebruik van 'test' zou leiden tot verwarring vanwege de bekendheid met de begripshantering van het woord 'toets' in die specifieke context.

### 2.3 Samengevat

Met de ordeningen en classificaties die met behulp van tests plaatsvinden, hangen keuzes samen. En die keuzes zijn per definitie niet neutraal. Vaak is aan de hand van een test zelf moeilijk te herleiden welke subjectieve keuzes erachter schuilgaan of welke politieke besluiten er als basis onder liggen. De eindtoets heeft zich in een lange historie ontwikkeld tot een belangrijke bepaler van onderwijskansen voor leerlingen en daarmee ook tot een bepaler van hoe naar onderwijs wordt gekeken en hoe (kwaliteit van) onderwijs wordt gezien. De eindtoets is van beleidsinstrument uitgegroeid tot een institutie in de politieke en sociologische betekenis van het woord. In de volgende hoofdstukken wordt de politieke carrière van de eindtoets nader onderzocht. Als eerste wordt in het volgende hoofdstuk onderzocht hoe de eindtoets zich verhoudt tot de ambities van samenleving en onderwijs, hoe de eindtoets werkt en uitwerkt en hoe naar de waarde van de eindtoets in termen van validiteit gekeken kan worden.



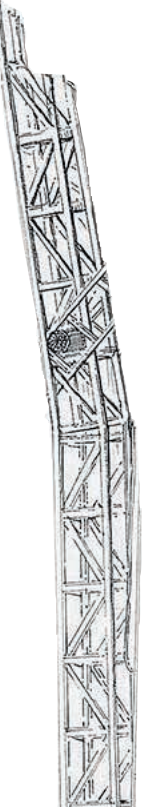






# HOOFDSTUK 3

Waarde en waardering van onderwijs



## Hoofdstuk 3 Waarde en waardering van onderwijs

*'Het onderwijs zal gericht zijn op de volle ontwikkeling van de menselijke persoonlijkheid en op de versterking van de eerbied voor de rechten van de mens en de fundamentele vrijheden. Het zal het begrip, de verdraagzaamheid en de vriendschap onder alle naties, rassen of godsdienstige groepen bevorderen.'*

*Uit: Universele verklaring van de rechten van de mens (1948)*

In onderwijs wordt weerspiegeld wat voor een samenleving van waarde is. Onderwijs, en in het verlengde daarvan toetsing, staat daarmee niet op zich. Doelen van onderwijs hangen samen met de rol die aan onderwijs wordt toegekend in een samenleving. Ontwikkeling van nationale onderwijssystemen en de ontwikkeling van natiestaten zijn nauw met elkaar verbonden. Onderwijs is voor natiestaten een belangrijke manier om te komen tot bredere sociale, culturele en economische ordening, die in elk land zijn specifieke kenmerken heeft. De relatie tussen natiestaten en onderwijs staat voortdurend onder druk door ontwikkelingen in de samenleving zelf en de wereld daaromheen (De Vijlder, 1996: 269). De Amerikaanse filosofe Nussbaum wijst op het feit dat onderwijsstelsels in met name de westerse wereld zich in de laatste decennia meer zijn gaan ontwikkelen richting economische groeimodellen. Zij spreekt van een 'stille crisis' (Nussbaum, 2011: 15), omdat die ontwikkeling plaatsvindt zonder stil te staan bij de vraag hoe zich dat verdraagt met democratie (Nussbaum, 2011: 43). Een te grote gerichtheid op toetsing heeft volgens Nussbaum gevolgen voor de democratische ontwikkeling van leerlingen. In plaats van de verbeeldingskracht te stimuleren en kritische vermogens te oefenen is de inhoud van onderwijs verschoven naar het voorbereiden op toetsen. Dat heeft gevolgen voor de manier waarop onderwijs wordt gegeven. Van onderwijs dat streeft naar het stellen van vragen en het stimuleren van individuele verantwoordelijkheid is het verschoven naar een vorm van onderricht dat kinderen volpropt met feiten om ze goede cijfers te laten behalen bij de examens (Nussbaum, 2011: 178). Dit verhoudt zich moeizaam tot het streven naar kansengelijkheid, want die is niet gediend bij het feit dat kinderen geen kans krijgen op het volgen van een stimulerende opleiding of een adequate voorbereiding op burgerschap (Nussbaum, 2011: 183). Vanuit het doel van onderwijs voor samenlevingen meer in het algemeen en de eigen samenleving in het bijzonder, kan de rol en functie van toetsing worden onderzocht en op waarde worden geschat. En omgekeerd. Daar waar het doel van onderwijs of meer specifieke doelen voor onderwijs ontbreken of niet erg expliciet zijn, kan toetsing inzicht bieden op wat van waarde wordt gevonden voor samenleving en onderwijs. Inzicht in de rol van toetsing in relatie tot onderwijs en samenleving kan ook op gespannen voet blijken te staan met intenties van beleid. Toetsing laat zien hoe effectuering van beleid gestalte krijgt en hoe dit zich verhoudt tot oorspronkelijke ambities van de overheid ten aanzien van onderwijs en samenleving. In dit hoofdstuk staat de relatie tussen samenleving, onderwijs en toetsing centraal. Om de waarde (de validiteit) van de eindtoets als test maar ook breder als beleidsinstrument voor de inrichting van onderwijs en samenleving, te kunnen bepalen is het van belang allereerst de vraag te beantwoorden welk doel onderwijs heeft of welk doelen met behulp van onderwijs (kunnen) worden nagestreefd. Door toetsing nader te onderzoeken op inhoud, vorm en uitwerking kan vanuit de instrumentele kant een beeld worden verkregen van dat wat blijkbaar van waarde wordt geacht voor onderwijs en samenleving. Zo ontstaat van twee kanten een beeld van hetgeen in Nederland van waarde wordt geacht en van hetgeen wordt gewaardeerd voor onderwijs en samenleving. Dit hoofdstuk start met een beschrijving van doel

en functie van onderwijs meer in het algemeen (3.1). Vervolgens wordt gekeken naar keuzes die samenlevingen hebben en maken in de vormgeving van sturing, structuur en kwaliteit van onderwijsstelsels (3.2). De keuzes die Nederland heeft gemaakt worden beschreven (3.3). Vervolgens wordt ingegaan op de rol die de eindtoets momenteel vervult in relatie tot onderwijs en samenleving en hoe deze toets daarop inwerkt en uitwerkt. De rol van de eindtoets wordt daarbij in samenhang met de het verplicht gebruik van een leerlingvolgsysteem gezien (3.4). Tot slot wordt de waarde van de eindtoets in termen van validiteit nader onderzocht. Het begrip validiteit wordt daarbij eerst nader verkend en beschreven alvorens de beide functies van de eindtoets nader op onderzochte en verklaarde waarde te onderzoeken, om zo de vraag te beantwoorden in hoeverre de eindtoets de functies die met de toets in relatie tot onderwijs en samenleving worden beoogd weet te onderbouwen en waar te maken, redenerend vanuit het concept van validiteit (3.5).

### 3.1 Onderwijs en samenleving

Sociaal antropoloog Gellner beschouwt onderwijs als een kenmerk van en onlosmakelijk verbonden met de moderne samenleving. Hij maakt onderscheid tussen een-op-een vormen van opleiding die meer traditionele samenlevingen gebruiken voor het opvoeden en opleiden van leden van de gemeenschap waarbij sprake is van een vorm van 'acculturatie'. Daarnaast onderscheidt hij de meer gespecialiseerde training die skills vereisen die buiten de gemeenschap liggen, waarmee wat hij noemt vormen van 'exo-training' ontstaan; vormen van opvoeden en opleiden die niet langer door de gemeenschap zelf kunnen worden vervuld. 'Exo-training' is een bij de moderne samenleving passende gecentraliseerde methode van reproductie: een los van de lokale gemeenschap staande instantie van specialisten voedt en leidt kinderen op zodat ze passen bij de kenmerken van de moderne, geïndustrialiseerde samenleving: 'Exo-socialization, the production and reproduction of men outside the local intimate unit, is now the norm, and must be so.' (Gellner, 2008: 37). Alleen de onderwijsmachine is in staat die brede basistraining te verzorgen die vereist is voor het leggen van een algemene, culturele basis. Gellner stelt dat daarmee alleen de staat kan waken over de inhoud en kwaliteit van die basis waarmee de rol van de staat in onderwijs onvermijdelijk is geworden: exo-socialisatie is volgens Gellner de belangrijkste reden dat staat en cultuur met elkaar verbonden moeten zijn. Exo-socialisatie is voor een moderne samenleving de norm en onvermijdelijk in de optiek van Gellner (Gellner, 2008: 37).

Die verbondenheid zorgt voor een permanente wisselwerking tussen onderwijs en samenleving; de samenleving is bepalend voor de 'kleur' die onderwijs aanneemt en onderwijs 'kleurt' mede de samenleving. Voor een samenleving is het van belang nieuwe burgers in te voegen in de bestaande maatschappelijke orde. Daarmee is onderwijs een instrument in het socialisatieproces van burgers. Onderwijs is volgens pedagoog Biesta altijd te zien als een interventie in het leven van een andere mens: onderwijspraktijken zijn socialisatiepraktijken (Biesta, 2016: 14). Het onderwijs is bij uitstek de plaats waar kinderen kennismaken met de bestaande wereld waarin zij zich moeten voegen. Politiek filosoof Hannah Arendt spreekt in dat kader over 'het in de wereld laten komen' van kinderen en jongeren. Arendt beschouwt onderwijs als een 'tussenruimte' tussen gezin en samenleving (Berding, 2016: 186). Afhankelijk van een samenleving kan socialisatie, het in de wereld laten komen, verschillend worden gezien en uitgewerkt. Biesta wijst er nadrukkelijk op dat dit niet hoeft te betekenen dat onderwijs 'louter de dienstknecht is van de bestaande orde' (Biesta, 2016: 14). Het is aan de samenleving om door een voortdurend debat over doel en functie van het onderwijs te komen tot een steeds passende balans. Want het gaat in on-



derwijs niet alleen om het invoegen in en laten aanpassen aan de bestaande maatschappij van een nieuwe generatie. De focus moet ook gericht zijn op het individu, op de cultivering van de menselijkheid van iedere individuele mens (Biesta, 2016: 14). Er zijn op zijn minst drie domeinen waarin het onderwijs functioneert en waarin onderwijsdoelen kunnen worden geformuleerd: kwalificatie, socialisatie en subjectificatie (Biesta, 2015: 20). Kwalificatie heeft betrekking op het vergaren van kennis en vaardigheden nodig voor het functioneren op de arbeidsmarkt, socialisatie is gericht op het kunnen functioneren in de samenleving meer in de breedte en subjectificatie heeft betrekking op de ontwikkeling van een individu in het ontdekken hoe het zich verhoudt en wil verhouden tot de bestaande samenleving. Die ontwikkeling plaatst Biesta in de voortdurende interactie tussen de aanbieder van onderwijs en de ontvanger van het onderwijs waarbij de uitkomst van dat proces per definitie niet voorspelbaar kan zijn. Deze meer pedagogische definiëring van onderwijs heeft volgens Biesta in de afgelopen decennia langzaam plaats gemaakt voor een meer op economische en marktmodellen gebaseerde manier van denken over onderwijs. In het onderwijs heeft een op het oog vooral talige verschuiving plaatsgevonden van 'opvoeding en onderwijs' naar 'leren'. Maar die verschuiving is volgens Biesta meer dan een talige exercitie; het heeft een nieuwe definiëring van het onderwijsproces mogelijk gemaakt. Een proces waarbij onderwijs meer gezien wordt als een economische transactie: de lerende kan gezien worden als de potentiële consument en de docent c.q. de school als de vervuller van wensen en behoeften. Het onderwijs is zelf handelswaar geworden: een ding dat door de docent of school wordt geleverd en door de lerende wordt geconsumeerd (Biesta, 2016: 34). Dit denken over onderwijs heeft ruimte gemaakt voor vragen naar efficiëntie en effectiviteit van het onderwijsproces. Biesta wijst op het gevaar van deze ontwikkeling omdat het in zijn optiek op de langere termijn de mogelijkheid voor democratische discussie over doelen en bedoeling van onderwijs ondermijnt (Biesta, 2016: 39).

De spanning die er is ten aanzien van doel en functie van het onderwijs, variërend van een meer economisch georiënteerd model tot een meer democratisch model is terug te zien in wat onderwijssystemen verkiezen voor de sturing van hun onderwijs in termen van inhoud en doelstellingen. De Vlaamse pedagoog Standaert spreekt van een doorlopende schaal waarop de inkleuringen van leerinhouden en doelstellingen zich begeven. Dit kan enerzijds worden ingegeven door een meer economisch-technische benadering die Standaert 'technocratisch' noemt, of anderzijds meer worden ingekleurd vanuit het individu, de pedagogisch-didactische benadering die Standaert definieert als 'leerlinggericht' (Standaert, 2016: 123). In het eerste geval staan de opbrengsten van het onderwijs centraal, uitgedrukt in voorgeschreven curricula, opgelegde toetsen en examens en zich op basis daarvan verantwoordende uitvoerders en deelnemers die met elkaar in competitie zijn over de te krijgen beloningen. In het tweede geval wordt uitgegaan van een onvoorwaardelijk recht van een jongere op een zo goed mogelijke, brede vorming die de leerling volgt, op maat, waarbij vormingen kunnen verschillen maar niet in waarde en de uitkomsten van onderwijs grotendeels onvoorspelbaar zijn (Standaert, 2016: 124). Onderwijs ingebed in een technocratische maatschappij is volgens Standaert te herkennen aan hoe de inhoud van het onderwijs wordt gekozen, hoe de structuur van het onderwijs is opgebouwd en hoe de kwaliteitszorg en kwaliteitsbewaking zijn vormgegeven. Hij werkt de kenmerken van een technocratisch onderwijssysteem uit in waarneembare aspecten zoals hieronder weergegeven (Standaert, 2014: 32-37):

<b>Kenmerken</b> <b>technocratisch</b> <b>onderwijssysteem</b>	<b>Aspecten</b>
<b>Inhoud</b>	<ul style="list-style-type: none"> <li>- Gescheiden leerinhouden/vakkensplitsing met een duidelijke hiërarchie in vakken</li> <li>- Sterke identificatie met een vak</li> <li>- Nadruk op cognitieve doelstellingen ten koste van affectieve, morele en waarde- vorming en muzische, artistieke en expressieve vorming</li> <li>- Toetsen bepalen in hoge mate het onderwijs (toetsconstructeurs als definieerders van curricula)</li> </ul>
<b>Structuur</b>	<ul style="list-style-type: none"> <li>- Gebrekkige, structurele samenhang met relatief weinig overstapmogelijkheden</li> <li>- Utilitaire en concurrentiële selectiesystematiek gebaseerd op objectiviteit en ge- loof in meetbaarheid, uitgaand van meritocratische gedachtegoed dat leidt tot (ex- terne) selectie ontleend aan een statische opvatting over intelligentie en leidend tot een vroegtijdige indeling in vroege en vaste onderwijsvormen.</li> <li>- Gebaseerd op het marktprincipe: keuzevrijheid voor ouders/kinderen, wat in com- binatie met selectieprincipes leidt tot homogene scholen en een dualisering op basis van sociaaleconomische achtergrond</li> </ul>
<b>Kwaliteitszorg/ kwaliteitsbewaking</b>	<ul style="list-style-type: none"> <li>- Toename externe controlemechanismen met behulp van (digitale) controlesyste- men waarmee de eigen professionaliteit wordt verwaarloosd</li> <li>- Toename van wantrouwen wat leidt tot 'shaming and blaming'</li> <li>- Hoe technocratischer hoe meer inzet van externe controlemechanismen als: <ul style="list-style-type: none"> <li>• gedetailleerde leerplannen,</li> <li>• centrale examens,</li> <li>• externe periodieke peilingen,</li> <li>• beperking van de inspectiefunctie tot 'aanvinkinspecties',</li> <li>• bekostiging van scholen op basis van resultaten,</li> <li>• gesofisticeerde evaluaties van leerkrachten en directies,</li> <li>• premies en loonsverhogingen op basis van behaalde resultaten,</li> <li>• het publiceren van de schoolresultaten en het rangschikken van scholen,</li> <li>• grotere macht van ouders en externen in de scholen ten nadele van de leraren,</li> <li>• het bevorderen van marktmechanismen bij het rekruteren van leerlingen,</li> <li>• subsidies en sponsoring door het bedrijfsleven, het toekennen van 'onderwijsbonnen' (vouchers),</li> <li>• centraal opgelegd rapporteringssysteem,</li> <li>• verplichting tot gedetailleerde schoolwerkplannen met meetbare doelstellingen,</li> <li>• ongelimiteerde vrije schoolkeuze, oprichting van scholen door privébedrijven met commerciële belangen.</li> </ul> </li> </ul>

Tabel 1 Kenmerken technocratisch onderwijssysteem

Het onderwijs in bijvoorbeeld Engeland typeert Standaert als 'een bijna zuiver technocratisch onderwijs' (Standaert, 2014:32). Europese onderwijssystemen zijn in de optiek van Standaert de afgelopen jaren in meer of mindere mate onder de invloed van een meer technocratische onderwijsopvatting gekomen.

Bulterman en de Muynck spreken van twee paradigma's van waaruit onderwijs kan worden vormgegeven: het instrumentele paradigma of het vormingsparadigma (Bulterman & De Muynck, 2014: 34).

Instrumentele benadering	Vormingsbenadering
1) De student wordt gevormd door het onderwijs volgens een vastgesteld traject;	1) De student vormt zichzelf in een sociaal-culturele omgeving waar het onderwijs deel van uitmaakt. De uitkomsten staan niet vast;
2) Het onderwijs is gericht op toetsbare kennis en uiterlijk waarneembaar gedrag;	2) Het onderwijs betreft de hele mens en is gericht op het volle leven;
3) Het onderwijs is gericht op objectieve kennis;	3) Het onderwijs is gericht op kritisch denken en leren oordelen;
4) In het onderwijs wordt op basis van objectieve criteria geselecteerd en gedifferentieerd.	4) In het onderwijs wordt niet geselecteerd, vorming is voor allen.

Tabel 2 Onderwijsparadigma's

Het paradigma van waaruit wordt gedacht, geeft invulling aan onderwijs: aan onderwijs als stelsel maar ook aan de inhoud en aanpak van het onderwijs, het toezicht erop en de tevredenheid erover. In de praktijk komen 'zuivere' voorbeelden van ofwel het ene ofwel het andere paradigma niet gemakkelijk voor. En vaak is lastig te achterhalen welk paradigma overheerst, omdat beleid kenmerken van zowel het ene als het andere paradigma kan hebben. Ook kan het ene paradigma beleden worden terwijl in de praktijk het andere paradigma blijkt te overheersen. Uit de inrichting van een stelsel, de sturing van het onderwijs en het toezicht erop kan wel een beeld ontstaan welk paradigma leidend is of op welk punt van het continuüm van meer 'technocratisch' naar meer 'leerlinggericht' een onderwijssysteem zich bevindt of in welke richting het zich ontwikkelt.

### 3.2 Onderwijs: organisatie van een stelsel

De organisatie van een onderwijsselsel ontwikkelt zich in de loop van de tijd onder invloed van de specifieke context van een land. Daarmee worden onderwijsselsels al snel 'begrijpelijk' voor wie vertrouwd is met die specifieke context. Maar dit betekent niet dat er onderweg geen keuzes zijn gemaakt. Keuzes die door andere landen in andere (beleids)contexten anders zijn uitgevallen. Het is van belang zicht te hebben op de specifieke contexten waarbinnen landen zich ontwikkelen, zoals we in hoofdstuk 4 met name zullen zien. Tegelijkertijd is het ook relevant te weten op welke punten onderwijsselsels andere keuzes kunnen maken en hebben gemaakt. Ook dit leidt tot inzicht in de soms verscholen, soms openlijke doelen die met onderwijs in een bepaald land worden nagestreefd.

Het stellen dat onderwijsselsels van elkaar verschillen is eenvoudiger dan het daadwerkelijk vaststellen van verschillen. Dit omdat het vergelijken van onderwijsselsels lastig is en veel kennis

van de context vraagt van ieder onderwijssysteem dat is geworteld in de eigen cultuur en historie van een land. Daarbij komt dat wat aan de oppervlakte gelijk lijkt te zijn of hetzelfde klinkt, bij nader onderzoek toch verschilt en andersom. Voor het beschrijven van onderwijssystemen meer in het algemeen en Nederland in het bijzonder wordt hier gekozen voor het hanteren van het analysekader dat door Standaert is ontwikkeld en gebruikt voor zijn landenvergelijkend onderzoek naar de onderwijssystemen van Duitsland (deelstaat Hessen), Engeland, Frankrijk en Nederland (Standaert, 2007: 145-9). Standaert heeft onderwijssystemen vergeleken op basis van de drie elementen 'sturing', 'structuur' en 'kwaliteitsbewaking/kwaliteitszorg'.

### 3.2.1 Sturing

Sturing kan verlopen via een mix van middelen die in meer of mindere mate samenhangen met de benadering die men met het onderwijs voor ogen heeft. Een eerste sturingselement is de vorm van het curriculum. Zo kan een curriculum meer open of meer gesloten zijn. In een meer open stelsel wordt gewerkt met meer open doelen zoals het geval is in Nederland waar gekozen is voor het formuleren van zogenaamde 'streefdoelen'. Er zijn stelsels die kiezen voor een meer gesloten sturing, zoals bijvoorbeeld Vlaanderen waar is gekozen te werken met minimumdoelstellingen.

Een ander element van sturing is de inzet van toetsen en examens. Belangrijk daarbij is waar het eigenaarschap van toetsing ligt. De volgende vragen zijn daarbij relevant:

1. wordt de inzet van toetsing overgelaten aan de professional die toetsing inbedt in zijn eigen handelingsrepertoire;
2. gaat het om toetsing die van buitenaf aan het onderwijs wordt opgelegd;
3. gaat het om toetsen die extern gevalideerd (moeten) zijn;
4. worden de resultaten openbaar gemaakt?

Een belangrijk onderscheid is daarbij of de toetsing normgerelateerd of criteriumgerelateerd is vormgegeven. Bij normgerelateerde toetsing is het doel leerlingen, scholen of stelsels te vergelijken met elkaar want met elkaar bepalen zij de norm. Bij criteriumgerelateerde vormen van toetsing staat de vraag centraal of bepaalde doelen al dan niet behaald zijn.

Een derde sturingselement dat zorgt voor variatie in sturing is het toezicht. Hierbij gaat het bijvoorbeeld om het al dan niet hebben van een inspectie, de vorm en reikwijdte van het toezicht, het al dan niet gebruiken van (toets)instrumenten en de consequenties die al dan niet met de beoordeling gepaard gaan. Toezicht alleen kan al leiden tot tal van variaties in het sturen op inhoud en kwaliteit van onderwijs. Daarnaast zijn er landen die werken met centraal voorgeschreven leermiddelen en zo de inhoud en kwaliteit van onderwijs bepalen.

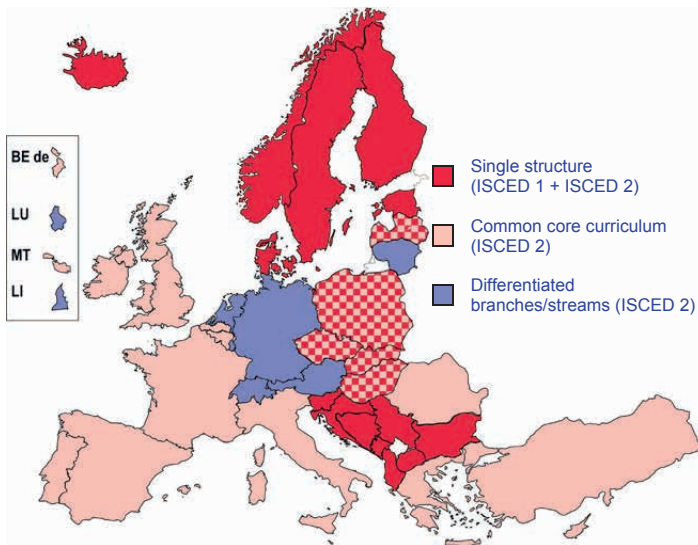
Met deze vormen van sturen op inhoud en kwaliteit van onderwijs zijn tal van combinaties mogelijk die leiden tot tal van variaties die gekozen kunnen worden om grip van overheden op onderwijsstelsels mogelijk te maken. Grip op onderwijs laat grote diversiteit zien tussen landen. Dit maakt het vergelijken van landen met elkaar lastig mogelijk. Dit feit moet ook leiden tot grote voorzichtigheid in het 'overnemen' van elementen van andere stelsels zonder het grote geheel goed in ogenschouw te nemen.

### 3.2.2 Structuur

Overgangsmomenten in een stelsel en de manier van selecteren op die overgangsmomenten wordt in veel beschrijvingen van stelsels als een belangrijk kenmerk gezien. Internationaal hanteert de Organisation for Economic Co-operation and Development (OECD) en ook de Raad van Europa dit als een belangrijk onderscheidend criterium voor de beschrijving van onderwijsstelsels. Ook in Nederland wordt een overgang op zich en de manier waarop de overgang plaatsvindt binnen het stelsel als een kenmerk van belang gezien (Nahuis & Verbruggen, 2009; Bronneman-Helmers, 2011; Van der Werfhorst, 2019).

De Raad voor Europa onderscheidt Europese onderwijsstelsels op basis van drie basisorganisatiemodellen. Het eerste model, 'Single structure education', is een model dat geen overgang kent tussen basis- en voortgezet onderwijs, waarin algemeen funderend onderwijs wordt gegeven aan alle kinderen. Het tweede model, 'Common core curriculum provision', is het model dat getypeerd wordt door een doorlopend, algemeen curriculum waarbij alle leerlingen aan het eind van het basisonderwijs overgaan naar het voortgezet onderwijs waar ze vervolgens allemaal een algemeen curriculum volgen. Het derde model, 'Differentiated lower secondary education' is het model waarbij leerlingen na het gezamenlijk volgen van het basisonderwijs met een gezamenlijk curriculum doorstromen naar verschillende vormen van voortgezet onderwijs, ofwel direct aan het begin van het vo ofwel in de loop van het vo<sup>3</sup>. De scheidslijnen zijn niet altijd even goed te trekken. Onderwijssystemen worden dan ook veelal getypeerd als 'meer' of 'minder' op een doorlopende schaal van 'integratieve vormen van onderwijs' (model 1 en 2) naar 'gesegregeerde vormen van onderwijs', gekenmerkt door een vroege

#### Main models of primary and lower secondary education (ISCED 1-2) in Europe, 2017/18



Figuur 1 Onderwijsmodellen Europa

3] European Commission/EACEA/Eurydice, 2017. The Structure of the European Education Systems 2017/18: Schematic Diagrams. Eurydice Facts and Figures. Luxembourg: Publications Office of the European Union.

selectie. Door de OECD wordt Nederland getypeerd als een onderwijsstelsel met relatief vroege selectie met relatief veel 'tracks' waarin leerlingen geplaatst kunnen worden (OECD, 2016: 22), namelijk 7 verschillende vormen van voortgezet onderwijs. Op de leeftijd van 12 jaar vindt de selectie daarvoor plaats. Soms worden leerlingen direct in een onderwijsvorm geplaatst en soms verloopt plaatsing via een brugklas die een combinatie vormt van meestal twee vormen van voortgezet onderwijs.

Binnen Europa kennen behalve Nederland ook Duitsland, Oostenrijk, Hongarije, Tsjechië en Zwitserland een vroege selectie. In Duitsland zelfs al op de leeftijd van 10 jaar. In Duitsland is in de meeste deelstaten het advies van de basisschool bepalend voor de plaatsing in het voortgezet onderwijs. Er wordt geen landelijk voorgeschreven gestandaardiseerde toets gebruikt. Het voortgezet onderwijs in Duitsland wordt voor een belangrijk deel op deelstaatniveau bepaald. Het onderwijs kent drie vormen van vervolgonderwijs: Hauptschule, Realschule en Gymnasium. Daarnaast is er combinatie van Hauptschule en Realschule in de vorm van een Gesamtschule als een aanbod van een brede scholengemeenschap naast het categoriale Gymnasium. Wie niet direct een advies krijgt voor het Gymnasium kan via bijvoorbeeld de Hauptschule of Gesamtschule alsnog naar alle vormen van vervolgonderwijs.

De meeste landen in Europa zijn in de vorige eeuw overgegaan naar meer integratieve vormen van voortgezet onderwijs. De noordelijke landen van Europa en meerdere Balkanlanden hebben gekozen voor het eerste model, een vergaande vorm van integratief onderwijs. De overige West-Europese landen hebben in meerderheid gekozen voor het tweede model. Het Verenigd Koninkrijk, dat tot in de jaren zeventig bekend stond als het stelsel met de meest stringente selectie op 11-jarige leeftijd met behulp van het zogenaamde 'elf-plus examen' is nu te typeren als een meer integratief onderwijsstelsel (model 2).

De keuze voor meer integratieve onderwijsstelsels (model 1 en 2) hangt in veel West-Europese landen samen met de democratiseringsgolf van de jaren zestig. Bronneman-Helmers stelt dat het streven naar gelijke kansen en een hoog onderwijsniveau voor zo veel mogelijk kinderen in deze landen een belangrijke overweging was om te kiezen voor een uniform aanbod voor alle jongeren (Bronneman-Helmers, 2011: 250). De Middenschool is te typeren als een poging om ook in Nederland een meer geïntegreerd onderwijssysteem te creëren.

### **3.2.3 Kwaliteitszorg/kwaliteitsbewaking**

Wat verstaan wordt onder 'kwaliteit' hangt in hoge mate samen met doel en functie van het onderwijs en aan wie het is om dat te bepalen. Volgens Standaert hangt kwaliteit van onderwijs samen met drie belangrijke aspecten: ten eerste met wat de doelen van onderwijs zijn die waardevol geacht worden, ten tweede met wat daarbij gezien worden als efficiënte strategieën (proces) om die doelen te bereiken en ten derde met wanneer er voldoende resultaat is in het realiseren van de beoogde doelen, ofwel de effectiviteit van het onderwijs (output) (Standaert, 2001: 13). Het eerste punt, de bewaking van de kwaliteit van onderwijs, levert een spanningsveld op, want pas als er eenduidig zicht is op wat de kwaliteit van onderwijs betekent, kan worden gezocht naar passende vormen van toezicht of kan bestaand toezicht op zijn waarde worden beoordeeld.

Kwaliteit is een meervoudig begrip. Standaert ordent discussies over kwaliteit van onderwijs op basis van een zevental tegenstellingen die hij beschouwt als continua (Standaert, 2014: 92):

1. Kwantiteit tegenover kwaliteit (te vergaderen kennis)
2. Specialisatie tegenover basisvorming
3. Cognitieve vorming tegenover brede vorming
4. Vakgericht onderwijs tegenover totaliteitsonderwijs
5. Theorie tegenover praktijk
6. Systematische opbouw tegenover exemplarische (volledigheid van een vak versus goede en representatieve voorbeelden)
7. Onmiddellijk resultaat tegenover blijvend resultaat<sup>4</sup>

Of 'kwaliteit' van onderwijs voor- of achteruit gaat, is in het licht van deze meervoudigheid een lastig te beantwoorden vraag. Bovendien is het een vraag die raakt aan de drie niveaus waarop onderwijs plaatsvindt: op microniveau (in de klas), op mesoniveau (op school) en op macroniveau (onderwijssysteem). Om kwaliteit te kunnen bepalen, zijn doelstellingen de spil van denken. Standaert spreekt in dit verband over een 'essentiële kwaliteitsdriehoek'. Daarmee bedoelt hij dat ten eerste doelstellingen voor kwaliteit moeten worden bepaald, dat vervolgens moet worden nagegaan of de beoogde kwaliteit wordt bereikt of nagestreefd en tenslotte moet er ondersteuning worden gegeven, organisatievormen en procedures worden ontworpen om die doelstellingen te bereiken. Voor toezicht op kwaliteit moet vanuit de drie niveaus worden nagedacht en gehandeld vanuit de opvattingen die er in een land zijn ten aanzien van kwaliteit. Voor toezicht op alle niveaus ligt een combinatie van kwantitatieve en kwalitatieve beoordelingen voor de hand. Zowel de keuze van kwantitatieve instrumenten zoals toetsen of indicatoren voor kwalitatieve beoordelingen als de concrete invulling ervan is in hoge mate afhankelijk van de doelstellingen die men met onderwijs voor ogen heeft en de kwaliteit die wordt nagestreefd. Omgekeerd geldt dat we aan de keuze voor concrete instrumenten en indicatoren kunnen afleiden wat de visie op onderwijs en kwaliteit is waar in een land op gestuurd wordt. De economische blik op onderwijs waarbij onderwijs wordt gezien als een investering in economische groei, heeft in het kielzog een afrekencultuur met allerlei metingen meegebracht waarbij het bruto nationaal product (bnp) geldt als indicator voor kwaliteit van het onderwijssysteem (Standaert, 2014: 87).

De Volder heeft in 2012 een vergelijkend onderzoek gepubliceerd naar externe schoolevaluaties in Europa. Zij onderzocht met behulp van participatieve observaties zeven landen<sup>5</sup> op de inhoudelijke, organisatorische en methodologische kenmerken van elk inspectiesysteem. Zij komt tot de conclusie dat ieder systeem in hoge mate uniek is en dat er geen sprake is van een eenduidige typologie op basis van de 40 kenmerken waarmee zij inspectietyperingen genereert (De Volder, 2012: 233). Op macroniveau lijkt er sprake te zijn van dezelfde algemene kenmerken maar bij observaties van het werk in de praktijk blijken er bij uitgevoerde schoolevaluaties op mesoniveau sprake te zijn van andere specifieke kenmerken. Zij komt tot een onderscheid in twee groepen systemen: werkend volgens het ontwikkelingsperspectief met als belangrijkste typering dat er geen openbaarheid van resultaten is en de mogelijkheid tot sanctionering ontbreekt en de groep werkend volgens het verantwoordingsperspectief dat gekenmerkt wordt door openbaarheid van resultaten en mogelijkheid tot sanctionering. Het Nederlandse inspectiesysteem plaatst zij in de groep van systemen die het verantwoordingsperspectief als uitgangspunt neemt (zie ook Hoofdstuk 5).

4] Met blijvend resultaat wordt bedoeld dat het gedrag verandert. Daar zijn twee voorwaarden voor: daar moet veel oefentijd voor zijn en daar moet motivatie voor zijn.

5] De landen Nederland, Zweden, Tsjechië en de regio's Noordrijn-Westfalen (Duitsland), Catalonië (Spanje), Wales (Verenigd Koninkrijk) en Vlaanderen (België).

### 3.3 Het primair onderwijs in Nederland

Het Primair Onderwijs kent in Nederland een eigen wet, eigen doelstellingen en eigen toezicht. In 2018 bezoeken ruim 1,5 miljoen kinderen het primair onderwijs, verdeeld over ruim 6700 scholen met gemiddeld 226 leerlingen per school, verdeeld over 8 jaarklassen<sup>6</sup>. De leerlingenaantallen lopen al sinds een aantal jaar langzaam terug. Voor de komende jaren is de verwachting dat zeker ruim 150.000 kinderen jaarlijks het primair onderwijs zullen verlaten en door zullen stromen naar een vorm van voortgezet onderwijs. Hoe in Nederland de overheid grip heeft op het primair onderwijs wordt aan de hand van de elementen sturing, structuur en kwaliteitszorg nader uitgewerkt.

#### 3.3.1 Sturing

In Artikel 8 van de Wet op het Primair Onderwijs (WPO) is het uitgangspunt en doelstelling van het primair onderwijs als volgt geformuleerd:

*'Het onderwijs wordt zodanig ingericht dat de leerlingen een ononderbroken ontwikkelingsproces kunnen doorlopen. Het wordt afgestemd op de voortgang in de ontwikkeling van de leerlingen.'*

Dit is een formulering waarin het te bereiken doel is geformuleerd zonder nadere instructies over wat het proces of het resultaat moet zijn. De formulering van dit artikel is een zogenaamd 'doelvoorschrift' volgens jurist Timmer. Artikel 8 WPO is volgens hem te kenmerken als een doelvoorschrift omdat 1) de nadruk ligt op de te bereiken uitkomst (het doel), 2) de normadressaat de vrijheid heeft om op zijn eigen wijze de uitkomst van het voorschrift te bereiken; bij hem ligt het initiatief voor de aanpak om tot de uitkomst te komen, en niet bij de regelsteller, en 3) er geen – of slechts in beperkte mate – specifieke handelingen worden voorgeschreven (Timmer, 2011: 27). Doelregelgeving kan een kwalitatieve of kwantitatieve doelstelling bevatten. De kwalitatieve doelstellingen omvatten zorgplichten met als uitkomst een meer open norm (Timmer, 2011: 26). In het geval van de WPO artikel 8 is volgens Timmer sprake van een inspanningsverplichting die een glijdende schaal van doelbereik impliceert (Timmer, 2011: 26). De toe te passen norm bestaat daarbij uit twee onderdelen: het inrichten van het onderwijs en het geven van het onderwijs, waaronder begrepen het afstemmen ervan op de voortgang in de ontwikkeling van leerlingen. In de WPO blijkt geen eenduidige lijn voor het benoemen van de normadressaten (Timmer, 2011: 162). Dat betekent volgens Timmer dat uit de WPO niet duidelijk wordt bij wie (op welk niveau) het inrichten van het onderwijs is belegd en op welk niveau het geven van onderwijs. In zijn onderzoek kiest Timmer ervoor de individuele school als uitvoerder van de norm te zien (Timmer, 2011: 164).

#### Doelstellingen

Sinds 2006 wordt in het primair onderwijs gewerkt met de derde generatie kerndoelen<sup>7</sup>. De kerndoelen zijn de belangrijkste landelijke leerplankaders in het primair onderwijs. Kerndoelen geven aan wat leerlingen in een bepaald deel van hun opleiding moeten kennen en kunnen.

6] <https://www.onderwijsincijfers.nl/kengetallen/po/leerlingen-po/aantallen-ontwikkeling-aantal-leerlingen>

7] In 1993 zijn de eindtermen primair onderwijs vervangen door 122 kerndoelen die minder in aantal en minder gedetailleerd zijn dan de eindtermen. Twee jaar later zijn de kerndoelen verder teruggebracht tot 103 en in 2006 tot 58. Op dit moment wordt er gewerkt aan een nieuw curriculum.



Het zijn streefdoelen met veel ruimte voor eigen invulling. Er zijn 58 streefdoelen en ze bestrijken grotere leergebieden: de talen, rekenen/wiskunde, oriëntatie op jezelf en de wereld, kunstzinnige oriëntatie en bewegingsonderwijs. De Stichting Leerplan Ontwikkeling (SLO) heeft de afgelopen jaren ook tussendoelen en leerlijnen ontwikkeld. Zowel de tekst van de wet als de uitwerking in leerdoelen is te beschouwen als uitwerking van een open sturing op de inhoud van het onderwijs. Dit past bij de uitgangspunten van de Grondwet voor onderwijs waarin de vrijheid op richting, stichting en inrichting van het onderwijs is vastgelegd (artikel 23.) Sinds augustus 2010 zijn voor taal en rekenen naast de streefdoelen de 'referentieniveaus' verplicht. Dat zijn beheersingsdoelen die aangeven wat leerlingen moeten kennen en kunnen als het gaat om Nederlandse taal en rekenen/wiskunde op de verschillende overgangsmomenten in het Nederlandse onderwijsstelsel. Er wordt onderscheid gemaakt tussen fundamentele niveaus en streefniveaus. De referentieniveaus zijn bedoeld om over sectoren heen een 'samenhangend, efficiënt en effectief curriculum' te bieden (Expertgroep, 2008). De doelen zijn in een doorlopende lijn po-vo-mbo beschreven. Daarmee heeft het onderwijs in Nederland een combinatie van een meer gesloten curriculum in de vorm van 'minimumdoelen' zoals vastgelegd termen van referentieniveaus voor taal en rekenen en een meer open curriculum voor de andere onderdelen van het curriculum in de vorm van 'streefdoelen' zoals beschreven in de derde generatie kerndoelen.

### **Autonomie**

De autonomie van de scholen is in de afgelopen decennia een belangrijk punt van aandacht geweest in het onderwijsbeleid van de overheid. De brede dereguleringstendens van de overheid, ingezet in de jaren tachtig (zie ook hoofdstuk 4), manifesteert zich in het onderwijs onder het motto van 'autonomievergroting'. Scholen krijgen meer ruimte en autonomie om op die manier de aandacht te kunnen richten op vooral het verdelen van onderwijs en onderwijskansen. Achterliggende gedachte is dat kwaliteit van onderwijs gebaat zou zijn bij marktwerking; kwaliteit van onderwijs verbetert omdat concurrentie slecht onderwijs afstraft en zo een verhoging van het onderwijspeil stimuleert (Waslander, 1999: 7). Tegenover die juridische vrijheid die er feitelijk altijd al was en autonomie die in de praktijk vooral betrekking had op de ruimte meer zelfstandig de eigen financiën te beheren ('Lumpsum bekostiging') is meer sturing gekomen. De invoering van minimumstandaarden in de vorm van referentieniveaus zijn gepaard gegaan met de invoering van gestandaardiseerde, externe toetsen. Daarmee is een belangrijk aspect van regie, namelijk die over de inhoud van het onderwijs, weggehaald bij de scholen. Sinds het schooljaar 2013-2014 is het verplicht om alle kinderen aan het eind van het basisonderwijs deel te laten nemen aan een eindtoets basisonderwijs. Daarnaast is bij de laatste wetswijziging een verplichting tot het gebruik van een leerlingvolgsysteem<sup>8</sup>. (LVS) voor het volgen van de ontwikkeling van taal en rekenen opgenomen. Met het gebruik van een LVS kan tussentijds op bepaalde momenten tijdens het primair onderwijs de stand van zaken per kind en per school worden vastgesteld. De gegevens van de eindtoets zijn openbaar en worden gebruikt voor verantwoording en toezicht.

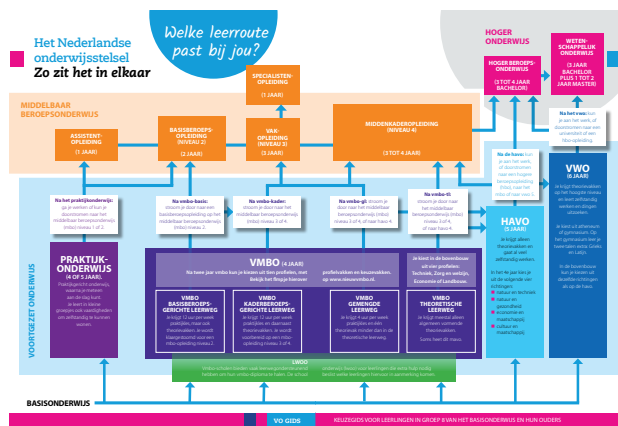
8] In Nederland geldt een verplichting de ontwikkeling van leerlingen systematisch te volgen met behulp van een LeerlingVolgSysteem (LVS). Op de markt zijn verschillende systemen van een LVS beschikbaar. Cito is tot op heden de grootste leverancier van een LVS. Hun LVS wordt door hen aangeduid als LOVS.

De afgelopen decennia is de sturing van het basisonderwijs in Nederland veranderd. In lijn met de grondwettelijke vrijheid van onderwijs was er aanvankelijk veel vrijheid en ruimte voor scholen om zelf inhoud en vorm te geven aan het onderwijs. Dit heeft door de invoering en toetsing van standaarden plaats gemaakt voor meer sturing van de kant van de overheid op zowel inhoud als resultaten.

### 3.3.2 Structuur funderend onderwijs

Er is in Nederland geen integratief onderwijsstelsel gekomen, niet in de vorm van de middenschool en ook niet in de vorm van basisvorming. Het funderend onderwijs kent een onderscheid in primair onderwijs en voortgezet onderwijs. Daarmee heeft Nederland een selectief onderwijssysteem (model 3) en een relatief vroege selectie. Bronneman-Helmers stelt dat de vroege selectie wordt gecombineerd met twee andere structuurkenmerken die de werking van de selectie versterken, namelijk een strikte scheiding tussen het algemeen voortgezet onderwijs en het beroepsonderwijs en een sterke gelaagdheid naar verschillende niveaus binnen de beide typen onderwijs (Bronneman-Helmers: 2011: 249). De overstapmogelijkheden tussen beroepsonderwijs en algemeen voortgezet onderwijs zijn beperkt, evenals de overstapmogelijkheden tussen de verschillende vormen van algemeen voortgezet onderwijs.

Aan het eind van het basisonderwijs wordt bepaald welk type vervolgonderwijs de leerling gaat volgen. Dit gebeurt op basis van het advies dat de leerling krijgt. Dit advies is gebaseerd op het oordeel van de school, aangevuld met de uitkomst van de afgelegde eindtoets die formeel gepositioneerd is als een ‘second opinion’. Leerlingen van het basisonderwijs en het speciaal onderwijs zijn sinds de wijziging van de WPO in 2011 verplicht om aan het eind van hun basisschoolperiode een eindtoets basisonderwijs af te leggen. Tezamen met het schooladvies vormt de uitkomst van de toets de basis voor doorstroom naar de volgende fase van het funderend onderwijs: het voortgezet onderwijs. Binnen het voortgezet onderwijs worden zeven varianten onderscheiden: praktijkonderwijs, vmbo in 4 varianten (twee beroepsgerichte, een gemengde en een algemene), havo en vwo. Leerlingen kunnen rechtstreeks instromen in een variant maar het merendeel van de leerlingen stroomt door naar een brugklas die een combinatie vormt van varianten, veelal twee. Het gebruik van de term ‘schoolkeuze’ suggereert dat ouders en kinderen de vrijheid hebben hun vervolgonderwijs te kiezen. Maar in de praktijk is die keuze zeer beperkt (Bronneman-Helmers, 2011: 249). Het schooladvies vormt de inkadering van de keuzevrijheid: gegeven het schooladvies, mede gebaseerd op het eindtoetsadvies, kunnen ouders en kinderen op zoek naar een school die de geadviseerde variant van voortgezet onderwijs aanbiedt. Daarmee is het advies niet vrijblijvend en de keuze van ouders en hun kinderen niet onbeperkt: de keuzevrijheid is zo breed als het gegeven advies. In het geval er een school is die



Figuur 2 Het Nederlandse onderwijsstelsel (uit keuzegids VO)

een 'brede' brugklas aanbiedt die de geadviseerde onderwijsvariant aanbiedt met nog een variant, kunnen ouders en kinderen het moment van selectie soms nog een jaar uitstellen maar ook dan binnen de bandbreedte van de mogelijkheden van die brugklas.

### 3.3.3 Kwaliteitszorg en kwaliteitsbewaking

In Nederland is de minister van onderwijs verantwoordelijk voor het waarborgen van de drie publieke belangen van onderwijs. Dat betekent dat de minister de kwaliteit van het onderwijs bevordert, zorgt voor de toegankelijkheid ervan en de doelmatige besteding van publieke middelen bewaakt (Onderwijsraad, 2013: 12). Het is aan de overheid om een goede balans te (blijven) bewaken tussen de drie publieke belangen die onder druk van economie en maatschappij de neiging hebben te veel de ene of de andere kant op te gaan. De autonomie van scholen brengt in eerste instantie met zich mee dat scholen zelf verantwoordelijk zijn voor de uitvoering van het onderwijs. De Inspectie van het Onderwijs richt zich met het toezicht, conform de Grondwet art. 23, op het rapporteren over de 'deugdelijkheid' van het stelsel. De kwaliteitseisen en maatstaven die de inspectie is gaan gebruiken bij het toezicht op schoolniveau hebben pas een wettelijke basis gekregen na de invoering van de Wet op het onderwijstoezicht (WOT) in 2002 (Bronneman-Helmers, 2011: 97). Daarmee is het toezicht op 'deugdelijkheid' zoals de Grondwet voorschrijft, verschoven naar een nieuwe interpretatie, namelijk toezicht op 'kwaliteit'. Kwaliteit als centrale notie in onderwijsbeleid en wetgeving heeft een ander gezicht gegeven aan toezicht. Het toezien op kwaliteit impliceert een meer inhoudelijke opvatting over wat goed onderwijs is (Peeters et al., 2013: 11). En een inhoudelijke opvatting over kwaliteit geeft ruimte voor normstelling op basis waarvan beoordeeld kan worden of er wel of geen sprake is van (voldoende) kwaliteit. De vraag is dan wel wie bepaalt wat kwaliteit is. In Nederland stelt de inspectie zelf het eigen inspectiekader vast en de minister keurt het (formeel) goed. Voor het primair onderwijs hanteert de Inspectie een apart toezichtkader. Toezichtkaders van de Inspectie zijn volgens onderwijsjurist Timmer formeel geen regels voor de scholen maar zijn bedoeld als een instructie voor de toezichthoudende inspectieteams; zij zijn daarmee dus formeel niet normstellend voor de instellingen. De Inspectie stelt hiermee geen eisen aan scholen, maar expliciteert in haar toezichtkaders de eigen werkwijze en stelt die vast. Timmer wijst erop dat scholen dit wat hij noemt 'subtiele onderscheid tussen toezichtkader en algemeen verbindende voorschriften' waarschijnlijk zal ontgaan. Dit komt volgens hem omdat de bepalingen in het toezichtkader gericht lijken te zijn op de onderwijsgevendenden en de school. En het feit dat het toezichtkader de Inspectie ertoe kan brengen een school te kwalificeren als een zwakke of zeer zwakke school en dit ook openbaar te maken, versterkt de indruk dat er wel degelijk sprake lijkt te zijn van een norm die met een sanctie wordt gehandhaafd (Timmer, 2011: 179). Voormalig hoofdinspecteur Primair Onderwijs Jonk zegt hierover in 2014: 'De inspectie baseert haar risico-analyse en oordeel over opbrengsten mede op de resultaten van leerlingen. Wij doen dat om onze (beperkte) capaciteit effectief in te zetten. Je zou kunnen zeggen dat wij de resultaten gebruiken om heel veel scholen niet te hoeven bezoeken.' (Jonk, 2014). In het 'Onderzoekskader 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs' (geldig vanaf 1 augustus 2020) stelt de Inspectie van het Onderwijs: 'De wet geeft aan dat scholen voldoende leerresultaten behoren te behalen (artikel 10a WPO). Er is sprake van voldoende leerresultaten als de resultaten op of boven de normering liggen (die is afgestemd op de samenstelling van de leerlingenpopulatie van scholen), zoals vastgelegd in de Regeling leerresultaten PO.' (Inspectie van het Onderwijs, 2016: 20).

Jarenlang gold de gemiddelde score op de eindtoets (gecombineerde score taal en rekenen) als uitgangspunt voor de kwaliteitsbeoordeling. Sinds de laatste verandering van het toezichtskader is dat nu het percentage leerlingen dat een bepaald referentieniveau behaalt. In beide gevallen gaat het om een beoordeling van de kwaliteit van de school op basis van leerlingprestaties, jaarlijks per school vast te stellen. Scholen zijn daarmee voor hun kwaliteitsoordeel mede afhankelijk van variabelen die samenhangen met de populatie die de eindtoets in een bepaald jaar maakt: de grootte van de groep en de samenstelling ervan. Omdat zich hier tussen scholen en tussen cohorten de nodige verschillen kunnen voordoen, is besloten een correctie toe te passen om de 'eerlijkheid' van de vergelijking te vergroten. In eerste instantie betrof het een correctie op leerlingniveau: elke leerling kreeg op basis van een aantal variabelen, vooral sociaaleconomische, een zogenaamd 'gewicht' mee. De gemiddelde score op de eindtoets werd 'gecorrigeerd' voor het 'leerlinggewicht': hoe lager het gemiddelde leerlinggewicht hoe hoger de gemiddelde score zou moeten zijn. Sinds kort heeft de inspectie gekozen voor een nieuw, aangepast model in de vorm van een 'schoolweging'. Het Centraal Bureau voor de Statistiek (CBS) berekent jaarlijks de schoolweging gebaseerd op drie voorgaande schooljaren<sup>9</sup>. In de schoolweging worden de potentiële scores van individuele leerlingen gecorrigeerd voor hun achtergrondgegevens die weer net anders zijn dan de indicatoren die van invloed waren op het eerdere 'leerlinggewicht'. Voor scholen is de schoolweging niet zelf te berekenen en ook niet te controleren.

Het CBS heeft onderzocht welke kenmerken het risico op een onderwijsachterstand van een leerling het best voorspellen. Dit zijn:

1. het opleidingsniveau van de ouders: van zowel de vader als de moeder wordt het opleidingsniveau meegenomen.
2. het gemiddelde opleidingsniveau van alle moeders op school.
3. het land van herkomst van de ouders: er wordt gekeken naar de herkomst van de ouders. De herkomstlanden zijn ingedeeld in een aantal categorieën op basis van de 'culturele afstand' tot Nederland.
4. verblijfsduur in Nederland: er wordt gekeken naar de verblijfsduur van de moeder in Nederland.
5. schuldsanering: er wordt gekeken of het gezin in de schuldsanering zit.

Het CBS heeft toegang tot deze gegevens en kan op basis hiervan de verwachte onderwijsscore van elke leerling berekenen. De schoolweging wordt bepaald op basis van de verwachte onderwijsscores van alle leerlingen op de school op peildatum 1 oktober. De inspectie rondt de scores lager dan 20 af naar 20 en hoger dan 40 naar 40. Hoe hoger de schoolweging hoe complexer de leerlingenpopulatie en hoe lager de verwachte leerresultaten voor deze school.

Voor meer informatie over de berekeningswijze van de schoolweging verwijzen we naar het CBS-bestand

<https://www.cbs.nl/nl-nl/maatwerk/2020/13/schoolweging-2017-2019>

### Referentieniveaus

Vanaf het schooljaar 2010-2011 zijn referentieniveaus voor taal en rekenen verplicht ingevoerd (OCW: 2010). In de wet op de referentieniveaus zijn voor taal en rekenen minimumnormen voor leerresultaten geformuleerd per schoolsoort. Hiermee is voor taal en rekenen een doorlopende leerlijn geformuleerd<sup>10</sup>. Hieraan is een bekostigingsvoorwaarde gekoppeld. De opbrengstgerichte benadering voor taal en rekenen is daarmee bij wet vastgelegd (Timmer, 2011: 223). Zo is er sinds 2015 een nieuwe, dubbele opdracht voor de eindtoets. Enerzijds is het de bedoeling dat de toets de populatie van eindtoetsleerlingen zodanig verdeelt dat er voldoende spreiding in scores, en dus moeilijkheidsgraad, in de toets is ingebouwd om selectie te kunnen realiseren. Anderzijds moeten er niveaus van beheersing worden vastgesteld waarbij het van belang is dat de opgaven

9] <https://www.onderwijsinspectie.nl/documenten/rapporten/2020/05/04/schoolweging-2017-2018-2019-2020>

10] <https://www.slo.nl/thema/meer/taal-rekenen/taal/>

niet verschillen in niveau maar juist voldoende opgaven van eenzelfde niveau bevatten om een goede beoordeling van de beheersing van de gevraagde niveaus mogelijk te maken. Vanaf het schooljaar 2020-2021 geldt een nieuw onder-



Figuur 3 Referentieniveaus en onderwijs

wijsresultatenmodel dat aansluit op deze verandering. De inspectie baseert het kwaliteitsoordeel op de schoolresultaten van de laatste drie jaar samen behaald op de beheersing van de referentieniveaus: 'Met het nieuwe model leidt een lage score op de eindtoets niet meer automatisch tot het oordeel 'onvoldoende'. Er volgt eerst nader onderzoek. Het oordeel kan alsnog positief worden als de school zich adequaat kan verantwoorden op basis van de leerlingenpopulatie.' (Ministerie OCW, 2019c). De beheersing van de referentieniveaus als uitgangspunt is wezenlijk anders dan de vergelijking en beoordeling van scholen gebaseerd op relatieve scores, uitgaand van de gemiddelde score. Met een dergelijk model is het lastig voortgang aan te tonen: de resultaten zijn immers van elkaar afhankelijk. Het besluit de standaardcores niet langer als uitgangspunt te nemen voor een kwaliteitsoordeel maar te werken met meer absolute normen, zoals de referentieniveaus, is een verandering van kwaliteitsbeoordeling op basis van relatief meten en beoordelen naar criteriumgerelateerd meten en beoordelen van kwaliteit van onderwijs op basis van opbrengsten, waarbij leerlingen wel of niet een vooraf bepaalde norm halen.

De uitkomsten op basis van de eerste jaren meten van de referentieniveaus laten zien dat meer leerlingen de beoogde referentieniveaus halen dan eerder gedacht (zie ook hoofdstuk 8). Dit is aanleiding geweest voor het bijstellen van ambities zoals blijkt uit de brief van de minister aan de Tweede Kamer: 'In plaats van een focus op risico's ligt nu de focus op ambities en verbetering van het onderwijs over de brede linie. De inspectie streeft er niet alleen meer naar om onvoldoende scholen op te sporen en die te ondersteunen bij het weer voldoende worden. De inspectie wil juist ook dat scholen die het al goed doen het nog beter gaan doen. Het inspectietoezicht staat in het teken van effectief toezicht voor beter onderwijs.' (Ministerie OCW, 2019c). Dit streven is als uitgangspunt nu opgenomen in het toezichtskader dat vanaf het schooljaar 2020-2021 ingaat; scholen dienen ernaar te streven dat zoveel mogelijk leerlingen 1F beheersen voordat zij de overstap naar het voortgezet onderwijs maken. De aanpassingen in het toezichtskader hebben niet de bedoeling de druk op prestaties bij scholen te verminderen; scholen en besturen moeten de resultaten gebruiken om te benchmarken ten opzichte van vergelijkbare scholen en besturen, want 'dit helpt bij het stimuleren van ambitie' (Ministerie OCW, 2019c). Daarmee is de overgang van relatief meten en normeren naar meer absoluut, criteriumgerelateerd, meten en beoordelen in een ander licht komen te staan; er wordt beoordeeld op basis van criteriumgerelateerde standaarden in de vorm van referentieniveaus, maar in de toepassing is alsnog vergelijkend meten en beoordelen de praktijk in het inspectietoezicht.

### **Schooladvies**

Scholen geven aan het eind van het basisonderwijs aan iedere leerling een schooladvies. Aanvankelijk waren er daarbij geen restricties voor de scholen voor het opstellen van het schooladvies. De eindtoets werd gepositioneerd als een 'second opinion' naast en na het schooladvies. Dit onafhankelijke karakter van het schooladvies staat onder druk. Scholen worden geacht hun schooladvies 'goed' vorm te geven. En onder 'goed' wordt verstaan: in lijn met de uitkomsten van de eindtoets. Driessen spreekt in dit kader over 'Over- en onderadvisering bij de overgang van het voortgezet onderwijs' waarbij blijkt dat 'correcte advisering' advisering is die aansluit bij de capaciteiten zoals gemeten via de eindtoets (Driessen, 2011: 3). Scholen moeten zich verantwoorden als ze met hun advies 'afwijken van het advies uit de eindtoets': bij onderschatting ten opzichte van de eindtoets worden scholen geacht hun advies naar boven bij te stellen, bij overschatting of zelfs het voordeel van de twijfel geven moeten ze zich verantwoorden ten opzichte van de Inspectie van het Onderwijs. Dat aan de uitkomst van de toets grote waarde wordt toegekend is af te leiden uit de brief van de staatssecretaris aan de Kamer: 'Hieruit blijkt dat het percentage leerlingen waarvoor het toetsadvies exact overeenkomt met het schooladvies gelijk is gebleven. Daarnaast blijkt dat in 2016 voor bijna driekwart van de leerlingen de afwijking een halve schoolsoort of minder bedroeg. Dit is bijvoorbeeld een leerling met vmbo-kl-advies die op de toets vmbo-bl/kl haalt.' (OCW, 2016). Hoewel de toets formeel bedoeld is als 'second opinion' zoals de staatssecretaris in dezelfde brief nogmaals benadrukt, blijkt dat de toets wel degelijk de norm is waaraan de staatssecretaris de kwaliteit van de schooladviezen en daarmee de scholen afmeet. Want als scholen hun advies niet bijstellen kan dat niet de goedkeuring krijgen van de staatssecretaris zoals blijkt uit dezelfde brief: 'Dit is zeer ongewenst en ook in strijd met de wet. Het bevoegd gezag heeft immers de wettelijke plicht om per leerling de afweging te maken of het advies bij een hoger toetsresultaat wordt bijgesteld. De inspectie zal het gesprek aangaan met de schoolbesturen waar niet of nauwelijks tot bijstelling wordt overgegaan, terwijl hier gezien de toetsresultaten wel alle redenen toe bestaat. Het doel hiervan is om hen te wijzen op het belang van heroverweging en te achterhalen welke afwegingen daarbij een rol spelen. Daarnaast zal de inspectie via haar communicatiekanalen alle besturen en scholen wijzen op de verplichting om bij hogere toetsresultaten het advies te heroverwegen.' (OCW, 2016).

Het positioneren van de toets als een 'second opinion' en het advies van de school als leidend blijkt met deze kaders in de praktijk beperkt feitelijk uitvoerbaar en houdbaar. Omdat de uitkomsten van de eindtoets een nauwe relatie hebben met de uitkomsten van het LVS van Cito (zie ook hoofdstuk 5), blijken scholen nu de uitkomsten van het LVS mee te wegen in hun schooladvies. Uit onderzoek van de Universiteit Twente en Oberon blijkt 98% van de scholen de resultaten van het LVS mee te nemen in het schooladvies (Luyten et al.: 2019: 25). Dit wordt ook nadrukkelijk geadviseerd door onder andere de PO-raad, de belangenbehartiger van de scholen voor Primair Onderwijs. De PO-raad stelt in de Handreiking Schooladvies uit 2015, verwijzend naar een rapport van het ITS hierover, dat toetsresultaten van het Cito-leerlingvolgsysteem 'een prominente rol spelen bij het opstellen van het basisschooladvies. Rekenen/wiskunde en begrijpend lezen wegen daarbij vaak het zwaarst. En dat is in lijn met de inhoud van de eindtoets. Sommige scholen baseren het advies op de scores van groep 6 tot en met 8, terwijl andere scholen zich beperken tot groep 7 en 8.' (PO-raad, 2015:2). De raad voegt daaraan toe op basis van onderzoek uitgevoerd door onderzoeksbureau Oberon dat 'de resultaten bij rekenen/wiskunde en begrijpend lezen de beste voorspellers van het succes zijn in het voortgezet onderwijs' (PO-raad,

2015:2). Gezien de hoge correlaties tussen scores op het LVS van Cito en scores op de eindtoets, wordt die laatste score het best 'voorspeld' op basis van de uitkomsten van het LVS. Ook de Inspectie van het Onderwijs hecht waarde aan het meewegen van de resultaten van toetsen uit het LVS bij het opstellen van het schooladvies zoals blijkt uit de feedback die leerkrachten krijgen na het spelen van een 'serious game' waarbij ze op basis van een profiel van henzelf en een profiel van een klas kunnen oefenen met het geven van schooladviezen (zie kader)<sup>11</sup>.

Op deze manier is formeel wel sprake van een toets als 'second opinion' zoals het wetsvoorstel voor ogen had, maar wordt er met de verplichting van het gebruik van een LVS nu een first opinion gegenereerd die heel dicht tegen de second opinion aan zal liggen. Daarmee wordt het aantal variabelen waarop het advies, zowel school- als toetsadvies, berust, ingeperkt tot voornamelijk de scores op gestandaardiseerde toetsen voor taal en rekenen, waarbij taal zich vooral beperkt tot begrijpend lezen. Voor de

brede kijk van de school op een kind die zou moeten worden meegenomen in het schooladvies, blijft zo weinig ruimte over. Daarmee ligt er voor de onderwijsinhoud een gerichtheid op de cognitieve elementen taal en rekenen zoals die met LVS en eindtoets worden gemeten, op de loer. Dit wordt versterkt door een beoordeling van kwaliteit van scholen op basis van een vergelijking van scholen met elkaar wat de gerichtheid op dat wat in de toets aan bod komt, verder vergroot.

Het Nederlandse onderwijs is met deze kenmerken naar sturing, structuur en kwaliteitszorg en kwaliteitsbeoordeling op dit moment te typeren als een meer technocratisch onderwijssysteem in de terminologie van Standaert. Het belang van de eindtoets bij de uitwerking van de sturing, de structuur en de kwaliteit van onderwijs is de afgelopen decennia enorm toegenomen.

### 3.4 Eindtoets als instrument van sturing, structuur en kwaliteit

De aanpassing van de WPO in 2011 heeft voor een aantal belangrijke wijzigingen gezorgd. Belangrijk is het besluit om de eindtoets ná het schooladvies te laten plaatsvinden. De bedoeling hiervan is de druk op de eindtoets (voor ouders, leerlingen en scholen) tegen te gaan. Omdat het schooladvies aan de toetsuitslag voorafgaat, kan het belang van de toets afnemen en daarmee de druk op alle betrokkenen. Daarnaast is vastgelegd dat het gebruik van een eindtoets verplicht is evenals als het gebruik van een LVS. Het College voor Examens (CvE), al verantwoordelijk voor de eindexamens voortgezet onderwijs, is vanaf de wijziging van de wet verantwoordelijk voor de eindtoets en verandert van naam in het College voor Toetsen en Examens (CvTE). Cito wordt belast met de ontwikkeling van de eindtoets die vanaf dat moment Centrale Eindtoets (CET) wordt genoemd en onder toezicht van het CvTE wordt ontwikkeld en afgenomen. Hiermee is de overheid verantwoordelijk voor de eindtoets en daarmee voor de beide functies die aan de eind-

#### FEEDBACK

##### Leerresultaten

U geeft aan dat u **de leerresultaten** van een leerling meeweegt in het advies. Uw advies lijkt ook meestal te worden beïnvloed door de leerresultaten, maar niet altijd.

Dat geeft te denken. Want u weet waarschijnlijk dat leerresultaten (en vooral van begrijpend lezen en rekenen) een belangrijke voorspeller zijn van schoolsucces in het voortgezet onderwijs. U moet goed inzicht hebben op de leerresultaten én de leerontwikkeling die een leerling doormaakte over een langere periode. Anders kunt u geen goede inschatting maken van de mogelijkheden van een leerling in het vervolgonderwijs!

11] <https://www.onderwijsinspectie.nl/onderwerpen/overgang/interactieve-simulatie-basischooladvies>

toets zijn toebedeeld: selectie en kwaliteitsbewaking. We werken de functies van de eindtoets en vervolgens het LVS eerst nader uit alvorens de rol van de eindtoets nader te beschouwen in relatie tot sturing, structuur en kwaliteit van het primair onderwijs.

### 3.4.1 De eindtoets nader beschouwd

In de verantwoording van het CvTE uit 2015 wordt de eindtoets, in lijn met de periode daarvoor, gekarakteriseerd als een 'Leervorderingentoets', wat volgens het CvTE betekent dat het een toets is die 'meet wat een kind in vergelijking met andere kinderen in acht jaar basisonderwijs geleerd heeft'. Op basis van wat een kind heeft geleerd in vergelijking met de andere kinderen in hetzelfde cohort, wordt dan vervolgens bepaald wie welk advies krijgt: 'Daarmee verschaft de toets onafhankelijke en objectieve informatie voor de keuze van een passend brugklastype. Want leervorderingen in het basisonderwijs zeggen iets over de vraag hoe goed een leerling de verschillende typen van voortgezet onderwijs aankan.' (CvTE, 2015). Dit onderbouwt het CvTE verder door te stellen:

*'Het lijkt gerechtvaardigd te veronderstellen dat leerprestaties uit het verleden iets zeggen over leerprestaties in de toekomst. Verschillen in leerprestaties op de onderdelen van de CE (Centrale Eindtoets) kunnen vermoedelijk voor een belangrijk deel worden verklaard door verschillen in eigenschappen als 'intelligentie', 'concentratie', 'motivatie' en 'doorzettingsvermogen'. Via de CE worden deze eigenschappen indirect en afgewogen gemeten; elke leerling heeft een 'persoonlijke mix'.'*

Deze tekst correspondeert met de laatste verantwoording die Cito zelf in 2011 heeft opgesteld. In die verantwoording wordt hieraan nog toegevoegd: 'Zo bezien zou de eindtoets wel eens veelzijdiger kunnen zijn dan op grond van oppervlakkige waarneming vaak gedacht wordt.' (Van Boxtel et al., 2011: 10). Het uitgangspunt van de eindtoets is dat resultaten uit het verleden een voldoende garantie lijken te bieden voor het voorspellen van kansen op toekomstig schoolsucces. Over welke resultaten het meer precies gaat en hoe de garanties worden ontleend aan scores, wordt hieronder nader uitgewerkt.

#### **Inhoud**

In de verantwoording van de Eindtoets Basisonderwijs uit 2011 staat dat 'de verdeling van de 200 opgaven (100 Taal, 60 Rekenen en 40 Studievaardigheid) over de verschillende kennisgebieden sinds jaar en dag vast ligt'. Dit blijkt bij nader onderzoek niet helemaal te kloppen. In de verantwoordingen van 1984 en 1991 rapporteert Cito een gelijke verdeling van opgaven over de drie onderdelen die samen leiden tot een standaardscore: taal, rekenen en informatieverwerking (60 opgaven per deelgebied). In 2011 rapporteert Cito een verdeling van 100 opgaven voor taal, 60 voor rekenen en 40 voor informatieverwerking dat dan 'studievaardigheid' wordt genoemd. Na de overname van de eindtoets door de overheid rapporteert het CvTE in de verantwoording van 2015 dat het onderdeel taal vanaf dat moment uit 135 opgaven bestaat en rekenen uit 85 opgaven. Het onderdeel Studievaardigheid blijkt sinds 2015 geen apart onderdeel meer te zijn van de Centrale Eindtoets. De Centrale Eindtoets van 2019 bestaat volgens de Terugblik Centrale Eindtoets 2019 uit 105 opgaven voor taal (45 opgaven lezen, 40 opgaven taalverzorging en 20 opgaven schrijven) en 60 opgaven voor rekenen. Binnen de onderdelen Taal en Rekenen hebben zich dus diverse wijzigingen voorgedaan, in aantallen items, in de relatieve verdeling van



de items over de onderdelen en ook inhoudelijk met het verdwijnen van het onderdeel Studietoetsvaardigheid. Het onderdeel Algemene Kennis is nooit vervallen maar heeft wel een andere naam gekregen: Wereldoriëntatie heet het onderdeel nu. Dit onderdeel is altijd facultatief geweest en heeft tot op heden nooit meegeteld in de eindscore. Al met al zijn de wijzigingen in de optiek van het CvTE zodanig dat het CvTE in de Terugblik Centrale Eindtoets 2015 vermeldt 'dat de inhoudelijke wijzigingen de vergelijking tussen de Centrale Eindtoets 2015 en de Cito-toets bemoeilijken' (CvTE, 2015:12).

### Ordenen

De centrale eindtoets werkt met 'percentielscores'. Dit betekent dat leerlingen op basis van scores geordend worden. Een individu dat een percentielscore van 70% heeft, kan dat lezen als dat 30% van de populatie hoger heeft gescoord dan hij en 69% lager. Op basis van zijn score krijgt elk kind een plek in de rangorde per toetsonderdeel. De scores op de verschillende toetsonderdelen worden vervolgens bij elkaar opgeteld tot een gecombineerde score. Voor de huidige eindtoets betekent dit dat de scores voor taal (lezen, taalverzorging en schrijven) en rekenen (getallen, verhoudingen, meten en meetkunde en verbanden) bij elkaar worden opgeteld om te komen tot een totaalscore.

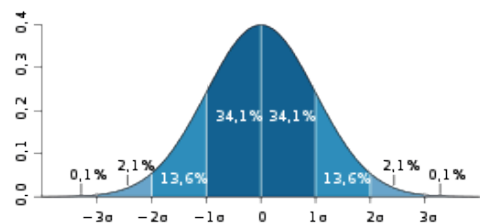
Vanaf 1977 is Cito, naast een rapportage in percentielscores, overgegaan op het hanteren van zogenaamde 'standaardscores'. Dit is gebeurd omdat er beperkingen zijn aan rekenkundige bewerkingen die met ruwe scores kunnen worden uitgevoerd. Een 'oplossing' daarvoor is volgens Drenth en Sijtsma de omzetting van percentielscores naar standaardscores; 'de hoofdbezwaren tegen percentiele normen [...] worden opgeheven in de categorie waarbij de bewerkte scores zijn uitgedrukt in standaardscore-eenheden' (Drenth & Sijtsma, 1990:91). Het werken met standaardscores maakt het mogelijk om scores verkregen met toetsen die jaarlijks kunnen verschillen in moeilijkheidsgraad, volgens eenzelfde schaal te rapporteren. Cito heeft daarbij gekozen voor een 'schaal met een gemiddelde van 535 en een standaarddeviatie van 10. De range van de schaal werd bepaald op 501-550. Deze wat ongewone schaal werd gekozen om iedere verwar- ring met schoolcijfers, intelligentiequotiënten en dergelijke te voorkomen.' (Wijnstra, 1984: 13). Cito voert een vorm van equivaleren in die het mogelijk maakt dat 'de scores op toetsversies uit verschillende jaren op eenzelfde schaal gebracht kunnen worden' (Wijnstra, 1984: 13). De equivaleringsprocedures zorgen ervoor dat de 'gemiddelde score en variantie daarin van jaar tot jaar vrijwel exact vergelijkbaar blijven' (Van Boxtel et al., 2011: 67). Het equivaleren van de scores gebeurt op basis van een lineaire transformatie. Dit is volgens Cito psychometrisch verantwoord onder de volgende voorwaarde:

*'We gaan uit van de veronderstelling dat er in de loop van een gering aantal jaren geen dramatische veranderingen plaatsvinden in het landelijke algemene niveau van de leer- vorderingen.... Aannemende dat het gemiddelde niveau in een groep constant is geble- ven kunnen de scores van twee eindtoetsen lineair met elkaar worden geëquivalereerd.'*  
(Van Boxtel et al., 2011:18).

Deze werkwijze is nog steeds de aanpak, blijkend uit de verantwoording van CvTE uit 2015: 'De totaalscore wordt via lineaire transformatie omgezet in een standaardscore. Dit houdt in dat de ruwe score wordt vermenigvuldigd met een getal (A) en dat er daarna een vast getal (B) bij wordt

opgeteld. De waarden van A en B worden jaarlijks bepaald. [...] De resulterende standaardscore heeft het karakteristieke gemiddelde van omstreeks 535. De verdeling heeft de linksscheve vorm.' (CvTE, 2015: 36). Om leerlingen te kunnen toewijzen aan verschillende categorieën is spreiding in scores gewenst en is het nodig variatie in moeilijkheidsgraad te realiseren en wel zodanig dat de spreiding rond het gemiddelde groot genoeg is. Vanaf het begin van de ontwikkeling van de eindtoetsen tot en met de eindtoets van 2019 is een hoge gemiddelde p-waarde niet het streven bij de (centrale) eindtoets. Heel expliciet stellen de verschillende verantwoordingen dat het streven is een goede afwisseling te hebben tussen moeilijke en makkelijke opgaven, passend bij het differentiërend karakter van de toets: 'p-waarden mogen spreiding vertonen van .45 à .50 - .90, met dien verstande dat de gemiddelde p-waarde uitkomt tussen .70 en .75 (een enkele p-waarde > .90 is toegestaan) (Wijnstra, 1984: 22). Ook voor de ontwikkeling van de huidige Centrale Eindtoets geldt dit uitgangspunt.' (CvTE, 2015: 35). De Groot en Van Naersen stellen in 1969 dat de te beogen (gemiddelde) moeilijkheidsgraad van een toets nauw samenhangt met het doel van de toets: 'Gaaf het om de vraag of de (minimum)doelstellingen met de hele groep bereikt zijn, dan hoopt men op een p-waarde<sup>12</sup> dicht bij 1; gaat het om de differentiatie tussen proefpersonen/leerlingen, dan is zo'n hoge p-waarde ongewenst.' (De Groot & Van Naerssen, 1969: 44). Bij de invoering van het vaststellen van referentieniveaus vanaf 2011 zou een vorm van meten passend bij het willen kunnen vaststellen van minimum-doelstellingen kunnen worden verwacht. Dit is niet gebeurd. Het psychometrisch ontwerpschema heeft geen veranderingen ondergaan en het differentiëren tussen leerlingen is het uitgangspunt gebleven.

Het ordenen van de leerlingen op basis van hun gecombineerde score van hoog naar laag scorend, op een toets die gericht is op differentiatie, leidt tot een Gaussische verdeling, ofwel een normaalverdeling. De meeste leerlingen zitten rond het gemiddelde, de 0 (535) en de overige leerlingen worden op basis van hun 'afwijkingen' van het gemiddelde verdeeld over de linker- en rechterkant van de klokvormige curve. Standaert noemt dit model 'bij uitstek een selectiemeting' (Standaert, 2014: 125). Het model maakt het mogelijk om 'de besten' te selecteren, zoals bijvoorbeeld de beste 20%.



Figuur 4 Gaussische verdeling

### Classificatie

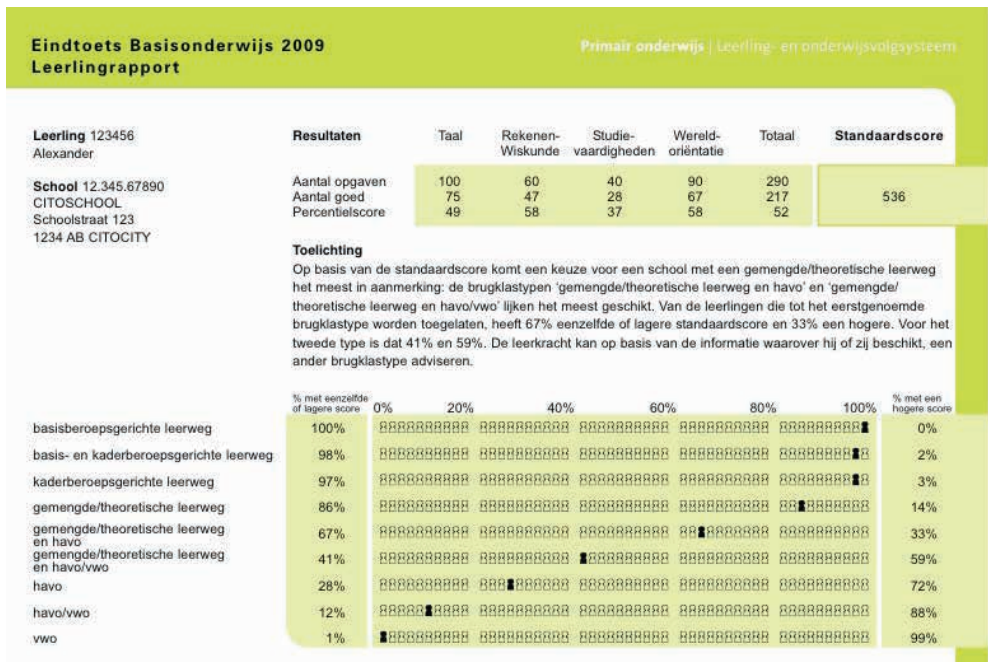
In de tijd van de ontwikkeling van de eerste versie van de eindtoets, de Amsterdamse Schooltoetsen, ligt in Amsterdam het percentage leerlingen dat naar het v.h.m.o.<sup>13</sup> doorstroomt, op een ongeveer een vijfde deel van de totale leerlingpopulatie: 'ongeveer 20% van de l.o.-6-populatie gaat naar het v.h.m.o.' (Samenwerkende Instituten, 1967: 136). Het werken met percentielscores,

12] P-waarde is de moeilijkheidsgraad van een opgave. Het staat voor het percentage leerlingen dat een vraag goed beantwoord heeft. Een hoge p-waarde geeft aan dat veel leerlingen de opgave goed hebben gemaakt en een lage p-waarde het omgekeerde.

13] Het v.h.m.o. of V.H.M.O. staat voor het voorbereidend hoger en middelbaar onderwijs. Dit is de vorm van onderwijs dat tot aan de invoering van de Mammoetwet de hogere vormen van middelbaar onderwijs omvat, waaronder gymnasium, lyceum, hbs, mms en handelsdagschool (deze laatste werd later omgezet tot hbs-a).

geordend in vijf categorieën van 20%, sluit aan bij de bestaande situatie in Amsterdam. Uit de verantwoording van de eerste Amsterdamse Schooltoetsen blijkt dat het streven is de uitkomsten van de toets in lijn te brengen met het gegeven schooloordeel dat als referentiepunt voor de validiteit als geschikt wordt gezien (zie ook hoofdstuk 4). Daarvoor zou de hoogst scorende 20% een op een overeen moeten komen met degenen die een v.h.m.o.-schooladvies hebben gekregen.

Voor wat betreft de selectie zien de eerste makers van de eindtoets vooral mogelijkheden door te kijken naar leervorderingen op basis van het verleden. Zij zien beperkingen ten aanzien van het doen van voorspellingen over toekomstig schoolsucces. Niet vanwege beperkingen van de toets zelf maar vooral vanwege de onbetrouwbaarheid van de cijfers die leerlingen krijgen in het voortgezet onderwijs op basis waarvan een juistheid van voorspelling dan zou moeten worden onderbouwd. Ook Cito is aanvankelijk aarzelend over het doen van voorspellingen ten aanzien van toekomstig schoolsucces. Uit de verantwoording van 1992 blijkt dat Cito er in de jaren tachtig voor kiest niet de leerling te volgen in zijn 'succes' in het voortgezet onderwijs, afgemeten aan schoolprestaties maar de leerlingen die op een bepaalde plek in het voortgezet onderwijs zitten als uitgangspunt te nemen voor een analyse van scores die zij destijds op de eindtoets hebben behaald. Zo hoeft er geen correlatie aangetoond te worden tussen de score op de eindtoets en prestaties in het voortgezet onderwijs. Er ontstaat een beeld van de scores waarmee leerlingen in mavo 3, of in vwo 3 zijn doorgestroomd. De aldus verkregen spreiding in scores voor de verschillende groepen van het voortgezet onderwijs, vormt dan het uitgangspunt voor de advisering van nieuwe leerlingen: 'Met de gegevens ... kan men een indruk krijgen of een leerling met de behaalde standardscore waarschijnlijk tot de betere, zwakkere dan wel de middenmoot zal behoren in een bepaald schooltype'. (Uiterwijk & Theunissen, 1996: 12). De score van een nieuw individueel kind wordt vergeleken met de populatie die al een bepaalde positie heeft weten te verwerven



Figuur 5 'Poppetjes-grafiek'

in het voortgezet onderwijs zodat een kind zich een beeld kan vormen met welke scores hij zich omringd zal weten bij plaatsing in een bepaalde brugklas. Daartoe gaat Cito gebruik maken van een zogenaamde 'poppetjes-grafiek' (Uiterwijk & Theunissen, 1996: 12). In de verantwoording van 2011 wordt ook een overzicht gegeven waaruit het gemiddelde en de spreiding van standaardcores per brugklatype blijkt: 'Per brugklatype is de gemiddelde score aangegeven en er wordt door middel van balkjes zichtbaar gemaakt waar de grenzen liggen van de volgende drie standaardcoregebieden (in termen van percentages): 5-24%, 25-75%, 76-95%': elk brugklatype krijgt een 'eigen' score-range (Van Boxtel et al., 2011: 56). Dit wordt overgenomen bij de overgang van de basiseindtoets van Cito naar CvTE. In de verantwoording van de Centrale Eindtoets<sup>14</sup> blijkt 'de poppetjesgrafiek' verdwenen en blijft de presentatie van verschillende brugklatypes met hun bijbehorende score-gebieden over. Daarmee wordt de score van een kind niet langer vergeleken met de scores van alle andere kinderen, geordend per brugklatype, maar zijn de vastgestelde intervallen per brugklatype de norm geworden waartegen de score van een kind wordt afgezet. Eerder kon nog gedacht worden in termen van 'advies' aan een kind en zijn ouders om het kind voor te bereiden op een situatie voor de toekomst, namelijk een inschatting van de populatie waarmee het kind zich bij een bepaalde keuze omringd zal zien in termen van scores op taal, rekenen en voorheen ook nog studievervaardigheid, gecombineerd te zien als een 'specifieke vorm van intelligentie' zoals Cito in de handleiding vermeldt. Met de komst van normtabellen ontstaat de suggestie dat een score gezien kan worden als een niveau van intelligentie dat 'hoort bij' of 'vereist wordt door' een bepaald schooltype.

De huidige normtabellen zijn tot stand gekomen op basis van 'toelatings- en doorstroomonderzoeken' die de afgelopen decennia zijn uitgevoerd door Cito in samenwerking met het CBS. Hierin zijn leerlingen gevolgd wat betreft hun plaatsing in het voortgezet onderwijs. 'Het onderzoek maakt het mogelijk een relatie te leggen tussen de standaardcore van een leerling en het school- of brugklatype waarop de leerling de grootste kans op succes heeft.' (Van Boxtel et al., 2011:18). Uit de gerapporteerde spreiding van de scores komt aanvankelijk een brede range per brugklatype naar voren en is er sprake van behoorlijke overlap in scores tussen de brugklatypes, deels verklaarbaar doordat er combinatiebrugklassen zijn maar zichtbaar is ook dat de ranges van scores niet alleen daardoor te verklaren zijn: leerlingen in de verschillende brugklassen blijken in die brugklassen geplaatst met grote, met andere brugklatypes overlappende verschillen in scores. Dit blijkt ook uit onderzoek dat Cito en CBS doen: 'Ruim zes op de zeven leerlingen gaan naar een brugklatype dat de Cito-toets aangaf. Vervolgens haalt het grootste deel van de leerlingen ook een diploma op het niveau van het geadviseerde schooltype. Maar leerlingen die gestart zijn in een brugklas met een hoger of lager schooltype halen ook vaker een diploma op dat hogere of lagere niveau.' (Hakkenes & de Wijs, 2012).

Op basis van onderzoek uit 2015 concludeert de Inspectie van het Onderwijs dat na 3 jaar voortgezet onderwijs er de nodige verschuivingen ten aanzien van de oorspronkelijke plaatsing zijn: 'Hoewel men ziet dat de meeste leerlingen op de door het advies geïndiceerde positie terechtkomen, ziet men tegelijk ook dat relatief veel leerlingen op een andere positie uitkomen.' (Inspectie van het Onderwijs, 2015: 47). Zo blijkt dat na drie jaar voortgezet onderwijs nog ongeveer 60% van de leerlingen in een onderwijstype zit waarvan het geadviseerde niveau deel uitmaakt,

14| <https://www.centraleeindtoetspo.nl/actueel/nieuws/2018/04/10/meer-weten-over-leerlingrapport-centrale-eindtoets-2018>

10% van de leerlingen is gedoubleerd, en op- en afstroom komt samen bij ongeveer 15% van de leerlingen voor (ongeveer evenveel af- als opstroom ten opzichte van het schooladvies) (Van Rooijen et al., 2017: 110, Inspectie van het Onderwijs, 2014, Inspectie van het Onderwijs 2016). De Inspectie wijst op het feit dat uit onderwijsonderzoek bekend is dat het hebben van hoge verwachtingen positief is voor de ontwikkeling van kinderen. Een hoog advies kan gezien worden als een vorm van het hebben van hoge verwachtingen. Dit betekent dat een hoger advies in het voordeel van leerlingen kan uitpakken. De Inspectie van het Onderwijs refereert hierbij aan de bevindingen van onderzoek van Timmermans, Kuypers en Van der Werf uit 2012 die stellen ‘dat een lager advies dan de toets indiceert, onvoordelig voor de leerling uitwerkt. Een hoger advies gegeven de toets werkt juist voordelig voor de leerling uit.’ (Inspectie van het Onderwijs, 2015: 49). Uit onderzoek van GION uit 2013 blijkt dat ‘de belangrijke conclusie is dat gemiddeld (of ‘per saldo’) onder advisering ongunstig en over advisering gunstig is voor een zo hoog mogelijke positie op de leerjarenladder. Het gaat daarbij om vrij grote tot grote verschillen. Men zou kunnen zeggen dat het advies ten dele het karakter heeft van een ‘selffulfilling prophecy’. Dit resultaat is vergelijkbaar met dat van Kuypers en van der Werf (2007) en dat van Kuypers en van der Werf (2012).’ (Timmermans et al., 2012:8).

In onderstaande tabel wordt zichtbaar hoe de ontwikkeling is verlopen van de rapportage van de gemiddelde score, de verschillende percentielgroepen en de ‘waargenomen range’, gebaseerd op onderzoek naar de instroom in leerjaar 1 in het schooljaar 2005-2006 bij leerlingen in verschillende brugklassen, naar score-intervallen zoals die vanaf 2011 gerapporteerd zijn door eerst Cito, daarna door het CvTE en tenslotte zijn opgenomen als addendum bij de wet. Is er in eerste instantie sprake van een rapportage van de range in scores die zich feitelijk voordeed in de verschillende brugklassen, vanaf 2011 gaat het niet meer om een rapportage van de range maar wordt gewerkt met score-intervallen die in de loop van de tijd zodanig worden ‘verfijnd’ dat van een overlap in scores tussen de verschillende onderwijsvormen geen sprake meer is. In de tabel is met een gele markerings aangegeven waar de getallen gewijzigd zijn ten opzichte van daarvoor. Zo wordt inzichtelijk dat de intervallen in de loop van de tijd de nodige aanpassingen hebben gekend die er uiteindelijk in hebben geresulteerd dat er nu eenduidige, elkaar niet meer

**Ontwikkeling van score-gebieden naar score-intervallen**

Brugklatype	Gemiddelde score	5 t/m 24 % (20%)	25- 75% (50%)	76- 95% (20%)	Waargenomen range	Score-interval 2011 Cito	Score-interval 2015 CVTE (CE 2014)	Score-interval 2016 CVTE (CE 2015)	Score-interval CVTL (2017)	Tekst staatscourant 25 juni 2018
Basisberoeps	518	508-513	514-522	523-528	508-528	501-520	501-520	501-520	501-518	
Basis- en kaderberoepsgerichte leerweg	522	511-517	518-527	528-533	511-533	519-525	519-525	519-525	519-525	505-524
Kaderberoepsgerichte leerweg	526	517-522	523-528	529-535	517-535	524-528	528-528	523-528	526-528	525-532
Gemengde/theoretische leerweg	531	522-527	528-534	535-540	522-540	529-533	529-533	529-533	529-532	-
Gemengde/theoretische leerweg en havo	534	526-530	531-538	539-542	526-542	532-536	538-536	535-536	535-536	535-539
Gemengde/theoretische leerweg en havo/vwo	538	527-533	534-542	543-549	527-549	535-541	-	-	-	-
Havo	539	531-535	536-541	542-546	531-546	537-540	537-540	537-540	537-538	-
Havo/vwo	542	534-538	539-544	545-549	534-549	540-545	540-544	540-544	540-544	540-544
Vwo	547	541-544	545-548	549-550	541-550	545-550	545-550	545-550	545-550	545-550

Tabel 3 Ontwikkeling van score-gebieden naar score-intervallen

overlappende adviescategorieën zijn overgebleven die gelden als 'harde' grenzen voor het al dan niet adviseren van een bepaalde vorm van onderwijs, bij wet vastgelegd.

De wetenschap dat grote verschillen in scores kunnen leiden tot eenzelfde diploma en omgekeerd, heeft niet kunnen verhinderen dat de score-intervallen steeds meer een determinerend karakter hebben gekregen. Overlap in scores per brugklasinterval, zoals jarenlang af te lezen was uit de 'poppetjesgrafieken' waar Cito mee werkte, heeft plaatsgemaakt voor een 'nette' opvolging van scores waarbij geen sprake meer is van overlap, twijfel of verwarring. Op deze manier is een 'advies' in termen van het geven van een beeld van wat een *kind* kan verwachten,

In de tabel is per brugklatype aangegeven wat het standaardscore-interval is dat noch op onderschatting noch op overschatting van de mogelijkheden van leerlingen lijkt te wijzen. Uit de tabel kan worden opgemaakt dat een leerling met de standaardscore 530 het toetsadvies best passend brugklatype 'vmbo kaderberoepsgerichte leerweg en vmbo gemengde/theoretische leerweg' krijgt.

**Tabel, per brugklatype het interval dat het uitgangspunt is voor de interpretatie van de standaardscore in het leerlingrapport centrale eindtoets.**

501-504	praktijkonderwijs en vmbo basisberoepsgerichte leerweg
505-524	vmbo basis- en kaderberoepsgerichte leerweg
525-532	vmbo kaderberoepsgerichte en vmbo gemengde/theoretische leerweg
533-539	vmbo gemengde/theoretische leerweg en havo
540-544	havo en vwo
545-550	vwo

*Staatscourant 2018, 39602 | Overheid.nl > Officiële bekendmakingen (officielebekendmakingen.nl)*

vervangen door een oordeel over een geschiktheid van een kind voor een schoolsoort in termen van intelligentie en daarmee een oordeel met gevolgen voor de plaatsing. De score-intervallen toegekend aan de brugklatypes die uiteindelijk zijn opgenomen in de Staatscourant geven het standaardscore-interval aan dat 'noch op onderschatting noch op overschatting van de mogelijkheden van leerlingen lijkt te wijzen' zoals de tekst luidt die als regeling in juli 2018 in de Staatscourant wordt gepubliceerd. (Staatscourant, 2018). De zogenaamde 'conversietabellen' hebben de afgelopen jaren aanpassingen ondergaan, 'zodat elke toetsscore nu leidt tot een eenduidig schooladvies', waarbij de twijfel, uitgedrukt in overlap in scoregebieden, heeft plaatsgemaakt voor elkaar uitsluitende categorieën: 'Deze wijziging zorgt voor meer helderheid doordat er bij bepaalde toetsscores niet langer sprake is van overlap in toetsadviezen.' (Staatscourant, 2018).

Hoewel de scores van kinderen in de verschillende vormen van onderwijs altijd de nodige overlap hebben gekend, zoals de jaarlijkse poppetjesgrafieken lieten zien, zal het nu een kwestie van tijd zijn of de huidige normtabellen in combinatie met toezicht op de strikte, verplichte toepassing ervan in combinatie met het verplichte gebruik van de uitkomsten van het LVS, zullen alle variatie die er nu nog is, eerst laten vervagen en dan geheel doen verdwijnen. Niet omdat de score een goede voorspeller is, maar omdat bewijs van het tegendeel gaat ontbreken: er zullen steeds minder kinderen zijn met een 'afwijkende' score die het goed blijken te doen in een onder-

wijstype dat volgens die score niet het meest passend is geweest. Het uitbannen van afwijkende scores uit de diverse schooltypes maakt de voorspelling vanzelf 'waar'.

### ***Van score naar allocatie***

Door de eindtoetsen vooraf te laten gaan door uitgebreide pretests met ankering en equivale- ring kunnen de uiteindelijke uitkomsten van de eindtoets in voorspelbare banen worden geleid. Door na de afname de data eerst te verzamelen en vervolgens de uitslagen te bepalen, kan voor eventuele afwijkingen in de definitieve afname alsnog worden gecorrigeerd. De beoogde gemiddelde moeilijkheidsgraad van tussen de .70 en .75 kan zo worden gerealiseerd, net als de spreiding van de moeilijkheidsgraad van de afzonderlijke items tussen de .40 en .90. Van alle opgaven die niet aan de criteria voldoen of interfereren met de beoogde opbouw van de toets als geheel, kan nog worden besloten ze niet mee te laten tellen in het totaal. Door de gemiddelde score over de 'bruikbare' set vervolgens te koppelen aan de standardscore van ongeveer 535, kan gezorgd worden voor een gecontroleerde spreiding van leerlingen over de totale range met een standaarddeviatie tussen 9 en 10, opgedeeld in score-gebieden met een bewuste, links- scheve verdeling: 'Het differentiërend vermogen en de nauwkeurigheid zijn daardoor [bedoeld wordt 'linksscheef'] in de lagere en gemiddelde regionen optimaal.' (Van Boxtel et al., 2011: 92). Het werken met een zo te sturen 'normaalverdeling', die bewust enigszins linksscheef uitpakt, zorgt ervoor dat de best scorende 20% zich met de meeste 'psychometrische zekerheid' aan het rechter uiteinde van de verdeling bevindt. Dit past bij de oorspronkelijke opzet van de toets: het selecteren van de beste 20%. Hoewel ook in deze situatie de zekerheid dat echt de besten zijn geselecteerd niet zonder twijfel is, al was het maar omdat het maar de vraag is of een test dat wel kan, is zorg over de 'juistheid' van de verdeling over de andere categorieën groter. Het is statistisch gezien lastiger om met zekerheid onderscheid te maken tussen leerlingen in de top van de grafiek. Rond het gemiddelde bevindt zich het grootste deel van de leerlingen en is het discriminerend vermogen van de toets het meest beperkt en de meetfout het grootst. Terwijl daar wel een, in de optiek van de samenleving, zeer beslissend onderscheid wordt gemaakt, tussen havo/vwo enerzijds en vmbo anderzijds. Onderwijspedagoog Terwel stelt dat 'De selectiepraktijk in Nederland is gebaseerd op een fundamenteel verkeerd idee over de verdeling van begaafd- heden bij leerlingen: wie niet theoretisch is moet wel praktisch zijn.' (Terwel, 2006: 11). Terwijl wie laag scoort op de Cito-toets, en dus afgewezen wordt voor havo-vwo, nog niet noodzake- lijkerwijs geschikt is voor het vmbo. Deze assumptie is volgens Terwel eerder een afgeleide van het bestaande schoolsysteem dan dat het gebaseerd is op een logische redenering of op feitelijk aanwezige talenten van leerlingen. 'Als dergelijke categorieën eenmaal geïnstitutionaliseerd zijn in 'leerwegen' of schooltypen bevestigen zij zichzelf: de hokjes moeten wel gevuld worden' zo stelt Terwel (Terwel, 2006: 11).

Het werken met opgaven die bewust verschillen in moeilijkheidsgraad, een gemiddelde moeilij- kheidsgraad kennen van .70 en het omzetten naar standardscores uitgaand van ongeveer 535 als punt op de schaal dat overeenkomt met de gemiddelde score, kan reguleren dat ongeveer de helft van de leerlingen een advies krijgt voor havo/vwo en de andere helft een vmbo-advies. En nog altijd krijgt, net als in 1965, ongeveer 20% een vwo-advies (destijds v.h.m.o.-advies). Het gebruik van normtabellen in combinatie met toezicht van de inspectie op het geven van 'afwijkende' schooladviezen wordt zo gezien een vorm van 'self fulfilling prophecy': resultaten van leerlingen uit het verleden, gebaseerd op vergelijkende toetsscores die in de loop van de

tijd anders zijn opgebouwd omdat de toetsen inhoudelijk wel degelijk zijn veranderd, zijn op deze manier bepalend geworden voor de mogelijkheden die nieuwe generaties leerlingen krijgen. Zo neemt een leerling niet alleen zijn eigen resultaten uit het verleden mee naar de toekomst maar ook de resultaten die andere leerlingen, jaren voor hem op inhoudelijk andere toetsen en onder andere omstandigheden, hebben behaald.

De tabellen suggereren 'harde' grenzen tussen de onderscheiden schoolsoorten waar kinderen gegeven hun score wel of niet in 'thuishoren'. Maar van thuisoren of passen is geen sprake. Niet de eventuele geschiktheid of ongeschiktheid van een leerling voor een bepaalde vorm van onderwijs wordt vastgesteld. Het is een psychometrisch geconstrueerde verdeelsleutel die bepaalt wie in welk 'hokje' terecht komt, op basis van de ordening die met de eindtoets gegenereerd en gecreëerd wordt. Deze aanpak blijkt ook op Curaçao te zijn doorgevoerd, zij het dat daar de percentages voor toelating tot havo/vwo lager liggen dan in Nederland.

De nadruk in de wetenschap om evidence-based te werken, of iets genuanceerder, evidence-informed, zoals in onderwijskundige kringen veelal aangeraden wordt, maakt hier plaats voor een nieuwe manier van werken: evidence-transformed. Met behulp van een rekenkundige formule worden de leerlingen verdeeld. Die verdeling wordt echter gezien en gevoeld als een geschiktheidsoordeel. De uitkomst van de eindtoets wordt door de leerling zelf en zijn omgeving gepercipieerd als een vaststelling van intelligentie in combinatie met inzet en motivatie met ver strekkende gevolgen richting toekomstig functioneren in de samenleving en op de arbeidsmarkt: een leerling gaat zich gedragen naar zijn beoordeling. Door nu de toewijzing strikt te reguleren en te onderzoeken vanuit die opgelegde verdeling, zal op termijn blijken dat de eindtoets een zeer goede 'voorspeller' blijkt: afwijkingen zijn bij voorbaat geëlimineerd en kunnen het tegendeel niet langer bewijzen.

### **Antilliaanse race om havo/vwo**

*De score op de eindtoets op Curaçao is een samengesteld cijfer: men middelt het cijfer voor Nederlands en voor Papiaments. Het cijfer dat hieruit komt wordt weer gemiddeld met het cijfer voor rekenen.*

*Edgar de Palm, rector Maria Immaculata Lyceum: 'Lange tijd stroomde elk jaar ongeveer 18% van alle Curaçaose leerlingen door naar een havo/vwo-school. Dat percentage was eenvoudigweg gebaseerd op het aantal beschikbare plekken. Het aanbod voor havo/vwo-plekken wordt groter. Daardoor kan ook het aantal leerlingen dat een havo-advies of hoger krijgt, stijgen. Er wordt momenteel een percentage gehanteerd van 22%. Het vaststellen van de verschillende cesuren gebeurt onder verantwoordelijkheid van een commissie, samengesteld door de gezamenlijke schoolbesturen. Zij geeft aan hoe hoog het percentage havisten en vwo'ers dat jaar kan zijn. Is dat 20%, dan krijgt 20% van de (beste) leerlingen voor de eindtoets een 8 of hoger, en daarmee een havo-/vwo-advies.' De Palm beaamt dat door deze werkwijze het gemiddelde niveau van de leerlingen het ene jaar wat anders uitvalt dan het andere jaar: 'Soms hebben we een zwakke lichting, zoals we dat noemen, en soms hebben we relatief veel sterke leerlingen. Het gevolg van deze werkwijze is dat leerlingen die wel de capaciteiten hebben niet per definitie op havo of vwo terecht komen. Aan de scores en percentages kunnen we dat dus niet zien.'*

*Uit: Toets!4, 2015*

### **Zicht op ontwikkeling**

Door leerlingen in hun eigen cohort jaarlijks met elkaar te vergelijken, te ordenen en middels



percentielscores gelijkmatig te verdelen en die verdeling te koppelen aan score-intervallen is het lastig door de jaren heen een constant beeld te hebben van de ontwikkeling van de taal- en rekenvaardigheid zoals die met de eindtoets wordt gemeten. De keuze voor een relatieve norm waarbij leerlingen met elkaar vergeleken en gerangschikt worden op een eenvoudige en inzichtelijke schaal, maakt een oordeel over de ontwikkeling van de taal- en rekenvaardigheid beperkt van waarde. Het bepalen van de stand van het onderwijs in termen van kwaliteit, al dan niet vertaald in het wel of niet behalen van bepaalde doelstellingen, verhoudt zich lastig tot een relatieve norm. Om met behulp van een relatieve meting nog enigszins iets te kunnen zeggen over longitudinale ontwikkelingen is het van belang dat de groepen die jaarlijks deelnemen vergelijkbaar zijn. Cito zegt hierover in 2011: 'Omdat het niet zeker is of beide steekproeven in de twee opeenvolgende jaren goed (of even goed) de Nederlandse schoolbevolking vertegenwoordigen, kunnen we ook geen harde uitspraken doen over een daling of stijging van de gemiddelde vaardigheid die door de Eindtoets wordt gemeten.' (Van Boxtel et al., 2011: 40). Deze beperking die samenhangt met vergelijkend meten, heeft er onder andere toe geleid dat nu *alle* basisschoolleerlingen verplicht aan de eindtoets moeten deelnemen. Hiermee wordt aan het steekproefprobleem tegemoet gekomen.<sup>15</sup> Tegelijkertijd is besloten meer absolute leerstandaarden in de eindtoets op te nemen in de vorm van de referentieniveaus. De vergelijking van scholen onderling en de resultaten longitudinaal kunnen dan gebaseerd worden op de uitkomsten van de scores op de referentieniveaus. Dit lijkt een logische stap, gegeven de beperkingen die relatief meten heeft voor een beeld op ontwikkeling door de jaren heen. De keuze voor de integratie van een meer absolute vorm van meten in een instrument met een relatief karakter brengt echter weer andere, nieuwe beperkingen met zich mee. De referentieniveaus suggereren een volgorde van opbouw van taal- en rekenvaardigheid conform de beschreven opbouw in niveaus. Uit tal van onderzoek is echter duidelijk geworden dat leren, ook van taal en rekenen, veelal grillig verloopt en eerder cyclisch, zoals Smits beschrijft aan de hand van de leercirkels van Kolb (1984), Zull (2004) en Gee (2017) dan volgens vaste leerlijnen (Smits, 2020). Daarnaast bieden de niveaus ruimte voor interpretatie. Dit bleek bijvoorbeeld toen ook andere toetsaanbieders referentieniveaus in hun eindtoetsen hadden opgenomen. Dit heeft de vergelijkbaarheid van de eindtoetsen belemmerd. Met behulp van uniforme ankersets<sup>16</sup>, wordt nu geprobeerd dit probleem aan te pakken. Een derde probleem is de dubbele functie in relatie tot de aantallen items: met minder opgaven moeten meerdere doelen worden gediend. De eindtoets moet zowel een passend selectieadvies geven én tegelijkertijd een kloppend oordeel over het behaalde referentieniveau, waarvan per vaardigheid twee referentieniveaus worden onderscheiden. Zo kan bij Rekenen beheersing van het hoogste niveau worden aangetoond bij een score van 80% correct op de totale itemset en diezelfde itemset moet ook onderscheidend genoeg zijn voor het kunnen verantwoorden, in combinatie met de taalscore, van zeven cesuurpunten die samenhangen met de acht verschillende schooladviezen die gegeven moeten worden.

De invoering van Referentieniveaus Taal en Rekenen als uitgangspunt voor de toetsing zou inhoudelijk hebben kunnen zorgen voor een beeld op de langere termijn over de ontwikkeling

15] De mogelijkheid dat ook andere testaanbieders een eindtoets mogen aanbieden, is aanvankelijk niet voorzien.

16] Dit zijn identieke opgaven die door alle toetsaanbieders verplicht in de eindtoetsen moeten worden opgenomen, waarvan de psychometrische waarden op voorhand bekend zijn en die als basis dienen voor de normering zodat vergelijkingen tussen de verschillende eindtoetsen mogelijk zijn.

van taal- en rekenvaardigheid. Dit is in de uitvoering nu zo ingewikkeld geworden dat het maar zeer de vraag is of het toevoegen van een leerstofonafhankelijk criterium in de vorm van referentieniveaus in de huidige constellatie hiervoor kan zorgen. Bovendien blijft aan de eindtoets een tweede doel gekoppeld, namelijk het doel van selectie. Met een beperkte set opgaven per deelvaardigheid moet nu een oordeel gegeven worden over het al dan niet beheersen van referentieniveau 1F, referentieniveau 2F/1S én de potentiële geschiktheid voor een van de acht verschillende schooladviezen. Het leunen op de uitkomsten van het LVS, zodat niet één moment zoals met de eindtoets (in politieke debatten regelmatig metaforisch aangeduid als 'foto'), maar een langduriger observatie zoals het LVS (meer een 'film' in deze politieke metafoor), lijkt daarom aantrekkelijk.

### 3.4.2 Leerlingvolgsysteem en eindtoets

In het wetsvoorstel van 2011 verplicht de minister naast het gebruik van een eindtoets het gebruik van een leerlingvolgsysteem (LVS) waarmee vorderingen op tenminste het gebied van taal en rekenen moeten worden gevolgd. Het gebruik van een LVS wordt gezien als een belangrijk middel om inzicht te krijgen in de leervorderingen en leerresultaten van leerlingen vanaf groep 3 tot aan het eind van groep 8. Op het moment dat het wetsvoorstel voorligt, zijn er twee LVS-systemen die de formele goedkeuring van de COTAN (Commissie Testaangelegenheden Nederland) hebben gekregen, waarvan het LVS van Cito verreweg het meest gebruikte is. Deze verplichting is in lijn met ingezet beleid gericht op meer rendement in termen van onderwijsopbrengsten dat in het onderwijs bekend wordt als 'opbrengstgericht werken'. De Onderwijsraad reageert positief op een verplichting van het gebruik van een LVS in het kader van opbrengstgericht werken. In de optiek van de Onderwijsraad betekent opbrengstgericht werken dat steeds opnieuw de stappen van de evaluatieve cyclus worden doorlopen: het stellen van doelen, het meten van resultaten, het analyseren van deze resultaten en het aanpassen van het onderwijsaanbod. Volgens de Onderwijsraad werkt in 2011 een derde van de scholen 'opbrengstgericht', in lijn met de eisen die de inspectie stelt. De Onderwijsraad acht een leerlingvolgsysteem 'onmisbaar voor opbrengstgericht werken' (Onderwijsraad, 2011: 13). De Raad pleit daarbij voor een prominentere plek voor de resultaten van het leerlingvolgsysteem bij de overdracht van leerlingen naar het vo.

De eindtoets en de uitkomsten van het LVS van Cito zijn op elkaar afgestemd. Ook het LVS meet taal- en rekenvaardigheid waarbij ordening van leerlingen in percentielscores het uitgangspunt is. Per toetsonderdeel worden de leerlingen geordend in ofwel percentielgroepen van 20%, oplopend van E naar A, of wel ingedeeld in percentielgroepen die geordend worden van V naar I, uitgaand van een indeling van 'onder', 'op' of 'boven' gemiddeld. Het LVS van Cito kent net als de eindtoets een relatieve normering. Het doel van de toetsen van dit LVS is het verdelen en ordenen van leerlingen van 'zwak' naar 'heel goed', gebaseerd op percentielscores. Het is daarmee onmogelijk dat alle leerlingen zich ontwikkelen tot een A-leerling of een I-leerling: de test is bedoeld om leerlingen te ordenen van zwak naar goed en dat volgens vaste percentielgroepen. Daarmee kan een vijfde deel van de leerlingen behoren tot de beste 20%, niet meer en niet minder. Dit geldt ook voor de zwakke groep. Ook die bestaat uit 20% van de leerlingen, niet meer en niet minder, ongeacht hoe zwak of sterk ze zijn. Een interpretatie die in de praktijk snel gemaakt wordt, is dat een score onder het gemiddelde hetzelfde is als een onvoldoende score. Maar een gemiddelde is niet hetzelfde als een na te streven doel in termen van 'voldoende'. En een A-score, dat wil zeggen een score gekoppeld aan de scores van de hoogst scorende groep van 20%,

Beste ouders, verzorgers,

In januari/februari en juni neemt de school van uw kind de LVS-toetsen van Cito af. Wij hebben de normen van een aantal toetsen herzien. Het gaat om de toetsen voor Begrijpend lezen, Rekenen-Wiskunde, Spelling en Woordenschat. In deze brief leggen we u uit wat dit betekent.

#### Waarom aangepaste normen?

Bij elke toets horen normen. Deze geven aan hoe de score van uw kind zich verhoudt tot de score van andere leerlingen. Leerlingen maken vandaag de dag de LVS-toetsen beter dan toen de toetsen met bijbehorende normen ontwikkeld werden. De oude normering geeft daarom een te rooskleurig beeld. Voor nu lijkt een hoger niveau leuk, maar wanneer uw kind straks naar het voortgezet onderwijs gaat, kunt u voor verrassingen komen te staan. Als u naar de oude normering kijkt, zou uw kind misschien bovengemiddeld scoren. Wanneer dan uit de centrale eindtoets opeens blijkt dat uw kind gewoon gemiddeld scoort, zult u verbaasd zijn. Met de update van de normering komt u bij andere toetsen en het schooladvies niet voor verrassingen te staan.

#### De resultaten op de LVS-toetsen van Cito

De toetsuitslagen op de LVS-toetsen bestaan uit twee delen: een score op de toets (de vaardigheidsscore) en de waardering van die score door middel van een letter (vaardigheidsniveau I t/m V of A t/m E). De letters geven aan hoe goed uw kind de toets heeft gemaakt vergeleken met zijn leeftijdsgenoten. Door deze vergelijking krijgen de resultaten betekenis. Zo kunnen we bijvoorbeeld zeggen dat een leerling die halverwege groep 5 een vaardigheidsscore van 25 behaalt, volgens de normering boven het landelijk gemiddelde van 22 scoort en dus niveau II behaalt. Als in de loop van de tijd het landelijk gemiddelde verschuift naar 25, dan is de conclusie dat de leerling met score 25 boven het gemiddelde scoort (II) niet (meer) juist. De leerling scoort nu immers gemiddeld. Vergelijk dit maar met de lengte: kinderen van nu worden gemiddeld groter dan hun ouders. Begin vorige eeuw zou je een reus zijn als je 1.70m lang was, nu ben je als man klein met zo'n lengte. Een lengte van 1.70m heeft nu dus een andere betekenis dan vroeger. Door de update van de normen trekt de school de juiste conclusies over de prestaties van uw kind ten opzichte van het gemiddelde in Nederland. Daarmee voorkomen we dat de resultaten op bijvoorbeeld de Entreetoets of de centrale eindtoets tegenvallen.

#### Wat betekenen de aangepaste normen voor uw kind?

Uw kind kan niet ineens minder goed lezen of rekenen dan voorheen. Wat wel is veranderd, is de waardering van zijn score.

*Uit: Ouderbrief Cito*

is ook niet de norm en zeker niet normaal of na te streven voor alle leerlingen. Want als te veel leerlingen te hoog gaan scoren, wordt de norm aangepast.

Omdat LVS-toetsen niet jaarlijks worden 'ververst' zoals wel het geval is bij de eindtoets, 'slijten' opgaven van het LVS en ook de toetsen, onder andere doordat opgaven langzaam bekend worden bijvoorbeeld omdat leerkrachten de toetsen bespreken, omdat er geoefend en voorbereid wordt op de toetsen, door kinderen die blijven zitten en de toetsen voor een tweede keer maken. Het kan zijn dat onderwijs daadwerkelijk verbetert en leerlingen steeds beter worden in de beoogde taal- en rekenvaardigheid. Hoe dan ook, de normering van de toetsen uit het LVS

wordt door Cito bijgesteld als blijkt dat 'te veel' kinderen 'te hoog' scoren op de toetsen van het LVS (zie kader met Ouderbrief Cito). Dit betekent volgens de brief niet dat bijstelling van de normen inhoudt dat een kind minder goed kan lezen maar alleen dat de waardering ervoor wordt aangepast: de werkelijkheid verandert niet volgens Cito, maar wel hoe ernaar wordt gekeken en hoe die wordt gewaardeerd. De waardering hangt af van wat de rest van de groep doet.

Aan de scores, afgeleid van het gemiddelde, zijn oordelen verbonden die oplopen in termen van 'zwak' tot 'goed'. En hieruit ontwikkelen zich (zelf)beelden waar niet alleen leerlingen zich naar gaan gedragen. Ook leerkrachten en docenten benaderen kinderen met verschillende scores anders. De 'opgelegde plicht' tot differentiëren waarop de Inspectie van het onderwijs in het kader van 'opbrengstgericht werken' toeziet vanaf het begin van deze eeuw, heeft ertoe bijgedragen dat al op jonge leeftijd behaalde scores en daaraan gekoppelde oordelen zich vastzetten in de geheugens van alle betrokkenen. Hiermee heeft het LVS gevolgen voor het kunnen inlopen van achterstanden door leerlingen die met achterstand ons onderwijs binnenkomen.

De in de wet opgenomen verplichting tot het gebruik van een LVS in combinatie met de 'druk' om als school het schooladvies in lijn te laten zijn met de uitkomsten van de later af te nemen eindtoets, heeft het LVS aan belang gewonnen. Onderzoeksinstituut Oberon heeft de waarde van LVS toetsen in relatie tot schoolsucces onderzocht. In de conclusies stelt Oberon dat het schooladvies 'sterk samenhangt met de scores op de toetsen uit het Leerling- en Onderwijsvolgsysteem' (Oberon, 2013: 16). Deze grote samenhang maakt het volgens Oberon mogelijk

een schooladvies 'te berekenen' (Oberon, 2013: 16). Dit is volgens Oberon het overwegen waard omdat het LVS van Cito een grotere samenhang heeft met het schooladvies dan de Centrale Eindtoets: 'Desondanks kan ruim 57% van de variantie in de positie op de leerjarenladder van de leerling in het vierde jaar van het VO worden verklaard op basis van LOVS-toetsen. De LOVS-scores bieden daarmee een iets minder goede voorspeller dan het leerkrachtadvies (zo'n 60%), maar wel weer beter dan de Cito-Eindtoets (zo'n 50%).' (Oberon, 2013: 16). Ook de inspectie van het onderwijs legt nadruk op het gebruik van de uitkomsten van het LVS bij het opstellen van het schooladvies. De PO-raad adviseert scholen de uitkomsten van de toetsen van het LVS van groep 7 en 8 mee te nemen in het schooladvies (zie kader).

#### **Adviezen voor basisscholen en leerkrachten**

- *Gebruik bij het opstellen en onderbouwen van het schooladvies in ieder geval de (ontwikkeling in de) vaardigheidsscores op de LVS-toetsen uit groep 7 en 8. De resultaten bij rekenen/wiskunde en begrijpend lezen zijn de beste voorspellers van het succes in het voortgezet onderwijs.*
- *Baseer aanvullende informatie over de sociaal-emotionele ontwikkeling, werkhouding, motivatie en dergelijke op meerjarige observaties. Leg deze informatie systematisch vast.*

*Uit: Handreiking Schooladviezen, PO-raad 2015*

De eindtoets vergelijkt leerlingen met elkaar op basis van een gecombineerde score van opgaven over de domeinen taal en rekenen. Daarbij is de eindtoets zodanig geconstrueerd dat de toets de leerlingen verdeelt over een door de jaren heen herkenbaar geworden score-schaal. Een score op die schaal heeft daarmee in de tijd betekenis gekregen. Een betekenis die in de beleving van alle betrokkenen niet zomaar een score is, maar een geschiktheidsoordeel is geworden waarmee kansen voor vervolgonderwijs worden bepaald. Het is nu een oordeel dat gevolgen heeft voor het beeld dat leerlingen, hun ouders en hun leerkrachten van hen hebben. Als leerlingen of scholen hard gaan werken kunnen ze hun de standardscore verhogen. Dit gaat wel ten koste van een en ander in de ordening. Er is niet voor iedereen plaats bij de 'beste 20%'. Dat is maar voor 20% van de leerlingen weggelegd.

De waarde van de toevoeging van de referentieniveaus als leerstofonafhankelijk criterium voor de beoordeling van de kwaliteit van onderwijs op school- en stelselniveau wordt beperkt door enerzijds de psychometrische vereisten die het selectie-doeleinde met zich meebrengt, maar ook door de gemaakte keuzes in het inspectietoezicht. De wens om scholen in competitie met elkaar en zichzelf te plaatsen is voor het beleid leidend. Zo is de eindtoets niet te typeren als een toets waarbij opbrengsten van onderwijs in termen van taal en rekenen kunnen worden vastgesteld. Het is een competitief systeem op basis van vergelijken, waarbij het wedstrijdement geacht wordt te fungeren als prikkel voor kwaliteit. De waarde van de eindtoets als bewaker van kwaliteit staat daarmee onder druk.

### **3.4.3 Beleidsambities voor eindtoets en LVS**

Uitkomsten van nationale en internationale toetsingen hebben de afgelopen decennia geleid tot meer aandacht voor de opbrengsten van onderwijs. In een brief van 1 maart 2011 aan de Tweede Kamer werkt minister van onderwijs Van Bijsterveldt-Vliegenthart uit dat ze opbrengstgericht werken en handelen op alle niveaus in het primair onderwijs wil versterken en dat ze daarbij de lat hoger wil leggen en prestaties wil verbeteren. Ze wil dat prestaties transparanter worden en

dat er heldere normen voor prestaties komen. Ze wil een 'landelijke eindtoets die volledig geënt is op de bestaande Eindtoets Basisonderwijs van het Cito, die jaarlijks op circa 85% van de basisscholen wordt afgenomen' waarbij 'het resultaat van de eindtoets wordt verplicht als het tweede, objectieve gegeven dat scholen naast de eindtoets gebruiken voor de overdracht naar het voortgezet onderwijs' en 'op het niveau van de school worden de toetsresultaten gebruikt om intern en extern een oordeel te kunnen geven over de onderwijsopbrengsten' (Ministerie OCW, 2011b). In haar Actieplan 'Basis voor Presteren', uit 2011, maakt minister Van Bijsterveldt-Vliegthart haar opbrengstambities concreet door aan te geven dat de leerprestaties omhoog moeten van een gemiddelde score van 535,4 (2010) naar 537 in 2015 en van een grensscore van de beste 20% leerlingen die in 2010 tussen 543-544 ligt naar 545 in 2015 en dat het aantal 'zwakke' scholen moet halveren, van 238 naar 119 (Ministerie OCW, 2011a). Ook gemeentes gaan hun ambities voor onderwijs formuleren in termen van het verhogen van gemiddelde scores op de eindtoets. Met een toets met een relatieve normering zoals de eindtoets kunnen dergelijke doelen alleen middels psychometrische aanpassingen gerealiseerd worden. Zelfs de grootste onderwijsinspanning kan het gemiddelde op een standaardscoreschaal niet veranderen omdat het werken met standaardscores met zich meebrengt dat een gemiddelde, hoe hoog of hoe laag ook, nooit anders zal corresponderen dan met een gemiddelde zoals de testontwikkelaar dat heeft voorzien voor de overzetting op de schaal van 501 tot 550 passend bij een beoogde normaalverdeling. In dit geval 536. Ook voor een gemeente geldt dat een gemeente wel een gemiddeld hogere score kan willen bereiken maar dat dit alleen kan 'ten koste van' andere gemeenten. Als alle gemeenten extra werk maken van taal en rekenen ligt teleurstelling in termen van behaalde standaardscores op de loer. In Rotterdam vat men de inspanningen na een aantal jaren samen: 'Wat betreft het verhogen van de citoscores, een van de speerpunten van het huidige beleid, is de winst betrekkelijk. Het streven om rond het landelijk gemiddelde te scoren is vooralsnog niet gelukt. De Rotterdamse scores zijn met pieken en dalen gestegen tot 532, maar liggen daarmee nog steeds onder het landelijk gemiddelde van 536.<sup>17</sup>

In de memorie van toelichting bij het wetsvoorstel voor de invoering van een centrale eindtoets(31-01-2012) stelt de minister van Onderwijs Van Bijsterveldt dat met het wetsvoorstel 'basisvoorwaarden voor opbrengstgericht werken: de centrale eindtoets en het werken met een leerling- en onderwijsvolgsysteem' gerealiseerd kunnen worden. Dat de overheid als eigenaar van de centrale eindtoets gaat fungeren, past volgens haar bij de taak van de overheid ten aanzien van onderwijs: 'De centrale eindtoets past bij een normerende taak van de overheid en bij de aanhoudende zorg van de regering voor het onderwijs. Het beoordelen van de leerresultaten door de school is immers een belangrijke functie van de eindtoets. De leerresultaten van de eindtoets zijn overigens van betekenis voor de verbetering van het onderwijs op alle niveaus: van het niveau van de leerling tot en met het niveau van het onderwijsstelsel, waar het toezicht op de scholen door de inspectie onder valt. Goed analyseren van resultaten, sterke en zwakke punten benoemen, doelen stellen, verbeteren van de zwakke punten en het borgen van sterke punten zijn daarbij de volgende stappen. Hieruit blijkt dat de eindtoets invulling geeft aan de stelselverantwoordelijkheid van de overheid voor kwalitatief goed onderwijs.' (Ministerie OCW, 2012).

Wat betreft de functie van de eindtoets stelt de minister dat die driedelig is:

17| <https://versbeton.nl/2014/02/rotterdams-onderwijs-is-bezig-met-een-inhaalslag/>

1. tweede, objectieve gegeven voor leerlingen, ouders en scholen bij de overdracht van het po naar het vo;
2. hulpmiddel voor de zelfevaluatie van scholen;
3. instrument voor de externe verantwoording en beoordeling van de leerresultaten van scholen.

De tweede en derde functie hebben beide betrekking op de kwaliteitsbewaking en de verantwoording. In het eerste geval, functie 2, betreft het een vorm van interne verantwoording en in het tweede geval, functie 3, een vorm van externe verantwoording. Als toelichting op de laatste verklaart de minister dat de Centrale Eindtoets de functie heeft 'van hulpmiddel om geobjectiverd te kunnen nagaan of de eindopbrengsten van deze vaardigheden op schoolniveau voldoen aan het bij de wet vastgestelde minimum qua onderwijsopbrengsten c.q. leerresultaten' (Ministerie OCW, 2012).

De minister ziet voor zowel de centrale eindtoets als het LVS een rol weggelegd in het bewaken van de kwaliteit van onderwijs door de Inspectie van het Onderwijs: 'De centrale eindtoets is een instrument voor het risicogerichte toezicht van de inspectie: het is een middel om inzicht te hebben in de ontwikkeling en kwaliteit van scholen.' (Ministerie OCW, 2012). In haar actieplan 'Basis voor Presteren' formuleert de minister het als volgt: 'Een centrale eindtoets maakt het verder mogelijk dat de inspectie de eindopbrengsten van scholen op een eenduidige manier kan beoordelen.' (OCW, 2011a: 6). Het onderliggende idee is dat onderwijsopbrengsten per school in kaart worden gebracht zodat scholen met elkaar vergeleken kunnen worden, waarbij de minister spreekt over 'leerwinst in beeld brengen' en 'welke bijdrage de school als geheel levert aan de ontwikkeling van de leerlingen'. En ten aanzien van het verplicht gebruik van een LVS stelt de minister dat een LVS 'scholen de mogelijkheid biedt om ook de longitudinale ontwikkeling van groepen te volgen en met elkaar te vergelijken. Het leerling- en onderwijsvolgsysteem heeft daardoor ook een meerwaarde bij zelfevaluatie van de school en bij kwaliteitszorg' (Ministerie OCW, 2012).

De verplichting om alle leerlingen met dezelfde of zeer vergelijkbare instrumenten te toetsen waarbij de resultaten niet primair bij de scholen liggen maar bij de overheid, hebben de positie van toetsen in het onderwijs wezenlijk veranderd. Het argument van de minister in de begeleidende brief bij het wetsvoorstel (Ministerie OCW, 2011b) dat op het moment van de indiening 'de landelijke eindtoets jaarlijks op circa 85% van de basisscholen wordt afgenomen' gaat voorbij aan de verandering van context door de verplichting. Tot dat moment is de keuze voor een eindtoets, ook voor een LVS, vooral een besluit dat een school zelf neemt en waar een school zelf van bepaalt wat er wel of niet met de scores gebeurt. Dat verandert nu: de scores op de eindtoets en de scores op het LVS zijn instrumenten geworden in handen van de overheid of ter overlegging aan de overheid in casu de Inspectie van het Onderwijs als vorm van verantwoording. Daarmee krijgt zowel de eindtoets als het LVS een ander karakter en verandert de context van het gebruik veel radicaler dan minister van Bijsterveldt doet voorkomen bij de introductie; de belangen en de belanghebbenden zijn veranderd. Een bestaande test inzetten in een andere context met andere beslissingen, kan wel degelijk gevolgen hebben voor de waarde, de validiteit van een toets. De validiteitsvraag is niet een vraag die alleen beantwoord moet worden bij de ontwikkeling van een instrument maar een vraag die voortdurende beantwoording vereist bij gebruik in de praktijk en zeker daar waar de praktijk en de functie van een test verandert (zie paragraaf 3.5).

Er zijn momenteel meer aanbieders van eindtoetsen, zoals bij aanneming van de wet mogelijk is gemaakt. Een daartoe in het leven geroepen 'Expertgroep' beoordeelt de kwaliteit van andere toetsaanbieders. Deze Expertgroep valt onder het ministerie. Zo is het ministerie zowel eigenaar en ontwikkelaar van de centrale eindtoets als beoordelaar van andere eindtoetsen. Vanaf het begin heeft het ministerie zich gebogen over de vergelijkbaarheid van de eindtoetsen onderling. Om die reden is in 2018 een serie opgaven opgenomen in iedere eindtoets, het 'anker', waarmee de vergelijkbaarheid van de gehanteerde normeringen kan worden gecontroleerd: 'Het is een mooi resultaat dat de eindtoetsen daardoor vergelijkbaar zijn geworden.' (Ministerie OCW, 2018). Tegelijkertijd meldt de minister in dezelfde brief dat 'de normering nog niet afdoende gelijk was in 2018'. De minister kondigt daarop aan dat hij met de eindtoetsaanbieders afspraken zal maken over het toepassen van de resultaten op basis van het anker: 'Dit heeft tot doel dat de normering gelijk wordt en de resultaten van de eindtoetsen daarmee ook gelijk worden.' In de eindevaluatie van de Wet eindtoetsing po van juni 2019 geeft de minister aan dat er stappen zijn gezet om ook de normeringen gelijk te maken: 'Zo zijn er bindende afspraken gemaakt met de toetsaanbieders hierover, zoals onder andere CvTE en inspectie per brief aan mij hebben verzocht. Dat zorgt ervoor dat het voor de resultaten niet uitmaakt welke toets de leerling maakt. Zo komen we tegemoet aan de zorgen die de inspectie en CPB hebben geuit.' (Ministerie OCW, 21 juni 2019). Vanuit een oogpunt van regie is het begrijpelijk dat de overheid wil zorgen voor vergelijkbare uitkomsten, ongeacht de eindtoets die gebruikt wordt. Maar bij de aanneming van de wet is indertijd nadrukkelijk beoogd dat er keuzevrijheid moet zijn voor scholen. Nu is er sprake van een overheid die ervoor zorgt dat de uitkomsten niet verschillen, waarbij de 'eigen' toets, de centrale eindtoets, leidend is voor de beoogde uniformering zoals blijkt uit de situatie die in 2019 is ontstaan bij de bekendmaking van de scores.

De eindtoets is in de afgelopen decennia een belangrijke rol gaan vervullen in de grip die de overheid heeft gezocht op de invulling van zowel inhoud als vorm van het onderwijs. De formele ruimte die er gezien de Grondwet artikel 23 is voor de vrijheid van onderwijs is daarmee sterk ingeperkt. De belangen voor alle betrokkenen die met de uitkomsten van de verplichte, gestandaardiseerde vormen van toetsen samenhangen, hebben het karakter van het onderwijs en toetsing veranderd. Dit veronderstelt dat de geldigheid van de eindtoets boven elke twijfel verheven is.

### 3.5 Geldigheid van de eindtoets

De eindtoets is te definiëren als een 'high stakes test' waarmee bedoeld wordt dat het hier gaat om een test waarbij de uitkomsten vergaande gevolgen hebben. Gevolgen die betrekking hebben op de toekomst van leerlingen en hun zelfbeeld. Voor een toets met een dergelijk karakter is een geldigheidsonderzoek van groot belang. Een geldigheidsonderzoek is wat in de testwetenschap wordt aangeduid als een onderzoek naar de validiteit van een test. Validiteit in de oorspronkelijke betekenis van het woord gaat over de waarde en waardering die aan iets gegeven mag worden, over deugdelijkheid en geldigheid. Bij de eindtoets blijft de discussie over validiteit veelal beperkt tot de waarde van het instrument zelf. Maar het is van belang de functie van de eindtoets in zijn volledige context te blijven zien om ook daar de rol en waarde van te bepalen. Madaus stelt dat de introductie van een high-stakes test bevraagd moet worden op wie er wint en wie er verliest. Hij ziet een test als een technologie die onderdeel is geworden van politieke besluitvorming

omdat de informatie die met een test gegenereerd wordt een vorm van macht creëert die vooral in het voordeel is van degenen die de controle erover hebben. Hij pleit ervoor te zorgen dat uit de evaluatie van het gebruik van tests naar voren komt welke individuen en groepen benadeeld worden (Madaus, 1994: 81). Daarmee bepleit Madaus een veel bredere interpretatie van een onderzoek naar validiteit dan gangbaar is en door testwetenschappers wordt bepleit.

### **Definitie en debat**

In de Internationale 'Standards for Educational and Psychological Testing'<sup>18</sup> (2003) wordt validiteit voor de context van toetsing gedefinieerd als 'the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests'. Deze definitie is een keuze uit de vele definities en invullingen die de laatste 100 jaar aan het begrip validiteit zijn gegeven. Want wat verstaan moet worden onder 'validiteit' is zeker geen uitgemaakte zaak en is nog altijd onderwerp van discussie. Overeenstemming lijkt er wel te bestaan over het feit dat validiteit geen vaste eigenschap of kenmerk is van een meetinstrument. Een test ontleent geldigheid aan de toepassing: 'Validity has always been considered to be a property of test use, not of the test itself.' (Shepard, 1982: 10). Een test zelf kan niet het predicaat 'valide' of 'niet-valide' krijgen. Zoals Cronbach (1971) het stelt: 'One validates, not a test, but an interpretation of data arising from a specified procedure' (Sheppard, 1993: 406). Onduidelijkheid over de definitie van validiteit heeft gevolgen voor het doen van onderzoek naar de validiteit van een test. Want wat precies onderzocht moet worden als het gaat om het aantonen of vaststellen van validiteit (of wellicht kwaliteit) van een test ligt daarmee niet eenduidig vast. Dit is zorgelijk omdat 'The Standards', de afspraken die wereldwijd gemaakt zijn voor de ontwikkeling van tests, stelt dat validiteit het meest belangrijke, vast te stellen aspect van een test is: 'Validity is the most important consideration in test evaluation.' (in: Standards 2003). Dit gebrek aan consensus heeft gevolgen voor het kunnen beoordelen van validiteit van tests, internationaal maar ook nationaal. Het biedt ontwikkelaars van tests veel ruimte om zelf te bepalen hoe zij validiteit zien en willen onderbouwen.

Messick pleit voor een ruime interpretatie van het begrip validiteit. Hij definieert validiteit als 'an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment' (Sheppard, 1993: 406). En hij gaat daarin verder dan tot dan toe gebruikelijk is. Belangrijk, en controversieel aspect van zijn blik op validiteit is dat Messick de sociale consequenties van testgebruik als een apart aspect van validiteit onderscheidt, een vorm van validiteit die volgers van hem definiëren als 'consequential validity'. Nichols en Berliner sluiten hierbij aan en komen daarmee tot vier soorten validiteit: content validity, construct validity, criterion validity en consequential validity. Deze laatste door hen onderscheiden vorm van validiteit is een uitwerking van het debat over validiteit in lijn met de manier van kijken zoals Messick die voorstelt. Nichols en Berliner definiëren 'consequential validity' als de vorm van validiteit die gaat over de consequenties en beslissingen die gekoppeld zijn aan de uitkomsten van tests. Bij onderzoek naar 'consequential validity' moet het gaan om onderzoek naar de gevolgen van tests en de afwegingen die daarmee samenhangen voor het al dan niet gebruiken van tests voor bepaalde doeleinden en contexten. Zij vatten dat samen met de volgende te beantwoorden vragen:

---

18] Ontwikkeld door de American Educational Research Association, American Psychological Association, Joint Committee on Standards for Educational, Psychological Testing (US), & National Council on Measurement in Education.



'Who gets hurt, who gets helped, and is it worth it?' (Nichols, 2007: 119). Van het 'validity model' dat Messick in 1989 introduceert en waar Nichols en Berliner voor kiezen, neemt Popham (2000) nadrukkelijk afstand. Popham erkent de waarde van Messick voor het vragen van aandacht voor de sociale consequenties die met tests samenhangen, maar hij wil hiervoor niet het begrip validiteit 'oprekken'. Popham pleit voor een terugkeer naar een meer beperkte inkleuring van het begrip validiteit door het als volgt te definiëren: 'Validity refers to the accuracy of the inferences (or interpretations) that are made based on student's performances on measurement devices.' (Popham, 2000: 114). Ook de Nederlandse psychometrici Borsboom en Mellenbergh zijn van mening dat consequenties die samenhangen met het gebruik van tests buiten de scope van validiteitsonderzoek vallen (Wools, 2015: 10). Borsboom, Mellenbergh en Van Heerden stellen dat de testpraktijk met het oprekken van het begrip 'validiteit' in de afgelopen honderd jaar niet gewonnen heeft aan eenduidigheid: 'This is because in the past century, the question of validity has evolved from the question of whether one measures what one intends to measure (Cattell, 1946; Kelley, 1927), to the question of whether the empirical relations between test scores match theoretical relations in a nomological network (Cronbach & Meehl, 1955), and finally, to the question of whether interpretations and actions based on test scores are justified—not only in the light of scientific evidence but with respect to social and ethical consequences of test use (Messick, 1989)' (Borsboom et al., 2004). Door te pogen alle aan een toets gerelateerde issues als relevant te beschouwen in het kader van validiteit, worden zowel de theoretische georiënteerde onderzoeker als de meer praktisch georiënteerde tester volgens hen niet geholpen in de uitvoering van hun dagelijks werk. Zij pleiten voor een terugkeer naar de basis zoals Kelley die in 1927 heeft gelegd door te stellen dat validiteit betrekking heeft op de vraag of een 'test meet wat hij beoogt te meten'. Zij hanteren daarbij de volgende redenering: 'If something does not exist, then one cannot measure it. If it exists but does not causally produce variations in the outcomes of the measurement procedure, then one is either measuring nothing at all or something different altogether. Thus, a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure. The general idea is based on the causal theory of measurement.' (Borsboom et al., 2004: 1061). Wools komt tot de slotsom dat wat er te zeggen valt over validiteit grote overeenkomst heeft met wat gezegd kan worden over het begrip 'kwaliteit': 'The definitions of validity and quality became very closely interrelated.' (Wools, 2015: 202). Zij vraagt zich af of het in de praktijk wel nodig is om onderscheid te maken tussen de beide begrippen. In de publicatie 'Toetsen op school' van Cito uit 2013 stelt Wools dat het bij validiteit gaat om de vraag of de toetscores de informatie bieden waarin we geïnteresseerd zijn en of op basis van deze scores de gewenste beslissingen genomen kunnen worden. Zij hanteert daarbij het begrip 'validiteitsbewijzen' waar zij ook 'betrouwbaarheid' onder schaaft (Cito, 2013). Drenth en Sijsma stellen dat er weliswaar een veelheid aan opvattingen en definities over validiteit is maar dat er wel degelijk een gemeenschappelijk basisprincipe te onderscheiden is dat zij samenvatten als 'de mate waarin een test aan zijn doel beantwoordt' (Drenth & Sijsma, 1990). Daarmee sluiten zij aan op de definitie van Kelley uit 1927. Zij onderscheiden een tweetal hoofdvormen van validiteit die direct te relateren zijn aan doel van een test. Van den Brink en Mellenbergh doen dat ook. Zij maken onderscheid tussen begripsvaliditeit en criterium-georiënteerde validiteit waarbij de laatste vooral van belang is in het kader van tests bedoeld voor selectie, diagnose, classificatie of een andere vorm van beslissend testgebruik en de eerste voor testgebruik met een meer beschrijvende functie (Van den Brink & Mellenbergh, 2006: 294). De COTAN, de Commissie

Testaangelegenheden Nederland, die onder andere is belast met de goedkeuring van eindtoetsen en LVS-toetssystemen, hanteert dezelfde tweedeling.

Wilson pleit er vanwege het risico op 'tautologie' en 'redundantie' voor om niet validiteit te bewijzen maar om de bewijsvoering om te keren en te komen tot een situatie waarin bij tests aangetoond moet worden dat er geen sprake is van 'invalidity' (Wilson, 1998: 184). Om te ervaren hoe anders er dan tegen de problemen van testing aangekeken kan worden heeft hij de zin uit 'The Standards' over het belang van validiteit omgekeerd. Dit brengt hem tot de volgende tekst: 'Invalidity is the most important consideration in test evaluation. The concept refers to the inappropriateness, meaninglessness, and uselessness of the specific inferences made from test scores. Invalidity or error estimation is the process of accumulating evidence to problematise and ultimately reject such inferences.' (Wilson, 1998: 183).

### ***Consequential validity***

Uit het voorgaande debat wordt duidelijk dat de testwereld ertoe neigt de validiteit van een test vooral te willen beperken tot de waarde en waardering van het instrument zelf in relatie tot wat men beoogd te meten. Door tests te beschouwen als beleidsinstrumenten kan echter niet voorbij gegaan worden aan de 'gevolgen' van testen in het domein waar zij een rol in moeten vervullen, de 'consequential validity' waar Nichols en Berliner over spreken. Bij tests als beleidsinstrumenten gaat de vraag verder dan 'Who gets hurt, who gets helped, and is it worth it?' (Nichols, 2007: 119). Een test als beleidsinstrument gaat ook over de belangen die met het gebruik van tests voor het verdelen van belangrijke kansen samenhangen. De vraag van Madaus naar 'who gains and who loses' is bij de inzet van een test als onderdeel van een politiek besluit relevant. Sociale en culturele groepen verschillen in de mate waarin zij de waarden delen die samenhangen met tests en indirect gepromoot worden door tests (Madaus, 1994: 79). Want tests zijn volgens Madaus 'scheppers' van dat wat belangrijk wordt gevonden in onderwijs en voor onderwijs. Als tests ingezet worden om bepaalde maatschappelijke of onderwijskundige doelen dichterbij te brengen, is het relevant het validiteitsdebat op die doelen te betrekken. Voor een high-stakes nationaal testprogramma formuleert Madaus een aantal vragen die in het kader van de waardering van de effecten van een testprogramma relevant zijn en op voorhand onderzocht moeten worden (Madaus, 1994: 81):

1. Leidt de inzet van het tests tot een onderwaardering van dat wat niet in de test aan bod komt?
2. Hoe zullen potentiële risico-leerlingen de gekozen inhoud van de test ervaren?
3. Wat zullen de gevolgen zijn voor hen die laag scoren?
4. Heeft de uitkomst gevolgen voor het volgen van vervolgonderwijs? Zo ja, wat zijn de gevolgen van het gebruik van de tests in relatie tot kansengelijkheid?

Het beschouwen van de validiteit van de eindtoets zoals Madaus dat voorstelt, wordt in Nederland 'gehinderd' door het feit dat het exacte doel van de eindtoets in relatie tot het doel van onderwijs door de overheid niet duidelijk omschreven is. De vraag, relevant in het kader van validiteit, of de eindtoets in zijn aard en uitwerking het doel van onderwijs bevordert of juist belemmert, is daardoor niet te beantwoorden.

Vanuit de beleidsambities die met de eindtoets nagestreefd worden kan worden onderzocht of en in hoeverre de eindtoets bijdraagt aan de daadwerkelijke realisatie van deze beleidsambities.

De twee onderscheiden functies die aan de eindtoets worden toebedeeld zijn als uitgangspunt genomen voor een nadere beschouwing van het vraagstuk van validiteit: de functie van selectie en de functie van kwaliteitsbepaler. Voor de beide functies van de eindtoets, de selectiefunctie en de functie van kwaliteitsbeoordeling, is onderzocht hoe de validiteit ervan nu wordt onderbouwd.

### 3.5.1 Validiteit en selectie

De eindtoets is vanaf het begin bedoeld als onderdeel van de toelatingsprocedure voor het v.h.m.o.. Het selecteren van de 'meest geschikten' is daarbij het uitgangspunt. In de verantwoording van de eerste Amsterdamse Schooltoetsen staat dat 'de toets alleen bedoeld is om door onderwijs verworven kennis en inzicht te meten; en dan nog met de specifieke bedoeling om tot een zekere differentiatie tussen leerlingen te leiden, zo, dat de uitkomsten in principe kunnen worden gebruikt voor beslissingen of adviezen over het te kiezen vervolgonderwijs' (Samenwerkende Instituten, 1967: 140). In de verantwoording van de eindtoets van 2010 spreekt Cito over de functie van de eindtoets als volgt: 'De primaire functie van de Eindtoets Basisonderwijs is het geven van onafhankelijke informatie voor de keuze van een passend brugklastype.' (Van Boxtel et al., 2011:63). Dit is de formulering die in de verantwoording door CvTE in 2015 is overgenomen. Zo geformuleerd heeft de toets vooral de bedoeling (gehad) om een rol te spelen in het plaatsingsproces. Op basis van 'communale doelstellingen'<sup>19</sup> wordt de inhoud van de eindtoets uitgewerkt waarbij een optelling van scores tot één totaalscore de basis vormt voor de ordening en de selectie. In de verantwoording van 1967 onderbouwen de Samenwerkende Instituten het werken met gecombineerde scores als volgt: 'afgaan op onderdelen kan misleidend zijn; totaalscores daarentegen, die op een groter aantal en meer gevarieerde opgaven betrekking hebben, zijn in het algemeen als meetuitkomst en prognostisch waardevoller' (Samenwerkende Instituten, 1967: 36). De eindtoets is niet eenduidig te positioneren als test (selectie-oogmerk) ofwel als (schoolvorderings)toets waarbij het doel is te achterhalen of leerlingen bepaalde kennis en vaardigheden beheersen. In het eerste geval is voorbereiding niet gewenst en in het tweede geval is voorbereiding passend. In het eerste geval kan voorbereiding worden gezien als een bedreiging van de validiteit en in het tweede geval is het niet voorbereiden van invloed op de geldigheid van de uitkomsten.

De voorspellende waarde, de predictieve validiteit, van de eindtoets is een belangrijk aspect van de eindtoets als selectie-instrument. Dit is vanaf het begin van de eindtoets onderbouwd door te stellen dat 'resultaten van voorafgaand onderwijs meestal zeer waardevolle aanwijzingen geven over wat er in daarop volgend onderwijs van een leerling kan worden verwacht' (Samenwerkende Instituten, 1967: 3). Cito en CvTE volgen die redenering: 'De toets meet wat een leerling in vergelijking met andere leerlingen in acht jaar basisonderwijs geleerd heeft. [...] En leervorderingen zeggen iets over de vraag hoe goed een leerling de verschillende typen van het voortgezet onderwijs aankan.' (Van Boxtel et al., 2011: 9). In de laatste verantwoording van CvTE blijkt de formulering in grote lijnen gelijk gebleven: 'Het lijkt gerechtvaardigd te veronderstellen dat leerprestaties uit het verleden iets zeggen over leerprestaties in de toekomst.' (CvTE, 2015: 12). Uit het feit dat taal en rekenen in de toets opgenomen zijn zou dan afgeleid kunnen worden dat een

19] Dit is de term die bij de start van de eindtoets wordt gebruikt om aan te duiden dat de inhoud van de eindtoets gebaseerd is op onderwijsdoelen waarvan men in die tijd dacht dat het gros van de scholen dat zou erkennen als relevante onderwijsinhoud en hier ook daadwerkelijk aandacht aan besteedde in het onderwijs.

gecombineerde score van taal en rekenen wordt gezien als goede voorspeller van toekomstig schoolsucces. De keuze voor een gecombineerde score van taal en rekenen als voorspeller van toekomstig schoolsucces is inhoudelijk echter niet met onderzoek onderbouwd. Cito verklaart dat deze keuze eerder pragmatisch is gekozen omdat de eerste eindtoets 'een analyse van de stand of het peil van het onderwijs mogelijk moest maken [...]. Daartoe moest de inhoud van de toets aansluiten bij onderdelen van het curriculum waaraan scholen vrijwel zonder uitzondering ongeveer dezelfde aandacht besteden.' (Van Boxtel et al., 2011: 10). Lang heeft naast taal en rekenen het onderdeel 'informatieverwerking/studievaardigheid' meegeteld in de bepaling van de standardscore. Dit onderdeel maakt de laatste jaren geen deel meer uit van de eindtoets maar de reden van het ooit toevoegen en later weer weglaten van dit onderdeel is niet met onderzoek onderbouwd en niet verantwoord.

Vanaf de eerste versie van de eindtoets uit 1968 wordt de validiteit van de eindtoets onderbouwd door te werken met correlaties van toetsscores met een tweede extern criterium. In de eerste versie van de eindtoets worden correlaties onderzocht met twee externe gegevens: de aanmelding voor het v.h.m.o. die al had plaatsgevonden op basis van het advies van de school en het ouderlijk milieu. Het doel bij de Amsterdamse Schooltoetsen is te komen tot een zo hoog mogelijke correlatie tussen ordening van leerlingen op basis van de toets en de adviezen van de scholen voor plaatsing in het v.h.m.o. (in Amsterdam ongeveer 20% van de leerlingen): 'Gegeven het feit, dat de aanmelding voor het v.h.m.o. in het algemeen al de voor het v.h.m.o. meest geschikte leerlingen eruit haalt, moeten de Schooltoetsscores, als zij óók 'goede voorspellers' zijn van later schoolsucces, een vrij hoge correlatie daarmee vertonen.' (Samenwerkende Instituten, 1967: 133). Van de tien procent beste scoorders (op basis van gecombineerde testscore) blijkt 84% van de leerlingen aangemeld voor het v.h.m.o. (Samenwerkende Instituten, 1967: 145). De Samenwerkende Instituten stellen in 1967 dat 'deze validiteitsbevindingen een krachtige steun verlenen aan de veronderstelling, dat (ook) de schooltoets in staat is de v.h.m.o.-geschikten eruit te halen' (Samenwerkende Instituten, 1967: 133). Een hoge correlatie van toetsscore met het door de school gegeven schooladvies wordt gezien als een belangrijke maat in het kader van de onderbouwing van validiteit. Nog altijd wordt ter onderbouwing van de validiteit van de eindtoets onderzoek gedaan naar correlaties met een tweede, extern gegeven in de vorm van het advies van de basisschool. Juist die adviezen, en dan vooral de vermeende subjectiviteit ervan, zijn tegelijkertijd op politiek, beleids- en publieke opinieniveau steeds aanleiding geweest om aan te sturen op een 'objectief' en 'onafhankelijk' oordeel in de vorm van toetsscores ontleend aan gestandaardiseerde metingen. Stellwag wees in 1955 al op de cirkelredenering waarvan bij valideringsonderzoek van IQ-tests voor de onderwijscontext destijds al sprake was: 'Want òf het een òf het ander: Het oordeel der onderwijzers en hun cijfers zijn geen betrouwbare criteria. Dan moet men dus naar criteria zoeken die wel betrouwbaar zijn. Hiervoor zou dan de intelligentietest in aanmerking kunnen komen. Maar het criterium voor de objectiviteit en de methodische juistheid van deze test wordt gevonden met behulp van iets, dat men vanwege de er aan verbonden onvermijdelijke subjectiviteit, verwerpt, nl. de schoolprestaties van het kind en de beoordeling daarvan door de school. Zo draaien we in kringetjes rond.' (Stellwag, 1955: 54).

De ontwikkelaars van de eerste versie van de eindtoets realiseren zich dat lang niet alles gemeten kan worden zoals het in theorie zou moeten en wat van belang is in het kader van het onderbouwen van validiteit van uitkomsten: 'In sommige gevallen van partiële dekking is in principe wel

een volstrekt dekkende operationele definitie voorhanden, maar deze is alleen via zo omslachtige of tijdrovende procedures toe te passen dat men met een benadering volstaat. Men weet dan dus wel precies wat men zou willen hebben (meten), maar neemt uit praktische overwegingen genoegen met een indirect-benaderende bepalingswijze.' (De Groot, 1961: 91). En soms is een criterium om mee te vergelijken niet direct voorhanden. Ook dan moet een oplossing worden gezocht: 'Van partiële dekking en benadering kan men bijvoorbeeld ook spreken, als bij onderzoeken over de voorspelling van studiesucces niet wordt gewacht totdat men, na zoveel jaren, definitief kan onderscheiden tussen gediplomeerden en niet-gediplomeerden, maar met een tussentijds criterium, dus met een benaderende operationele definitie van 'studiesucces' wordt volstaan.' (De Groot, 1961: 90).

Dit langeretermijncriterium, dat ontbreekt bij de eerste versie van de eindtoets, wordt door Cito later wel gebruikt: Cito onderzoekt met welke scores leerlingen tot welke brugklastypen zijn (aangemeld en) toegelaten. De bevindingen worden als volgt samengevat: 'In het algemeen worden leerlingen met hogere standaardscores op moeilijker schooltypen geplaatst. Er blijkt echter ook een flinke overlap tussen de schoolsoorten.' (Wijnstra, 1984: 14). Uit latere verantwoordingen van de eindtoets blijkt dat niet alleen de plaatsing onderzocht wordt in het eerste leerjaar maar ook de relatie tussen eindtoets en functioneren in het voortgezet onderwijs. Hierbij zijn er 'enkele problemen': 'Over welke periode moet schoolsucces voorspeld worden? Bijna nooit zijn objectieve toetsgegevens aanwezig, zeker niet die in alle schooltypen zijn afgenomen (restriction of range), en rapportcijfers blinken echter over het algemeen niet uit door stabiliteit en betrouwbaarheid.' (Wijnstra, 1984: 25). Voor het correleren van de eindtoets aan schoolprestaties bijvoorbeeld in de vorm van gemiddelde rapportcijfers acht Cito deze bron van informatie niet geschikt voor het onderbouwen van de voorspellende waarde van de eindtoets.

In 2011 voegt Cito aan de verantwoording uitkomsten toe van ander, uitgebreid correlatieel onderzoek. Hierbij is de eindtoets vergeleken met andere intelligentietests. Cito rapporteert hierover: 'Deze bevindingen tonen aan dat de Eindtoets en intelligentiemetingen ergens tussen de 50% en 60% van hun variantie gemeenschappelijk hebben.' (Van Boxtel et al., 2011:11). Cito noemt deze uitkomsten 'comfortabel' en geeft daarvoor twee redenen: 'De correlatie is enerzijds substantieel, en toont aan dat schoolse prestaties inderdaad samenhangen met intelligentie [...], maar anderzijds is de correlatie laag genoeg om een zelfstandig gebruik van de eindtoets te rechtvaardigen.' (Van Boxtel et al., 2011:11).

De afgelopen vijftien jaar wordt in het kader van de validiteitsvraag voor de eindtoets gerefereerd aan onderzoek dat betrekking heeft op het verband dat er is tussen het advies op basis van de toets en de plek die het kind daadwerkelijk heeft na 1 jaar (plaatsing) of 3 jaar in het voortgezet onderwijs (studiesucces). Ook CvTE kiest ervoor op basis van het toelatings- en doorstroomonderzoek de validiteitsvraag te beantwoorden, namelijk door de scores te vergelijken met de posities die leerlingen vervolgens innemen in het vervolgonderwijs, een onderzoek dat sinds enige jaren wordt uitgevoerd en gerapporteerd door het CBS. Het gaat hierbij om het berekenen van de hoeveelheid 'goede' voorspellingen, dat wil zeggen voorspellingen waarbij het advies en de daadwerkelijke plek in het vo na 1 of 3 jaar in lijn zijn met het destijds gegeven advies. In de verantwoording van 2011 meldt Cito dat 'in ongeveer 70% van de gevallen er sprake is van overeenkomst tussen het doorstroomadvies (bedoeld wordt: *van de leerkracht/school*) en het advies

dat wordt uitgebracht op basis van de standaardscore op de eindtoets. Dit betekent dat bij 30% van de leerlingen de uitkomst van de toets en het advies van de school niet in lijn zijn met elkaar. Uitgaand van 150.000 leerlingen betreft het 45.000 leerlingen waarbij school- en toetsadvies minimaal één categorie uit elkaar liggen. In drie procent van het totale aantal adviezen, uitgaand van 150.000 leerlingen betreft het dan 4500 leerlingen, gaat het om meer dan één categorie verschil: 'In ongeveer 97% van de gevallen (96.7% in 2010) liggen beide adviezen niet verder dan één categorie uit elkaar, zowel hoger als lager.' (Van Boxtel et al., 2011: 81).

De leeftijd waarop een leerling de eindtoets aflegt, kent grote verschillen. Dit kan het gevolg zijn van het 'overslaan' van een klas, van 'zitten blijven' maar ook binnen de 'reguliere' groep doen zich leeftijdsverschillen voor. Doornbos heeft de samenhang tussen de leeftijdspositie binnen de jaargroep en het schoolsucces van leerlingen onderzocht. Uit zijn onderzoek blijkt een 'onmiskerbare benadeling van de jongere kinderen in een klas' (Doornbos, 1971: 1). Hij concludeert dat 'een toevallige omstandigheid – de geboortemaand – in een groot aantal gevallen een beslissende invloed blijkt uit te oefenen op de gang van de leerlingen door ons schoolstelsel' (Doornbos, 1971: 171). Sharp, Hutchison en Wetton concluderen op basis van hun onderzoek in het Verenigd Koninkrijk dat in de zomer geboren kinderen aan het eind van 'key stage 1' (leeftijd 6 -7 jaar) minder goed presteren dan kinderen die in de herfst geboren zijn (Sharp et al., 1994: 107). Omdat het schooljaar in het Verenigd Koninkrijk op 1 september start, zijn kinderen die in de herfst geboren worden bijna een jaar ouder dan hun klasgenootjes die in de zomer geboren zijn. Dit maakt dat de onderzoekers ervoor pleiten hier rekening mee te houden bij de interpretatie van resultaten van metingen (Sharp et al., 1994: 117). In haar onderzoek naar de plaatsing van leerlingen in het voortgezet onderwijs in verschillende Europese landen komt Korthals tot de conclusie dat relatief jonge kinderen een grotere kans hebben in een lagere 'track' geplaatst te worden. Dit effect is groter in landen waar sprake is van 'early tracking' zoals in Nederland. Jongere kinderen behoren in de lagere tracks van het voortgezet onderwijs dan wel weer tot de beter presterenden terwijl de relatief oudere kinderen vaker tot de minder goed presterenden behoren in de track waarin zij geplaatst zijn. Korthals stelt dat het kijken naar eerdere schoolprestaties voor het opstellen van het schooladvies geen optimale plaatsing geeft van leerlingen over de tracks (Korthals, 2015: 131). Zij stelt dat er sprake is van een bias in het nadeel van relatief jonge leerlingen als eerdere prestaties bepalend zijn voor schooladvisering (Korthals, 2015: 132).

Wilbrink stelt dat 'verschillen waarop men selecteert aantoonbaar verband moeten houden met hetgeen waartoe men selecteert' (Wilbrink, 1997: 345). In hoeverre de beheersing van taal (en de daarbij nu aangebrachte beperking tot taal als lezen, taalverzorging en schrijven) en rekenen geschiktheid voor de verschillende vormen van voortgezet onderwijs kunnen voorspellen, is geen onderwerp van onderzoek geweest maar is gebaseerd op aannames. Wilbrink concludeert dat 'de voorspellende geldigheid van de citotoets nooit adequaat is onderzocht, en dus onbekend is' (Wilbrink, 1997: 364). Een geldigheidsonderzoek zou 'dubbelblind' moeten gebeuren verwijzend naar de manier waarop onderzoek naar medicijnen plaatsvindt (Wilbrink, 1997: 364). 'In het geval van classificatie of selectie zoals in een onderwijssysteem met vroege 'tracking' is de onbekendheid van de voorspellende geldigheid een ernstige zaak' (Wilbrink, 1997: 364). Meijnen stelt dat het systeem zelf de toetsing van de validiteit van de selectie verhindert, omdat het 'zichzelf waarmakende voorspellingen genereert' (Meijnen, 2004: 83).

Nader onderzoek naar de onderbouwing van de validiteit van de eindtoets in het kader van de selectiefunctie van de eindtoets laat zien dat er de nodige haken en ogen zijn ten aanzien van de waarde van de technieken die van oudsher worden gebruikt om de validiteit van tests bedoeld voor selectie en voorspelling van toekomstig functioneren te onderbouwen.

Voor de beoordeling van de validiteit van het gebruik van de eindtoets voor vraagstukken van kwaliteitsbeoordeling zijn andere onderbouwingen aan de orde.

### 3.5.2 Validiteit en kwaliteit

Uit de wettekst is te herleiden dat de eindtoets betrekking moet hebben op het vaststellen van de taal- en rekenvaardigheid van leerlingen, in relatie tot de vastgestelde referentieniveaus. Noch de wettekst, noch de toelichting geven een inhoudelijke onderbouwing ter verantwoording van deze keuze in relatie tot de functie van kwaliteitsbewaking. De overheid positioneert de eindtoets als een onderdeel, verplicht, van het adviesproces van de school zelf: 'De basisschool dient de eindtoets te gebruiken als tweede objectief gegeven in de schooladvisering.' (Ministerie OCW, 2017). Een expliciet doel voor de toets zelf formuleert de overheid niet.

Voor de vraag naar de waarde van de eindtoets als indicator van onderwijskwaliteit is tot op heden geen onderzoek bekend. Bij de eerste versie van de eindtoets, de Amsterdamse Schooltoetsen, wordt in de verantwoording gemeld dat men vanwege het ontbreken van algemene leerdoelen kiest voor een inhoudelijke selectie van wat men destijds als 'basisleerstof' beschouwt, namelijk Nederlandse taal en rekenen, omdat, zo stelt de verantwoording, men daar 'vermoedelijk de meeste onderlinge overeenkomst (bedoeld wordt: tussen scholen) mag verwachten' (Samenwerkende Instituten, 1967: 34). Volgens Cito wordt onder deze 'communale doelstellingen' verstaan: 'algemeen aanvaarde doelstellingen die voor alle leerlingen van het basisonderwijs geldig worden geacht, betreffende alleen cognitieve vaardigheden' (Wijnstra, 1984: 7). Ook CvTE stelt in de verantwoording van 2015 dat 'de opgaven een operationalisering zijn van communale doelstellingen' (CvTE, 2015: 13). Voor wat betreft de verdeling van opgaven over taal en rekenen stelt CvTE in 2015 dat 'de verdeling van de opgaven over de verschillende domeinen een afspiegeling vormt van de doorsnee onderwijsbesteding voor taal en rekenen (afgeleid van PPO-balansen)' (CvTE, 2015: 24).

De dubbele ambitie van zowel selectie als evaluatie heeft de eindtoets van het begin af aan parten gespeeld. De Groot denkt in 1965 nog heel optimistisch dat het moet lukken beide ambities te verenigen, maar al snel blijkt de evaluatieve ambitie ten aanzien van de opbrengsten van onderwijs zich moeizaam te verhouden tot de psychometrie gericht op selectie. Uit de verantwoording van 1984 blijkt Cito zich bewust van de lastige combinatie van selectie en evaluatie: 'Wat betreft de evaluatie van het onderwijs: het besef is ontstaan dat een toets als deze te beperkt is van opzet om te functioneren in een onderzoek naar de stand van het onderwijs, met name omdat slechts een deel van het totale curriculum wordt bestreken en dan nog met alleen meerkeuzevragen.' (Wijnstra, 1984: 10). Wat wel kan, volgens Cito, ten aanzien van de evaluatieve functie van de eindtoets, is wat ook in eerste instantie gebeurt: het vergelijken van scholen onderling. Vanaf de jaren '70 komt Cito met schoolrapporten waarin scholen zichzelf afgemeten zien ten opzichte van eerst regionale en later landelijke gemiddelden. En op de rangschikking volgt een beoordeling van scholen ten opzichte van elkaar: 'De laagste en hoogste 14% kregen als oordeel 'laag', respectievelijk 'hoog', de overige gemiddelden werden als 'gewoon' aangemerkt.' (Wijnstra,

1984: 17). In de verantwoording van 1996 wordt gesteld dat de eindtoets 'informatie levert ten behoeve van de evaluatie van het gegeven onderwijs' (Uiterwijk & Theunissen, 1996: 9).

Daartoe ontvangt een school een evaluatierapport waarmee de school gegevens heeft 'waarmee kan worden nagegaan welke onderdelen van het onderwijsprogramma dat jaar meer of minder effectief zijn geweest' (Uiterwijk & Theunissen, 1996: 14). De verantwoording van CVTE uit 2015 zet evaluatie neer als doel (CVTE, 2015: 14). Op basis van een zeer smalle set onderwijsdoelen wordt een breed kwaliteitsoordeel over de effectiviteit van scholen uitgesproken.

Door leerlingen in hun eigen cohort jaarlijks met elkaar te vergelijken, te ordenen, met behulp van percentielscores gelijkmatig te classificeren en die verdeling te koppelen aan score-intervallen is het lastig over de jaren heen zicht te houden op de ontwikkeling van de taal- en rekenvaardigheid zoals die met de eindtoets wordt gemeten. Van het monitoren van een longitudinale ontwikkeling van de kwaliteit van onderwijs is gezien het relatieve karakter van de eindtoets dan ook geen sprake, ook niet waar het gaat om taal en rekenen.

Hoewel van diverse kanten steeds geprobeerd is de breedte van onderwijsdoelstellingen onder de aandacht te brengen en gepoogd wordt de ontwikkeling van kinderen in een breder perspectief te zien, zorgt de eindtoets, en de belangen die eraan gekoppeld zijn voor wat betreft de plaatsing in het voortgezet onderwijs, voor een stevige verankering van taal en rekenen in het basisonderwijs. Zolang plaatsing in het voortgezet onderwijs wordt gebaseerd op toetsing van (een beperkte set van) taal en rekenen en de beoordeling van de kwaliteit van scholen gebaseerd wordt op toetsscores, zal de toets, en dan met name de onderdelen die in de toets aan bod komen, veel aandacht krijgen in het onderwijs zelf, in het toezicht op het onderwijs en in de waardering van het onderwijs (zie ook hoofdstuk 5).

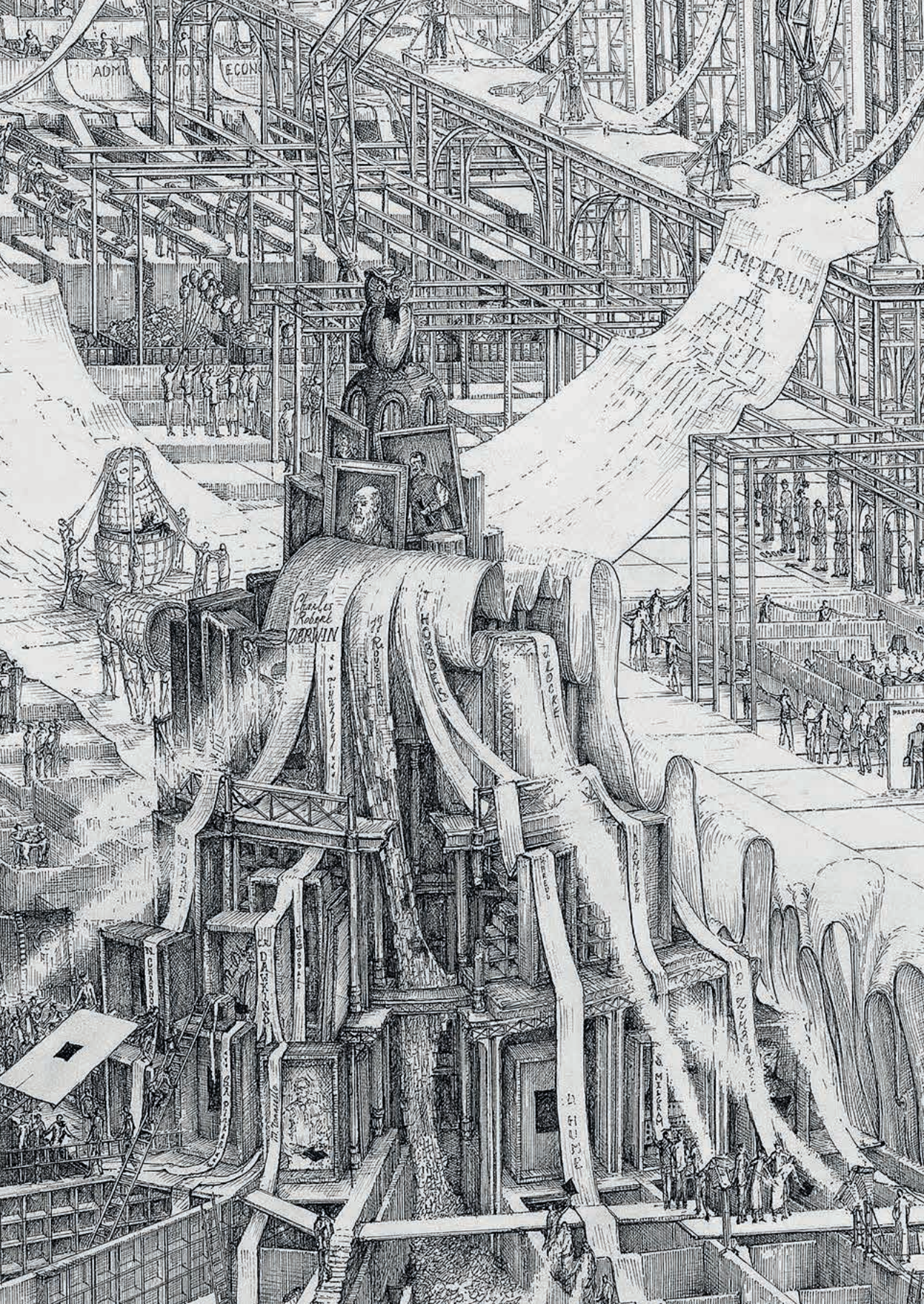
### 3.6 Samengevat

Nederland heeft, in vergelijking met de meeste landen in Europa, een stelsel dat vroeg selecteert voor een sterk gedifferentieerd voortgezet onderwijs. De vroege selectie in combinatie met een veelheid aan 'tracks' is een stabiel kenmerk van het Nederlandse onderwijsstelsel. Het funderend onderwijs heeft formeel als doel een ononderbroken ontwikkelingsproces te bieden met behulp van kerndoelen geformuleerd als streefdoelen, maar van een ononderbroken ontwikkelingsproces is door de allocatie aan het eind van het basisonderwijs feitelijk geen sprake. De referentieniveaus, vanaf 2011 vastgelegd in toetsen en toezicht, hebben voor een substantieel deel de streefdoelen vervangen. De overheid is met het toegenomen belang van de eindtoets zowel verantwoordelijk voor de allocatie als voor het toezicht op kwaliteit die afgelezen wordt uit de onderwijsopbrengsten in de vorm van toetsscores, gecorrigeerd voor achtergrondvariabelen. Subtestscores voor taal en rekenen worden gecombineerd tot één score waarmee alle leerlingen eenvoudig kunnen worden geordend van hoog naar laag scorend en waarbinnen relatief eenvoudig grenzen kunnen worden aangegeven die toelating tot de ene vorm van vervolgonderwijs onderscheiden van de andere vorm van vervolgonderwijs. Daarmee is voor alle betrokkenen enerzijds een eenvoudig en transparant systeem voor toewijzing naar vervolgonderwijs mogelijk en anderzijds een eenvoudig en transparant systeem voor het vaststellen van onderwijskwaliteit gerealiseerd. De vraag naar de geldigheid van deze simplificatie van de werkelijkheid is daarbij lastig te beantwoorden omdat voor het bepalen van de validiteit van tests geen eenduidige de-



finitie bestaat noch eenduidige criteria waarmee de validiteit vastgesteld kan worden. Beschouwen we de effecten van de beide formeel onderscheiden functies van de eindtoets tezamen, dan kan worden gesteld dat de eindtoets een belangrijk onderdeel is van een maatschappelijke competitie waarin zowel degenen die het onderwijs verzorgen als degenen die het onderwijs volgen gezamenlijk gebaat zijn bij zo hoog mogelijke resultaten met een nooit te winnen race als gevolg. Van het systeem gaat de suggestie uit dat de competitie ervoor zorgt dat alle betrokkenen het beste uit zichzelf halen. Maar het relatieve karakter van de eindtoets leidt tot vergelijkingen waarbij altijd een vast percentage gekenmerkt kan worden als 'zwak' en een vast percentage als 'excellent'. Want wat betreft de selectie wordt de stabiliteit van de selectie geborgd met de verdeelsleutel die samenhangt met de psychometrische uitgangspunten van de eindtoets. En voor de bewaking van de kwaliteit zorgt de relatieve meting in combinatie met een voortdurende druk tot het verbeteren van prestaties voor een permanente druk op scholen, terwijl het systeem van relatieve metingen tegelijkertijd voorkomt dat *alle* scholen een predicaat van 'voldoende' kunnen krijgen. Met dit relatief simplistische systeem van ordenen, classificeren en selecteren heeft de overheid de grip op onderwijs en samenleving in termen van sturing, structuur en kwaliteit stevig in handen terwijl de bewijsvoering voor de geldigheid, van de test zelf en van de test als beleidsinstrument, beperkt is. De validiteit van de eindtoets kan niet boven elke discussie worden verheven. De vraag naar de validiteit van de eindtoets kan uitsluitend worden beantwoord in relatie tot het meer algemene doel van onderwijs waarbij de vraag is of de eindtoets bijdraagt aan het dichterbij brengen of realiseren van het doel van onderwijs. Deze vraag kan niet beantwoord worden. Uit de beoogde functies van de eindtoets in termen van selectie en borging van kwaliteit ontstaat ook geen eenduidig beeld van de validiteit van de eindtoets. Tegelijkertijd wordt duidelijk dat de overheid door het opleggen van een verplichting tot het gebruik van een eindtoets in combinatie met een leerlingvolgsysteem in hoge mate bepaalt hoe voor het primair onderwijs de sturing, de structuur en de kwaliteit van onderwijs gezien wordt. De grondwettelijke vrijheid van onderwijs is wordt door de eindtoets in combinatie met LVS sterk aan banden gelegd zonder dat grondig onderzoek is gedaan naar 'who gains and who loses'.

Uit de analyse van wat het doel is van onderwijs in relatie tot samenleving ontstaat een dubbel beeld. Enerzijds rijst uit wetteksten een beeld op van vrijheid en autonomie van scholen als het gaat om de invulling en uitvoering van onderwijs. En anderzijds blijkt uit de rol en waarde van toetsing dat de overheid streeft naar inperking van die door de wet beoogde vrijheid. Duidelijk blijkt dat de overheid met de combinatie van de verplichte afname van eindtoets en LVS streeft naar grip en controle op onderwijs en de toeleiding naar de samenleving. De uitgangspunten voor beleid in relatie tot selectie en beoordeling van kwaliteit zoals die in en met de eindtoets hun beslag hebben gekregen, hebben elk hun eigen historie. In de volgende twee hoofdstukken worden beide functies in hun eigen historische context nader onderzocht. In het volgende hoofdstuk, hoofdstuk 4, staat selectie centraal. In hoofdstuk 5 staat het denken over kwaliteit centraal.



ADMIRALTY ECONOMY

IMPERIUM

Charles Rogée DE BRUXELLES

L'ÉCONOMIQUE

L'ÉDUCATION

L'AGRICULTURE

L'INDUSTRIE

L'ARTS

L'ÉTAT

L'ÉCONOMIQUE

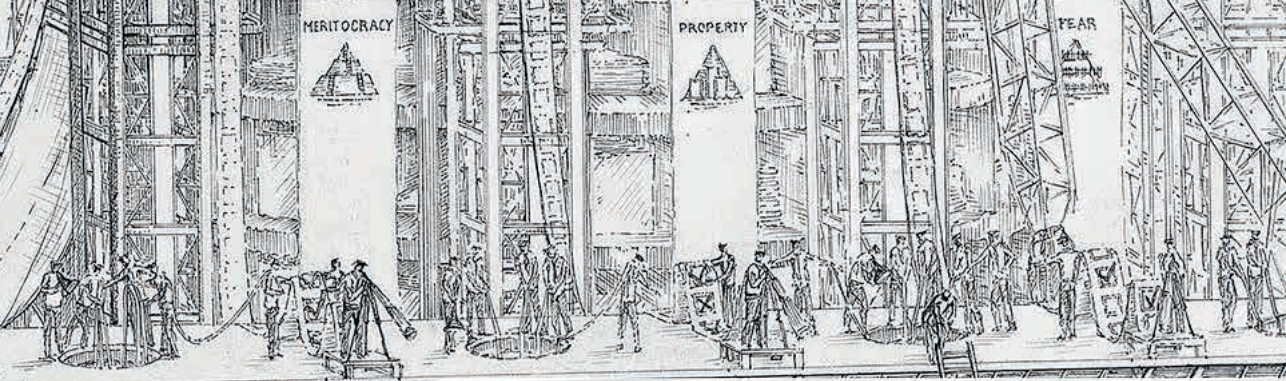
L'ÉDUCATION

L'AGRICULTURE

L'INDUSTRIE

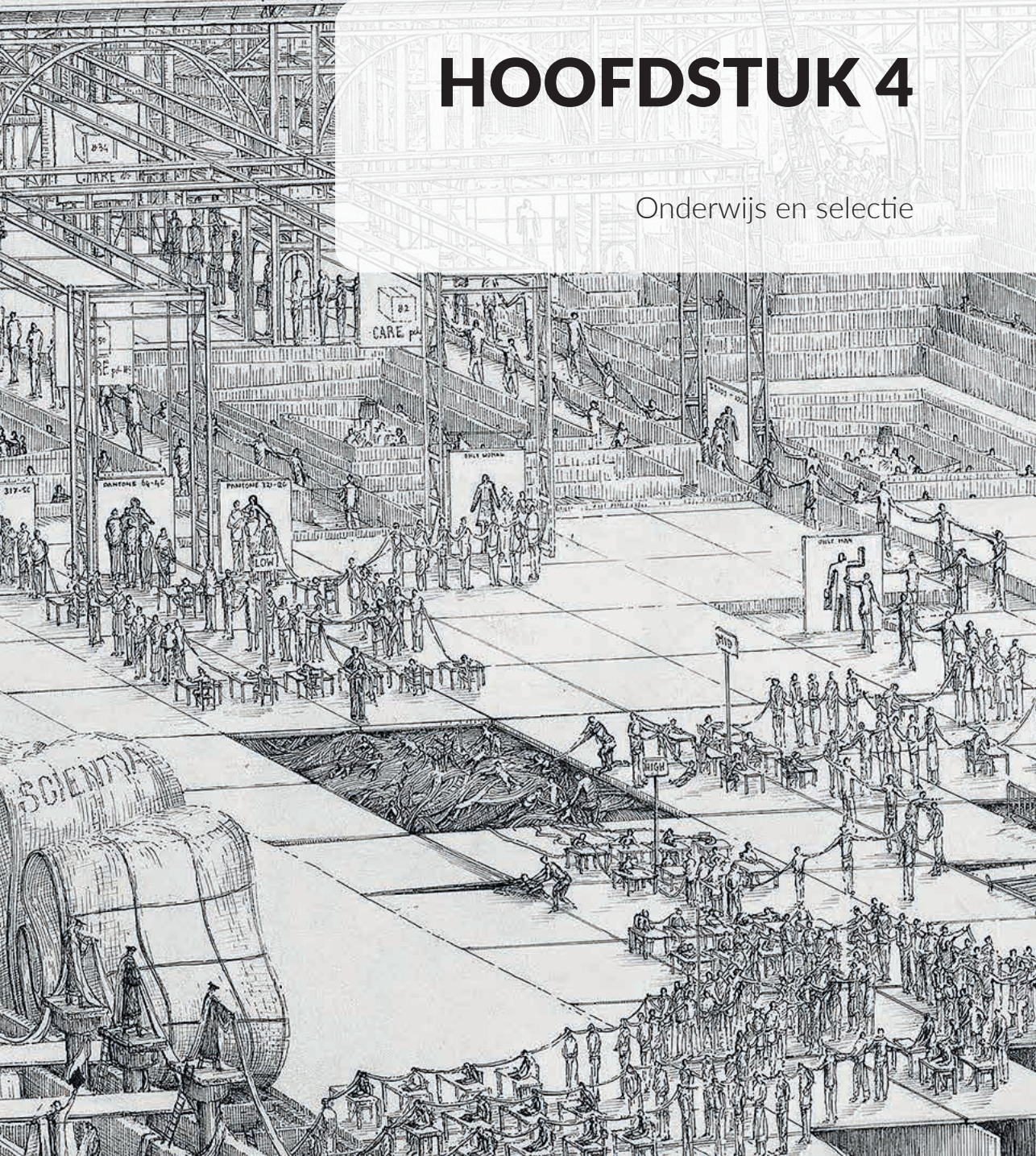
L'ARTS

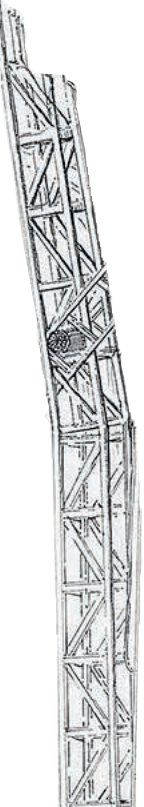
L'ÉTAT



# HOOFDSTUK 4

Onderwijs en selectie





## Hoofdstuk 4 Onderwijs en selectie

*'Beste mevrouw,*

*U zult zich niet eens meer mijn naam herinneren, u heeft er zoveel laten zakken. Ik heb echter vaak moeten denken aan u, uw collega's, het instituut dat u school noemt en aan de jongens die u 'afwijst'. U verwijst ons naar het land en de fabriek en u vergeet ons.'*

*Uit: Kinderen van Barbiana, 'Die rotschool van U; brief aan een onderwijzeres' (1967)*

Het 'aaneensluitingsvraagstuk', zoals de overgang van lager naar middelbaar onderwijs aan het begin van de vorige eeuw wordt genoemd, blijkt een beleidsthema met een lange historie. In Nederland is dit vraagstuk tegen verschillende achtergronden, in verschillende beleidscontexten en met verschillende bedoelingen, steeds een onderwerp op de politieke agenda geweest. Aanvankelijk gaat het vooral om de vraag wie verantwoordelijk moet zijn voor de toelating tot het middelbaar onderwijs. Vervolgens komt daar de vraag bij hoe die overgang het beste te regelen. Het vertrouwen in de wetenschapsonwikkeling meer in het algemeen en in de psychodiagnostiek in het bijzonder leidt tot aandacht voor meer gestandaardiseerde vormen van meten. In combinatie met de ontwikkelingen in het denken in termen van rendement en opbrengsten creëert dat ruimte voor een test die zich positioneert als enerzijds een oplossing voor het selectievraagstuk en anderzijds als een indicator voor opbrengsten en rendement. Vanaf 1968 krijgt toetsing een zelfstandige organisatie waar het denken over toetsen en examens verder kan worden ontwikkeld, het Cito (Centraal Instituut voor Toetsontwikkeling). Dit hoofdstuk start met een korte, historische schets van de ontwikkeling van het funderend onderwijs in Nederland en de toelating daartoe (4.1). De uitval uit het voortgezet hoger middelbaar onderwijs (v.h.m.o.) is lange tijd aanleiding voor een zoektocht naar instrumenten en beleidsmaatregelen om hierin verandering aan te brengen (4.2). Dit alles vindt plaats tegen de achtergrond van internationale ontwikkelingen gericht op het vaststellen en meten van intelligentie (4.3). In de periode van de wederopbouw wordt zorg om de uitval aangevuld met zorg over de benutting van talent, niet alleen om Nederland weer op te bouwen maar vooral ook om Nederland mee te kunnen laten komen in de vaart der volkeren, waarbij de gedachte is dat onderwijs daarbij een cruciale rol speelt. In de jaren vijftig en zestig komt er een economisch motief bij: doordat kinderen steeds langer naar school gaan, ontstaat druk op de begroting. Die vergrote instroom biedt mogelijk een potentieel aan extra talent maar dat kan alleen benut worden als dat potentieel 'gevonden' kan worden (4.4). De democratiseringsgolf eind jaren zestig roept onder andere op tot een andere kijk op onderwijs, een waar de ontplooiing van een individu meer centraal kan staan. Er ontstaat ook druk op vernieuwing van het onderwijsstelsel (4.5). En er is spanning tussen dit soort doelen voor onderwijs en meer arbeidsmarktgerelateerde onderwijsambities (4.6). De koerswijzingen van de overheid vanaf de jaren tachtig, waarin nadruk komt op effectiviteit en efficiëntie van beleid, hebben ook hun weerslag op onderwijs. In onderwijs vertaalt efficiëntie zich vooral in een verdere en strakkere streaming, waar selectie een belangrijk beleidsinstrument bij blijkt te kunnen zijn en effectiviteit van onderwijs vooral wordt ingebed in zorgen om en controle op de opbrengsten van onderwijs (4.7). Tegen de achtergrond van deze ontwikkelingen komt de eindtoets tot stand en groeit uit tot een instrument dat door het overgrote deel van de scholen ingezet wordt als onderdeel van de overgangprocedure van basis- naar voortgezet onderwijs (4.8).

#### 4.1 Ontwikkeling van onderwijs en toelating

Met de eerste onderwijswetten uit 1801, 1803 en 1806 begint het tijdperk waarin onderwijs toegroeit naar een vorm van collectieve voorziening. Belangrijke kenmerken hiervan zijn dat de kwaliteit van het verstrekte is gestandaardiseerd, dat deugdelijkheidseisen kunnen worden gesteld en dat de toegankelijkheid van die voorziening voor de bevolking is gegarandeerd (Wansink, 1992: 43). Een collectieve voorziening heeft de aanwezigheid en betrokkenheid van een staat nodig (Wansink, 1992: 43). In zijn 'Schets van het Nederlandse schoolwezen' stelt (onderwijs)jurist en statisticus Idenburg dat de bemoeienis van de staat met het onderwijs al van eerder dateert, blijkend uit de 'Algemene Schoolordre' van 1625. Een document waarin de staat zich volgens Idenburg 'ten volle openbaart als opper-pedagoog' (Idenburg, 1960: 31). Aan het einde van de 18e en aan het begin van de 19e eeuw, ten tijde van de Bataafse Republiek, is onderwijs een belangrijk instrument bij de vorming van de eenheidsstaat die dan tot stand komt, gebaseerd op een scheiding van kerk en staat. In het Volksonderwijs, de voorloper van lager onderwijs, staat vooral de vorming tot burger van de nieuwe republiek voorop: de opvoeding 'tot volgens algemeen christelijke beginselen levende Nederlanders' (Wansink, 1992:44). Als de rol van de staat begin 19e eeuw in de eerste onderwijswetten wordt vastgelegd blijkt uit de formulering de betrokkenheid van de staat bij het onderwijs: 'Het onderwijs is een voorwerp van de aanhoudende zorg van de regering', de formulering die ook nu nog in artikel 23 lid 1 van de Grondwet staat. In Nederland komt voor het eerst een verplichting tot het volgen van Lager Onderwijs na de invoering van de Leerplichtwet in 1900. De bemoeienis van de overheid met de inhoud van het onderwijs dat meer en meer gericht wordt op de 'opvoeding tot deugzaamheid' en zo te werken aan een uniforme en homogene cultuur op liberale grondslag, stuit op verzet. Gelovigen maken zich in de 19e en begin 20e eeuw hard voor 'eigen' onderwijs. Met de zogenaamde 'pacificatie' eindigt in 1917 formeel de schoolstrijd. Vanaf dat moment zijn bijzonder onderwijs, onderwijs op confessionele basis, in de financiering gelijkgeschakeld aan openbaar onderwijs en in ruil daarvoor wordt zeggenschap en invloed van de kant van de overheid geaccepteerd. De grens tussen overheid en onderwijs is er vanaf dan een die onder een vergrootglas ligt en bewaakt wordt door de scholen die zoveel mogelijk zelf willen bepalen hoe ze aan wie wat onderwijzen en door de overheid die zich verantwoordelijk stelt voor de deugdelijkheid van het onderwijs.

##### **Toelating**

Idenburg stelt dat onderwijs in het begin vooral bedoeld was 'om 'het lagere volk' op te heffen uit de 'sociale verwording', waarin het was neergezonken en welke tot een gevaar voor de samenleving dreigde te worden, maar het denkbeeld van sociale mobiliteit was eraan vreemd. Dat gold voor alle onderwijs. Het werd beheerst door de gedachte der continuïteit. De sociale herkomst der leerlingen en hun bestemming waren, standsmatig gezien, vanzelfsprekend dezelfde.' (Idenburg, 1960: 153). De kinderen van de hogere stand gingen niet naar school maar kregen 'huize-lijke' opvoeding (Idenburg, 1960: 143). De standen zijn een gegeven en 'vrijwel niemand dacht anders dan dat deze orde bewaard moest worden' (Idenburg, 1969: 149). Tot ver in de 20e eeuw blijft het denken over onderwijs verbonden met het denken in bestaande maatschappelijke structuren: het onderwijsstelsel is in lijn met de opbouw van de samenleving. Er is voor iedere stand een 'eigen' vorm van onderwijs. De Grondwet van 1815 maakt onderscheid tussen 'hoog' en 'laag' onderwijs maar dat heeft een betekenis die met de standensamenleving samenhangt: het 'hoge' onderwijs was voor de opleiding van de 'geleerde stand'. Dat maakt dat de Latijnse scholen, de latere gymnasia, gezien worden als 'hogere' onderwijs. Dat onderwijs was bedoeld voor de

elite. Voor de 'eenvoudige burgerstand' en de 'gezetten burgerstand' is in die tijd geen passend onderwijs na de lagere school (De Rooy, 2018: 123). Als leider van zijn tweede kabinet (1862-1866) is Thorbecke degene die daarmee aan de slag gaat. Hij komt met specifiek onderwijs ná het lager onderwijs, bedoeld voor de 'talrijke burgerij', naast het gymasiaal onderwijs voor de 'hogere' stand. Zijn wet uit 1863 kan worden gezien als de eerste, meer concrete invulling van het begrip 'middelbaar onderwijs' zoals dat in de Grondwet van 1814, vooruitlopend op de werkelijkheid, al wel is opgenomen, maar dan nog niet is uitgewerkt. Voor de vorming van de 'nijvere burgerij' onderscheidt Thorbecke drie varianten: een tweejarige burgerschool, een driejarige en een vijfjarige hogere burgerschool (De Rooy, 2018: 126). Latijnse scholen en de gymnasia vallen buiten de wet en worden niet gezien als middelbaar onderwijs; zij blijven beschouwd worden als 'hoger' onderwijs, bedoeld als opstap naar de universiteit en bedoeld voor hen voor wie dat qua afkomst en financiën passend is. Zoals Idenburg het formuleert: 'Zo werd het stelsel van scholen aan de behoefte der burgerij aangepast. Men denkt hiërarchisch. De scholen waren bestemd ter bevestiging van de sociale status van het gezin, niet voor opklimming tot een hogere sociale laag.' (Idenburg, 1960: 155). Het middelbaar onderwijs is 'hooger volksonderwijs', bedoeld als eindonderwijs gericht op algemene vorming als vervolg op het lager onderwijs. Hiermee wordt het middelbaar onderwijs vooral 'iets voor de 'burgerstand' voor wie de lagere school niet langer voldoende was' (De Rooy, 2018: 125). In zijn boek 'Een eeuw middelbaar onderwijs 1863-1963' noemt Bartels de grote vrijheid die Thorbecke beoogt met zijn nieuwe wet 'opvallend' (Bartels, 1963: 64): 'Thorbecke wilde de grootst mogelijke vrijheid toestaan: de wet bevatte geen andere aanwijzing omtrent het leerplan dan de opnoeming der vakken.' (Bartels, 1963: 64). Hij citeert Thorbecke: 'Op dit gebied komt geene alleenheerschappij van een eenvormig model te pas; in het belang van het middelbaar onderwijs zelf moet verscheidenheid van behoefte en inzicht zich ongehinderd kunnen doen gelden.' (Bartels, 1963: 64). Die ruimte staat Thorbecke niet alleen voor ogen voor de scholen maar ook voor de leerlingen. Wat hem betreft zijn leerlingen niet verplicht 'andere lessen bij te wonen, dan die hij goed vindt te volgen'. Met datzelfde gevoel voor ruimte acht Thorbecke 'toelatingsexamens onder alle omstandigheden verwerpelijk' (Bartels, 1963: 94). Bij de behandeling van de Staatsbegroting voor 1872 stelt hij duidelijk teleurgesteld: 'Hetgeen ik niet wilde is ingeslopen; in de jongste jaren is de heerschappij wederom gekomen aan een schoolsch begrip van dwang, waardoor het onderwijs, met name het middelbaar onderwijs, waarmede wij nu te doen hebben, eene soort van fabriek zou worden. Mijne verklaarde meening was steeds, dat het genot van het onderwijs aan de hoogere burgerscholen vrij moet zijn gelijk aan de Universiteiten.' (Bartels, 1963: 95). De door Thorbecke beoogde ruimte in programmering en vrijheid voor het vormen van eigen leertrajecten blijkt in de praktijk weerbarstig en stuit op politieke weerstand. In 1873 komt er een Koninklijk Besluit waarbij de toelating en bevordering worden aangepast: 'In het vervolg zou niemand tot enige klasse der school worden toegelaten dan na het afleggen van een openbaar examen, afgenomen door den directeur en de leeraren, waaruit blijke, dat hij de kundigheden bezit, vereischt om het onderwijs in die klasse met vrucht te kunnen bijwonen. Hetzelfde geldt ten aanzien van de bevordering naar een hogere klasse.' (Bartels, 1963: 98). Ook komt er vanaf het schooljaar 1916-1917 een min of meer uniform leerplan voor de hbs (Bartels, 1963: 75). Vanwege de grote uitval en de twijfel aan de adequaatheid van de diverse toelatingsexamens van de verschillende middelbare scholen wordt besloten meer ruimte te geven voor het advies van het hoofd van de lagere school. Vanaf 1902 mogen middelbare scholen alleen op deze verklaring afgaan en hoeven geen toelatingsexamen af te nemen maar niet veel scholen maken van die mogelijkheid gebruik (De Regt, 2004: 299). In

de Lager Onderwijswet van 1920 wordt deze mogelijkheid een verplichting. Als het hoofd van de lagere school een positieve verklaring van geschiktheid aflegt, is een toelatingsexamen niet toegestaan. 'Het gevolg van deze bepaling was, dat de toelating op grond van de "verklaring" de gewone loop van zaken werd; slechts bij uitzondering werd een examen afgenomen.' (Bartels, 1963: 104).

Rond 1920 kan wel gesproken worden van een succesvolle introductie van 'hoger burgeronderwijs' maar is het idee van 'eigenaarschap voor de leerling' waarmee Thorbecke de introductie van dit nieuwe onderwijs heeft willen omgeven, ingeperkt en ingekaderd met leerplannen en welomschreven regels voor toelating en overgang.

## 4.2 Toelating en uitval

Het eerste, grootschalige onderzoek naar uitval en zittenblijven verschijnt in 1926 en is uitgevoerd door de Hongaars-Nederlandse psycholoog Révész. Hij heeft de schoolloopbaan van 3000 hbs-leerlingen onderzocht en stelt vast dat 'slechts de helft' door het eindexamen komt. En van deze helft heeft twee derde de school doorlopen, zonder een keer te zakken. De uitkomsten zijn op dat moment geen opzienbarend nieuws en doen weinig stof opwaaien omdat de hoogte van de uitval in 1926 een stabiel en bekend gegeven is. Statistische gegevens wijzen uit dat al in 1865 de situatie niet anders was (Deen, 1969: 16). Bovendien loopt het onderzoek van Révész een beetje achter de feiten aan, omdat hij onderzoek heeft gedaan bij leerlingen die dan nog wél een toelatingsexamen hebben gedaan, terwijl bij het verschijnen van zijn onderzoek er geen toelatingsexamens meer zijn omdat het hoofd van de lagere school sinds 1920 over de toelating gaat. Révész zelf ziet de hoge uitval vooral als een 'schiftingsprobleem' en ziet de oplossing in een meer 'doeltreffende zeef' om zo het probleem van de uitval te kunnen oplossen. Hij brengt het probleem terug tot een selectie-vraagstuk. Deen stelt dat Révész denkt dat 'met het vinden van een doeltreffende zeef, die de voor het middelbaar onderwijs geschikten uit de massa der ongeschikten zeeft, het rendementsprobleem der middelbare school op te lossen zou zijn' (Deen, 1969: 16). Die denkrichting van Révész sluit niet aan op de vanaf 1920 ingezette beleidslijn waarmee besloten is het advies over toelating juist niet instrumenteel op te lossen maar in handen te leggen van in eerste instantie het hoofd van de lagere school, omdat de lagere school zich door de jarenlange ervaring met de kinderen het beste een beeld kan vormen van de geschiktheid van een kind. Dit stuit op bezwaren en verzet bij het middelbaar onderwijs: 'De verantwoordelijkheid ... was gelegd op de onderwijzer van de lagere school, die deze niet alléén kan dragen, omdat een objectieve maatstaf voor de beoordeling ontbreekt en het bovendien voor onderwijzers, die het middelbaar onderwijs niet kennen, moeilijk is te oordelen over de bekwaamheid en geschiktheid voor het volgen van middelbaar onderwijs.' (Bartels, 1963: 104). Er komt een Staatscommissie, de Commissie Rutgers, die zich moet buigen over het probleem van de toelating, met name omdat de uitval en het aantal zittenblijvers na de wetswijziging van 1920 toenemen. Die Commissie beveelt aan artikel VI te laten vervallen en het aan gemeentebesturen en besturen van hogere burgerscholen over te laten te beslissen op welke wijze zij de toelating willen regelen (Bartels, 1963: 104). Het advies leidt tot een nieuw Koninklijk Besluit (KB) in 1928 dat de toelating tot de 5-jarige hbs als volgt aanpast: Het hoofd van de lagere school geeft een advies, het hoofd van de middelbare school onderzoekt de leerling schriftelijk op zijn kennis en geeft ook een advies, zijn de adviezen in lijn met elkaar dan is er geen probleem, verschillen de adviezen dan komt er een mondeling onderzoek in handen van het middelbaar onderwijs dat dan



komt tot een eindoordeel, al dan niet weer in overleg met het hoofd van de lagere school. Deen noemt dit KB een compromis waarbij 'enerzijds de verantwoordelijkheid voor de toelating wel aan het middelbaar onderwijs werd opgedragen, maar dat anderzijds een nauwe samenwerking tussen het lager en het middelbaar onderwijs met betrekking tot de toelating bevorderde' (Deen, 1969: 105).

#### 4.2.1 Verbeteren van de aansluiting

Het KB van 1928 maakt echter geen einde aan de discussie over de aansluiting. Er komen eind jaren twintig twee nieuwe rapporten uit over de manier waarop de aansluiting het best is aan te pakken: een rapport van de door de gemeente Den Haag ingestelde 'Commissie Verdenius lager onderwijs', die moet adviseren over het probleem van de uitval in relatie tot de selectie en een rapport van het Nutsseminarium voor Pedagogiek van de universiteit van Amsterdam, onder leiding van de pedagoog en hoofd van het Nutsseminarium, Philip Kohnstamm. De Commissie Verdenius heeft lange tijd nodig voor een advies (van 1924-1928) en blijkt uiteindelijk te komen met lastig uitvoerbare oplossingen. Ze pleit onder andere voor de inrichting van centrale vijfde, zesde en zevende leerjaren waarop een commissie moet toezien, daarmee de selectie feitelijk vervroegend én verlengend. De Commissie Verdenius kijkt ook naar meer wetenschappelijke, rationele oplossingsrichtingen om het 'subjectieve oordeel van de schoolhoofden te objectiveren' (Deen, 1969: 19) en stelt dat psychologisch onderzoek, in die tijd psychotechniek genoemd, deel kan uitmaken van het selectieproces. Zowel Révész als de Commissie Verdenius zien de oplossing van de uitval in het verbeteren van de selectie. Ze stellen niet de vraag wat het v.h.m.o. zelf bijdraagt aan 'de mislukking van zoveel van de toegelaten leerlingen' (Deen, 1969: 19). Dat doet het Nutsseminarium in zijn advies wel. In het eerste rapport uitgebracht door het Nutsseminarium in 1928 betreffende het vraagstuk 'De Aaneensluiting tusschen Lager en Middelbaar (Gymnasiaal) Onderwijs' laat het Nutsseminarium zich bij monde van Kohnstamm en Van Veen kritisch uit over het middelbaar onderwijs zelf. Kohnstamm en Van Veen pareren de kritiek vanuit het middelbaar onderwijs op de slechte aansluiting met de tegenvraag 'Met welk recht mag men in het v.h.m.o. verwachten dat de lagere school haar onderwijs zal afstemmen op één-twintigste deel van haar leerlingen?' Kohnstamm en Van Veen stellen dat de aansluitingsproblemen vooral worden veroorzaakt door de totaal verschillende aanpak wat betreft didactiek en inrichting van het lager en het middelbaar onderwijs. Dat zien zij als een in de eerste plaats door het onderwijs zelf op te lossen probleem, want, zo vat Deen hun denken samen: 'Het schoolsysteem is er ter wille van het kind' en het schoolsysteem heeft de taak 'ieder kind op de voor het kind juiste plaats te brengen' (Deen, 1969: 21). En vanuit een meer pedagogisch perspectief pleiten zij in hetzelfde rapport voor een volledige herstructurering van het onderwijs omdat 'de school', het lager en middelbaar onderwijs tezamen, 'één organisch geheel dient te zijn, waarvan de delen in voortdurend contact staan': een eenheidsschool die de verantwoordelijkheid aanvaardt voor de opgenomen leerling (Deen, 1969: 20-21).

#### 4.2.2 Tests en selectie

Ondanks hun pleidooi voor een grondige aanpak van het onderwijsstelsel gaan Kohnstamm en Van Veen in 1928 wel mee in het verkennen van meer wetenschappelijke, rationele oplossingsrichtingen zoals de psychotechniek die wellicht kan bieden voor het verbeteren van selectie. Het rapport doet naast meer fundamentele ook praktische voorstellen. Het terugvallen op een toelatingsexamen uit te voeren door het v.h.m.o., is geen reële optie meer na het onderzoek van

Révész (Deen, 1969: 35). Dat heeft onvoldoende resultaat laten zien in het kader van uitval. Kohnstamm en Van Veen stellen in het rapport wel mogelijkheden te zien in zowel de 'intelligentietest als de scolastic test (schoolprestatietoets of achievementtest)' en vooral in de combinatie ervan (Deen, 1969: 30). In die tijd staan Van Veen en Kohnstamm nog open voor de inzet van dit soort instrumenten en zien de potentiële waarde daarvan wel in: 'De mogelijkheid om buiten de leerstof om de begaafdheid der kandidaten vast te stellen, wordt door de schrijvers dan nog niet in twijfel getrokken.' (Deen, 1969: 26). In 'Op de juiste plaats' (1995) stelt psycholoog Haas dat het Amsterdamse rapport gekarakteriseerd kan worden als meer psychologisch georiënteerd door te stellen dat toelatingsexamens in de bestaande vorm, zowel de brede kennistoets als het schrijven van een opstel, ontoereikend zijn om de geschiktheid van individuele leerlingen voor het volgen van voortgezet onderwijs te bepalen, maar wel mogelijkheden zien voor het toepassen van intelligentietests en schoolvorderingentests (scolastic tests) (Haas, 1995: 162). Dit in weerwil van het feit dat de eerste experimenten met intelligentiemetingen in Nederland uitgevoerd door medewerkers van het Nutsseminarium zelf, niet veelbelovend zijn. Want onderzoek naar toepassing van intelligentieproeven voor de voorspelling van toekomstig schoolsucces door Kuiper en Van Lennep van het Nutsseminarium heeft dan geleid tot teleurstellende resultaten: leerlingen die op de intelligentieproeve slecht presteren, blijken in de praktijk van het voortgezet onderwijs wel over te (mogen) gaan en andersom. 'Er blijkt, dat de schoolrangorde een betere prognose geeft dan de test.' (Kuiper, 1927: 334). Kuiper en Van Lennep komen tot de slotsom dat de intelligentietest in de aansluitingsprocedure slechts een bescheiden plaats kan innemen (Deen, 1969: 32).

Op zowel 'het Haagse' als 'het Amsterdamse' rapport komt de nodige kritiek, onder andere van testpsycholoog Luning Prak, een groot voorstander van de afname van groepstests voor intelligentie, die in een artikel in *Paedagogische Studien* beide rapporten negatief beoordeelt. Hij betitelt ze als naïef en elkaar tegensprekend met betrekking tot de voorspelbaarheid van succes op de middelbare school op basis van resultaten behaald bij het kerstrapport en het toelatingsexamen. Luning Prak betoogt dat 'zonder het toepassen van psychologische tests het onderwijs zelf "als een weinig economische en zeer tijdrovende" zeef zou blijven functioneren' (Haas, 1995: 179). Mede naar aanleiding van zijn kritiek wordt er van overheidswege een nieuwe commissie ingesteld, de Commissie Bolkestein, waarvan ook Kohnstamm, Van Veen en Verdenius deel uitmaken (Haas, 1995:162). Op 14 mei 1934 komt de Commissie met haar rapport. De commissie wil voorkomen dat het voortgezet onderwijs zelf als zeef fungeert voor ongeschikten: geschiktheid moet vooraf worden bepaald. De commissie pleit voor een toelating bestaand uit een tweetal onderzoeken. Het eerste onderzoek moet gebeuren door een in januari van ieder schooljaar te benoemen commissie van leraren en onderwijzers. Dit onderzoek bestaat uit een schriftelijk examen en onderzoekt a) de mate, waarin de verstandelijke aanleg ontwikkeld en gevormd is en b) de vertrouwdheid met de voor het volgen van het middelbaar onderwijs vereiste begrippen op het gebied der Nederlandse taal, rekenen, aardrijkskunde en geschiedenis. Vlak voor de zomervakantie volgt dan een tweede, nader onderzoek naar de vaardigheid in het gebruiken van de kennis in de bovengenoemde leervakken (Verseput, 1956). De Commissie Bolkestein ziet het stilleesstuk als bruikbaar alternatief voor een intelligentietest. Een stilleesstuk is een korte (vertellende of informatieve) tekst waarbij de leerling een aantal vragen voorgelegd krijgt waarvan de antwoorden na aandachtig lezen uit de tekst zijn af te leiden. De Commissie kent deze vorm van toetsing een belangrijke rol toe in het onderzoek naar de schoolgeschiktheid (naast

een schoolvorderingentoets en het advies van de Lagere School). Het rapport vormt de basis voor de toelatingsregeling van 1938. Het stillezen krijgt niet alleen een plek bij het onderdeel Nederlandse Taal 'maar ook Rekenen, Aardrijkskunde en Geschiedenis worden in de vorm van stilleesoefeningen gepresenteerd' (Deen, 1969: 53). Het geschiktheidsonderzoek is 'volgens de stilleesmethode gedacht. Daardoor heeft dit gehele "eerste onderzoek" veel meer het karakter van een intelligentie-onderzoek dan van een schoolkennis-onderzoek gekregen.' (Deen, 1969: 53). Maar wel een intelligentie-onderzoek in een andere geest dan de testpsychologie wellicht had gehoopt: 'Toen het rapport verscheen bleek de testpsychologie aan het kortste eind te hebben getrokken: geen klassikale intelligentietests, wel schoolvorderingentests en een weliswaar psychologisch onderbouwd, maar door onderwijzers door te voeren geschiktheidsbeoordeling. Voor insiders was het duidelijk dat vooral Kohnstamm zijn stempel had gezet op het rapport.' (Van Strien, 2003). De adviezen van deze commissie worden niet helemaal een op een overgenomen in het KB van 1938 over de toelating. Wel worden enkele cruciale elementen overgenomen zoals de 'stilleestest' ter bepaling van het algehele geschiktheidsniveau en de schoolvorderingentest voor het vaststellen van de parate kennis. Het besluit reflecteert daarmee de verschuiving van de aandacht en het vertrouwen in het testen van parate kennis naar het inschatten van de capaciteiten van kinderen. Het idee van selectie krijgt meer het karakter van 'determinatie' of 'verwijzing' (De Regt, 2004: 301). Het Koninklijk Besluit van 1938 kwam in de ogen van Haas nog 'het dichtst bij de idealen van de pedagogen' al was het niet echt gelukt 'een adequaat beleid voor de aansluitingsproblematiek te voeren' (Haas, 1995: 163). Dit KB wordt uiteindelijk door het uitbreken van de Tweede Wereldoorlog niet concreet ingevoerd. Deen stelt desalniettemin: 'Dat de invloed van het rapport groot is geweest, mogen we afleiden uit het feit dat het nog in het toelatingsbesluit van 22 juli 1965 (...) vrijwel op de voet werd gevolgd.' (Deen, 1969: 54).

De Regt wijst erop dat de discussies over de moeizame aansluiting tussen het lager en het middelbaar onderwijs en de rol van een toelatingsexamen tot in de jaren vijftig voorbijgaan aan het feit dat dit probleem slechts betrekking heeft op een fractie van alle leerlingen die het lager onderwijs verlaten, namelijk alleen voor degenen die voorbestemd waren voor het middelbaar onderwijs. Middelbaar onderwijs is in die tijd nog altijd het privilege van de hogere en middenklasse: 'Alleen zij werden op de lagere school voorbereid op het toelatingsonderzoek. Sommigen van hen gingen naar zogenaamde opleidingsscholen die uitgebreidere leerstof aanboden, sommigen gingen op gewone lagere scholen naar een speciale opleidingsklas of kregen in een klein groepje afzonderlijk les om hen te trainen voor het toelatingsexamen.' (De Regt, 2004: 300). Voor kinderen uit de lagere klassen die 'goed konden leren' was de mulo bestemd en daar was geen toelatingsexamen en dus geen extra voorbereiding voor nodig (De Regt, 2004: 300).

### 4.2.3 Rangorde of cesuur

Naast de zoektocht naar een passend instrumentarium voor de overgang van lager naar middelbaar onderwijs is er tevens aandacht voor het vraagstuk van de cesuur: de grens voor aanname of afwijzing, zakken of slagen, overgaan of blijven zitten. Het draait daarbij om de vraag of een cesuur relatief of absoluut moet worden gesteld. In het eerste geval gaat het om een norm waarbij scores vergeleken worden met de scores van andere kandidaten zoals bij de intelligentietests algemeen gebruik is en in het tweede geval gaat het om een norm waarbij te behalen doelstellingen het uitgangspunt vormen.

Uit het onderzoek van Révész uit 1926 blijkt bijna een kwart van de leerlingen het eerste jaar in het v.h.m.o. niet door te komen. Kohnstamm en Van Veen leggen direct een verband met de Gaussische normaalverdeling. In de optiek van Kohnstamm en Van Veen vertonen de scorepatronen van de uitval uit het v.h.m.o. deze verdeling: 25% valt af, 50% scoort rond het gemiddelde en 25% zit daar boven. In mededeling 3 van het Nutsseminarium constateren Kohnstamm en Van Veen dat van de 50% die in de middelbare school de eindstreep niet halen, al 25% in de eerste klasse is gestruikeld. Dit vinden zij niet wenselijk en niet efficiënt: 'èn ter wille van 't onderwijs, èn terwille van deze leerlingen zelf, èn ter wille van een economisch onderwijsbeheer is het van belang om deze 25% bij voorbaat te kunnen elimineren.' (Deen, 1969: 55). Uit het feit dat zij er in eerste instantie voor pleiten de 25% slechtste leerlingen de toegang tot het v.h.m.o. te ontzeggen, blijkt dat zij de normaalverdeling volgens de curve van Gauss een aanvaardbaar criterium achten voor de groepering van leerlingen (Deen, 1969:29). Als het doel is om zo snel en goed mogelijk 25% te elimineren is het zaak te streven naar een instrument dat een zo zuiver mogelijke normaalverdeling waarborgt zodat op voorhand de zwakste 25% de toegang tot het middelbaar onderwijs kan worden ontzegd. Kohnstamm en Van Veen willen toe naar een 'rangorde, die het mogelijk maakt de leerlingen op grond van een spreidingsmaatstaf te selecteren' (Deen, 1969: 55). Tegelijkertijd maakt Kohnstamm zich in mededeling 3 maar later ook in mededeling 4 zorgen of de selectie in het voortgezet onderwijs zelf dan wel stopt of dat het voortgezet onderwijs alsnog weer overgaat tot het afstoten van een kwart van de leerlingen. Hij realiseert zich dat middelbare scholen hun cijferbepaling vergelijkend doen en zich meestal niet baseren op objectieve normen. Dat zou betekenen dat selectie na toelating op dezelfde manier verder gaat en dat is niet zijn bedoeling: 'Voor de leerlingen, die een school eenmaal heeft opgenomen, behoort zij zich ook verantwoordelijk te voelen. De tijden waarin een school vrijuit ging, die zichzelf als een "zeef" beschouwde, die langzamerhand de "ongeschikten" uitzeeftde, zijn voorbij.' (Deen, 1969: 56).

In het middelbaar onderwijs blijkt een kwart van de leerlingen het predicaat 'ongeschikt' te krijgen. Dit is de gebruikelijke gang van zaken gedurende het gehele onderwijs; per test, per vak en bij iedere overgang. Deze wetmatigheid beschrijft Posthumus in 1940 in een artikel in *De Gids* (Posthumus, 1940: 24-42), wat sindsdien in de wandelgangen 'de wet van Posthumus' is gaan heten. Posthumus stelt: 'De gymnasia en hogere burgerscholen zijn sinds 1875 volkomen veranderd. Vakken, leerplannen, urentabellen, zijn gekomen en gegaan. Het aantal leerlingen is veeleer vervoerd. De meisjes hebben hun intrede gedaan. De maatschappelijke samenstelling der schoolbevolking is geheel veranderd; de wereldoorlog heeft mens en maatschappij vervormd; het gezinsleven, de sport, de vermaken, het is alles anders geworden. Maar de onvolledige statistieken die er zijn, tonen, dat het gedeelte uitvallers der middelbare scholen gelijk bleef.' (Posthumus, 1940:37).

Mededeling 15 (1930) beschouwt Deen als het keerpunt in het denken van Kohnstamm over cesuurbepaling. Kohnstamm laat cesuurbepaling gebaseerd op rangorde achter zich en zoekt naar cesuren met een meer kwalitatief karakter (Deen, 1969: 57). Dit is vooral ingegeven door het feit dat de onderzoeken van het Nutsseminarium laten zien dat er lang niet altijd een normaalverdeling verwacht kan worden bij de resultaten van leerlingenwerk (stillezen). Daar kan namelijk gegeven de aard van onderwijs geen sprake zijn van een 'toevalsresultaat', wat wel onderliggend is aan de theoretische onderbouwing van een normaalverdeling. Het Nutsseminarium is van mening

dat niet kan worden uitgegaan van een statische interval- of ordinale schaal waarop prestaties volgens de normaalverdeling van Gauss zijn geordend (Deen, 1969: 59). Kohnstamm 'blijft met overtuiging bepleiten om geen interval- of ordinale schaal te hanteren' (Nelissen, 2014: 23). Er wordt gezocht naar andere en verdere verfijningen van procedures. In zijn 'Inleiding tot de studie der paedagogische psychologie' uit 1957 stelt ook Langeveld dat de Gaussische verdeling in de context van onderwijs geen na te streven ideaal zou moeten zijn. In onderwijs is de zoektocht volgens Langeveld gericht op het onderscheiden van 'beslist voldoende' en 'beslist onvoldoende'. Bij selectie is het streven gericht op het onderscheiden van de duidelijke voldoende van de duidelijke onvoldoendes en een scherpe, goed te onderscheiden cesuur tussen beide groepen is daarbij van belang: 'Maar de kromme, die dan als ideaal voor den dag zou komen, is het omgekeerde van de kromme van Gauss. We krijgen niet de top in het midden met sterke glooiingen naar links en rechts, maar een duidelijk dieptepunt in het midden' (Langeveld, 1957: 499). Het uitgaan van 20 à 25 % onvoldoendes 'legt daarmee het verloop van de kromme vast, hij kan op die manier een kromme van het type van die van Gauss forceren' (Langeveld, 1957: 499).

Pedagogen en psychologen als Kohnstamm, Van Veen, Langeveld en Stellwag zijn niet de enigen die worstelen met cesuurbepaling en normeren. Ook Idenburg besteedt hieraan aandacht in zijn 'Schets van het Nederlandse Schoolwezen'. In 1960 is de wetmatigheid zoals Posthumus die in 1940 heeft vastgesteld volgens hem nog altijd geldig: ongeveer een kwart van de leerlingen valt uit in het middelbaar onderwijs. Dat ligt volgens Idenburg niet aan het feit dat een vierde deel van de leerlingen niet aan de eisen voldoet, maar, zo stelt Idenburg: 'dat is de opzet van de schifting; zij is geen natuurwet, die uit de ervaring is geput, maar een besluit dat op de ervaring wordt toegepast; zij is geen eigenschap van de beoordeelde maar een gewoonte van de beoordelende groep'. Er zijn nog altijd geen objectieve maatstaven die in het onderwijs gehanteerd worden: 'Er zijn geen vastgestelde eenheden, waarin wij vorderingen in de diverse schoolvakken kunnen uitdrukken. Schoolcijfers zijn dit niet. Ze geven alleen rangnummers aan. ... Het werk van de groep bepaalt de maatstaf.' (Idenburg, 1960: 511). Idenburg stelt dat 'deze leer een dwangmatigheid in ons schoolwezen blootlegt van niet geringer macht dan de mobiliteitsdrang, welke wij eerder signaleerden' (Idenburg, 1960: 512).

Hoewel het inzicht begin zestiger jaren zeer zeker aanwezig is dat het uitgaan van een relatieve norm en een normaalverdeling niet goed past bij de context van onderwijs, is er geen beproefd alternatief voorhanden. Deen vat de periode samen door te stellen dat er al met al én geen bruikbaar instrument voor selectie is ontwikkeld maar ook geen criterium voor het stellen van een cesuur. Voor Deen blijft selecteren toch vooral een subjectieve aangelegenheid (Nelissen, 2014: 24).

De studies van Stellwag en Deen kunnen worden gezien als een terugblik op het zorgvuldig onderzoeken van het probleem van de aansluiting. Feitelijk komen beiden tot de slotsom dat alleen uitstel van selectie een optie is omdat alle varianten gericht op het verbeteren van de selectie lijken uitgeput. Die stand van zaken van dat denken is ook terug te vinden in het Plan Rutten van 1950 -1951: 'Daar het in hoge mate onwaarschijnlijk geacht moet worden, dat op een leeftijd van 12 à 13 jaar reeds voor alle leerlingen, die een overwegend theoretische intelligentie bezitten, kan worden beoordeeld of zij voor voorbereidend wetenschappelijk onderwijs geschikt zijn, is het niet juist, dan reeds een zeer scherpe scheiding aan te brengen tussen het voortgezet

onderwijs voor algemene doeleinden en het voorbereidend wetenschappelijk onderwijs.' (Stellwag, 1955: 37). Stellwag wijst erop dat het selectievraagstuk vooral moet worden gezien als een pedagogisch probleem: 'Begaafdheid is niet van tevoren, maar pas achteraf, te constateren.' (Stellwag, 1955: 21).

### 4.3 Intelligentie en tests

Vanaf de oprichting van de hbs is er de nodige ervaring met toelatingsexamens. Toch blijft de uitval hoog. Dit maakt dat er verwachtingsvol wordt gekeken naar de mogelijkheden die zich aandienen op het terrein van de geschiktheidsmetingen in de vorm van intelligentiemetingen die met name op de arbeidsmarkt, in de wereld van overheid en bedrijfsleven, meer en meer ingang vinden. Ook internationaal is er veel aandacht voor de mogelijkheden die intelligentiemetingen lijken te kunnen bieden.

De Franse arts Binet wordt aan het begin van de vorige eeuw door het Franse ministerie van onderwijs gevraagd een test te ontwikkelen waarmee leerlingen geïdentificeerd kunnen worden voor wie het volgen van regulier onderwijs te hoog gegrepen zou zijn (Bisseret, 1979: 17). Binet zoekt naar mogelijkheden om dit vast te stellen, los van het onderwijs en hun sociale omgeving. Hij onderzoekt kinderen met behulp van een op een afgenomen testjes. De geobserveerde scores op een reeks opgaven van een kind vergelijkt hij met de gemiddelde score van kinderen van dezelfde leeftijd. Binet introduceert de begrippen 'mentale leeftijd' en 'chronologische leeftijd'. Door beide met elkaar te vergelijken wordt zichtbaar wie voorloopt en wie achterblijft. In de systematiek van Binet is 'intelligentie dus een relatief begrip' (Stellwag, 1955: 64). De testjes van Binet trekken de aandacht. Ze worden al snel vertaald in het Engels en verder ontwikkeld door de Amerikaan Terman, psycholoog aan de Stanford Universiteit. Hij gaat op basis van de ontwikkelde testjes van Binet verder. Dit wordt de Stanford-Binettest, die model komt te staan voor alle Amerikaanse IQ-tests die daarna verschijnen (Hanson, 1993: 209). Hij komt tot het begrip 'quotient': hij deelt de verstandelijke leeftijd door de chronologische en introduceert daarmee het intelligentie-quotient, ofwel het IQ (Stellwag, 1955: 65). De ontwikkeling van de meerkeuzevraag door Kelly in 1915 opent de weg voor intelligentiemetingen op grote schaal. Belangrijk zijn de Army Alpha en de Army Beta die in de periode tot aan het einde van de eerste Wereldoorlog door meer dan 1,75 miljoen rekruten worden gemaakt (Hanson, 1993: 211). Niet langer is het meten van intelligentie gericht op zwakbegaafden of wordt het geassocieerd met toepassingen in de context van onderwijs, maar het is een instrument voor het nemen van beslissingen over de 'aptitudes and achievements of normal people' (Hanson, 1993: 212).

Hoewel de toepassing en uitwerking van IQ-tests een vlucht neemt vanaf de Eerste Wereldoorlog, blijkt er nog altijd geen definitie van het begrip intelligentie zelf te kunnen worden gegeven. In 1923 haalt een psycholoog van Harvard, Edwin Boring, de Gordiaanse knoop uit elkaar door te stellen dat intelligentie datgene is dat door een intelligentietest wordt vastgesteld. Daarna volgen diverse pogingen om het begrip intelligentie alsnog nader te duiden. In zijn boek 'Intelligentie' uit 1980 geeft psycholoog Vroon een overzicht van tal van definitiepogingen voor het begrip 'intelligentie' en komt tot de slotsom dat 'we voorlopig moeten besluiten dat de definities van intelligentie op een niet al te hoog niveau liggen en weinig ophelderen' (Vroon, 1980: 43). 'Men volstaat met het aantonen ervan door de score op een instrument te correleren aan een extern criterium. Maar correlaties en oorzaken hoeven niets met elkaar te maken te hebben.' (Vroon,

1980: 45). De Britse psycholoog Spearman, die zich begin vorige eeuw bezig hield met factoranalyses, gaat er vanuit dat intelligentie een 'algemeen en erfelijk vermogen' is dat hij aanduidt met de grootheid  $g$ . Zijn idee is dat verschillende factoren bijdragen aan de algemene intelligentie en dat metingen gericht op de verschillende factoren tezamen een beeld geven van die algemene intelligentie (Vroon, 1980: 49).

Het gebrek aan een inhoudelijk duidelijke definitie heeft de aantrekkelijkheid van het concept en het willen kunnen meten ervan niet gehinderd. Het grote publiek heeft van intelligentie een beeld dat zich centreert rond het 'vermogen te leren'. De antropoloog Allan Hanson onderscheidt drie kenmerken die worden toegekend aan intelligentie:

- Het is één ding: dat zie je aan het feit dat het wordt uitgedrukt in één getal hoewel het wel uit verschillende onderdelen kan bestaan.
- Het varieert in hoeveelheid: je kunt er veel of weinig van hebben, wat je kunt aflezen van de numerieke schaal die gebruikt wordt om het in uit te drukken (alleen kwantitatieve data kunnen zo uitgedrukt worden).
- Wat je ervan hebt, staat vast voor de rest van je leven en daarom meet je met een intelligentiemeting niet wat je al weet maar wat je kunt leren (ability to learn). Wat je er precies mee doet hangt samen met je kansen en je motivatie maar iemands ability wordt als vaststaand gezien en ook nog eens in hoge mate als erfelijk (Hanson, 1993: 255).

Door gebruik te maken van meerkeuzevragen bij intelligentietests is grote, groepsgewijze afname mogelijk geworden. En is de verwerking ervan bij grote aantallen afnames veel eenvoudiger geworden. Dit is volgens Hanson de belangrijkste verklaring voor het succes van intelligentietests; dat het technisch mogelijk is geworden intelligentietests op grote schaal toe te passen (Hanson, 1993: 253). Want al in de jaren twintig van de vorige eeuw dringt volgens psycholoog Danziger (1990) het besef door dat de praktische toepassing van intelligentietests beperkingen kent. Hoewel de Verenigde Staten op grote schaal geëxperimenteerd hebben met intelligentietests voor de rekrutering van soldaten voor het leger, komt men daar tot de slotsom dat de uitkomsten maar beperkt bruikbaar zijn (Danziger, 1990: 156) en gaat men op zoek naar andere, aanvullende te meten kwaliteiten zoals loyaliteit, stabiliteit en kracht. Het individu wordt gezien als een verzameling van goed te onderscheiden, stabiele en algemene kwaliteiten die identiek zijn in hun aard voor het ene en het andere individu en waarin individuen alleen verschillen in de hoeveelheid die ze van een eigenschap hebben (Danziger, 1990: 157). Ook voor onderwijs geldt dat intelligentietests maar beperkt samenhangen met academische prestaties. Daarom zoekt men volgens Danziger aanvullingen vallend onder 'karakter' of 'persoonlijkheid' (Danziger, 1990: 158). Danziger stelt dat de beperkte successen van intelligentietests niet leiden tot het ter discussie stellen van de methodologie die het fundament vormt van die metingen en die de basis blijft van nieuw te ontwikkelen instrumenten, vanwege de economische belangen die ermee gemoeid zijn: 'The new instruments copied the features that had made group intelligence testing such an economical proposition.' (Danziger, 1990: 159). Bij intelligentietests wordt gewerkt met meerkeuzevragen die goed of fout beantwoord kunnen worden, en waarbij een groter aantal goed beantwoorde vragen een goede weerspiegeling zou geven van een grotere hoeveelheid van een bepaald vermogen (Danziger, 1990: 159). Deze manier van denken is vervolgens de basis geworden voor het 'meten' van persoonlijkheid: 'The psychological reality that was being constructed here invol-

ved the positioning of features that were independently definable and isolable and that retained their identity across situations and across persons.' (Danziger, 1990: 159). Door uitkomsten te koppelen aan een individu wordt de persoon het object en niet het subject van de reflectie. En daarmee worden vervolgens empirische data verzameld die volgens Danziger niet voortkomen uit de data maar daar juist aan vooraf gaan (Danziger, 1990: 161). Het opsporen van verschillen wordt daarmee van grotere waarde dan wat zich in een persoon afspeelt. Persoonlijkheidstests reduceren complexe input van individuen tot een serie scores op simpele schalen, wat feitelijk meer een reflectie is van de bedenker van de test en de opdracht die wordt voorgelegd dan dat de scores een goed beeld schetsen van een individu. Eventuele ambiguïteit kan worden voorkomen door taken wel of niet op te nemen in de set. Daarmee wordt volgens Danziger een subject niet gezien in zijn verscheidenheid maar vooral als subject geconstrueerd (Danziger, 1990: 2).

### 4.3.1 Intelligentietests en Nederland

In Nederland, en ook Duitsland, is men aanvankelijk afwachtend en afhoudend met het meten van intelligentie. De Duitser Stern heeft in de vooroorlogse periode veel invloed op het denken van de pedagogen en psychologen in Nederland. Ook hij is geboeid door het onderzoek dat in Europa en de Verenigde Staten plaatsvindt op het gebied van intelligentiemetingen. Het gebruik van kennistoetsen bij de toelating tot het middelbaar onderwijs heeft laten zien beperkingen te kennen omdat de scores van kandidaten weliswaar inzicht geven in wat de kandidaat aan kennis heeft verworven, maar dat blijkt niet veel verband te houden met toekomstige resultaten. En zorg over uitval blijft zich vertalen in druk op het kunnen voorspellen van schoolsucces. De zoektocht naar 'de begaafdheid buiten de verworven schoolse kennis om' is aantrekkelijk voor exploratie en onderzoek, aanvankelijk ook voor het Nutsseminarium. Want de zoektocht naar bepaling van geschiktheid blijft de gemoederen bezighouden. Eigen en ander Europees onderzoek maken echter dat Kohnstamm in 1936 tot het inzicht komt dat intelligentie-onderzoek gericht op een onveranderlijk, in aanleg aanwezig gegeven voor hem onhoudbaar is geworden (Mededeling 30, Nutsseminarium, in: Deen, 1969: 32). De Duitser Otto Selz en zijn Mannheimse onderzoekers laten Kohnstamm zien dat 'het afnemen van een intelligentietest en het uitrekenen van een IQ alleen maar iets zegt over de intelligentie in actu en niet over de intelligentie in potentia' (Deen, 1969: 34). Zij tonen met onderzoek aan dat leerprestaties in een schoolklas verbeteren wanneer de betere leerlingen hun aanpak overdragen op hun klasgenoten (van Strien, 2003). Kohnstamm schuift op in de denkrichting dat intelligentie niet voor eens en voor altijd vastligt, dat de intelligentie-prestatie voor opvoeding, voor beïnvloeding vatbaar is. Hij stelt dat het niet fair is opleidingskansen van een kind afhankelijk te maken van een instabiele meting (Van Strien, 2003). De hoofdaandacht bij intelligentietests verschuift voor Kohnstamm van prestatie naar proces (Deen, 1969: 33). Dit leidt tot een positieve houding ten opzichte van het stillestuk dat gezien wordt als een gecombineerde taal- en denkoefening. 'Kohnstamm en Van Veen hebben belangstelling voor het 'stillezen' omdat zij die vorm van meten zien als een manier om geestelijke vermogens te toetsen waarbij parate kennis er minder toe doet. Terwijl dit soort oefeningen in de Anglo-Amerikaanse testgeschiedenis al spoedig in een objectief scorebare multiple-choice vorm wordt gegoten, behouden ze in de Europese traditie bij de uitbouw tot stillestuk hun kwalitatief karakter: sommige antwoorden zijn evident goed of fout, maar andere zitten daar ergens tussenin. De procesgerichtheid maakt dat Kohnstamm vooral heil ziet in het stillestuk als bepaler van schoolgeschiktheid. Er zijn immers geen kant-en-klare antwoorden en vooral het gesprek over de vragen moet de afnemer een beeld geven van het denkvermogen van de kandi-



daat. Degene die de proef als toets gebruikt moet zelf de kwaliteit van de antwoorden beoordelen.' (Van Strien, 2003). Stellwag laat zich aarzelder uit over de bruikbaarheid van stillezen als bepaler van algemene geschiktheid. Ze wijst op de problemen die samenhangen met een goede constructie van dergelijke tests en de manier waarop de beoordelaar komt tot een oordeel 'goed' of 'fout'. Met name het laatste probleem wijst zij als oorzaak aan voor het 'afzakkingsproces' dat zij ziet in de praktijk, waarmee ze bedoelt dat beoordelaars soepeler worden naarmate ze meer beoordelen. Ze komt tot de slotsom dat om een beeld te krijgen van de intelligentie van kinderen 'men echter niet alleen zal moeten afgaan op de gegevens door middel van het stilleesstuk verkregen, maar daarnaast ook opgaven zal moeten voorleggen die overeenkomst vertonen met intelligentietests. Dus formele denkvragen. Maar deze zijn op zichzelf niet voldoende om zich een begrip te kunnen vormen van de genuanceerde capaciteit van een intelligent kind. Daarnaast zal men werk moeten geven om te zien of en op welke wijze het kind de leerstof der lagere school heeft verwerkt. Dus examenvragen.' (Stellwag, 1955: 239).

Hoewel het Nutsseminarium positief is over het stillezen als intelligentiemeting heeft het instituut nooit zelf onderzoek ernaar gedaan. Het kleinschalige onderzoek in de vorm van een dissertatie dat de pedagoog Langeveld (1934) heeft uitgevoerd en een later vervolgonderzoek brengen Langeveld tot de slotsom ten aanzien van tests in het kader van de aansluiting tussen lager en middelbaar onderwijs 'dat een selectie-systeem eigenlijk niet de aangewezen weg is' (in: Deen, 1969: 52). Deen zelf stelt: 'voor een conclusie dat het stillezen als zodanig een betere voorspeller van toekomstig schoolsucces zou zijn dan andere middelen, ontbreken de gegevens' (Deen, 1969: 52). Hij verklaart de opkomst van het stillezen vooral vanuit het beoogde terugslaggeffect op het onderwijs: 'Hier ligt, wetende welke invloed van examens op het onderwijs uitgaat, m.i. de voornaamste reden om het stillezen in iedere toelatingsprocedure op te nemen.' (Deen, 1969: 52).

Testpsycholoog Prak onderzoekt de mogelijkheden die intelligentie-metingen bieden voor de context van het onderwijs. Hij wijst erop dat moderne intelligentietests eigenlijk in wezen gestandaardiseerde proefwerken zijn die 'het verstand van de leerling zo goed mogelijk peilen'. Hij ziet zeker moeilijkheden bij het vaststellen van intelligentie door de mogelijke interferentie vanuit wat in het onderwijs geleerd is en milieu maar stelt dat de invloed daarvan beperkt is (Prak, 1952: 14). Hij pleit voor een vergroting van de efficiëntie in het onderwijs met behulp van tests. Hij maakt daarbij een analogie met de scheepvaart waar dankzij instrumenten als kompas, radiopeiling, echolood en radar men veel veiliger is gaan varen en pleit voor het gebruik van instrumenten om onderwijs te verbeteren en 'sociale misstanden en persoonlijke déraillementen' te voorkomen: 'Nu door het genie van pioniers als Galton, Binet, Terman, Burt, Lewis en anderen, de navigatiemiddelen zijn ontdekt en tot een zeer bevredigende graad van verfijning en doelmatigheid zijn opgevoerd, is het de hoogste tijd, dat elk schip er een dankbaar gebruik van maakt.' (Prak, 1952: 403).

Stellwag ziet geen heil in groepsgewijs afgenomen vormen van intelligentietests. Zij ziet mogelijkheden voor het bepalen van schoolgeschiktheid in de interpretatie en combinatie van gegevens op basis van tests en andere gegevens die het onderwijs zelf levert: 'We moeten ons dus beven in fenomenologische, structureel-constitutieve of analytische intelligentie-theorieën.' (Stellwag, 1955: 112). Zij benadrukt de beperking die IQ-tests hebben voor de context van on-

derwijs en wijst op de gevaarlijke gevolgen van IQ-tests in de aansluitingsproblematiek: 'Het zou echt paedagogisch en didactisch onjuist zijn wanneer de school, door de grote waarde die men aan het I.Q. toekent, zich in de richting van de test-intelligentie zou gaan specialiseren, dus een bloedarme formeel-abstracte vorming als hoofdzaak zou zien, ten koste van de inhoud van het onderwijs. We zouden dan tot de meest primitieve, vóór-wetenschappelijke opvatting van onderwijs zijn teruggekeerd.' (Stellwag, 1955: 110). Zij ziet wel de waarde van schoolvorderingstoetsen als informatiebron voor onderwijzers, maar zij acht die vormen van toetsing onbruikbaar als selectiemiddel voor het voortgezet onderwijs: 'Eenzijds omdat het onderwijs zich dan te veel zou richten op het onderwijzen van datgene wat uiteindelijk zou worden getoetst en anderzijds omdat de voorspellende waarde ervan voor het succes van de leerling in het voortgezet onderwijs gering is. Bovendien zou de persoonlijke verantwoordelijkheid van de onderwijzer worden uitgeschakeld.' (Haas, 1995: 188). Stellwag ziet meer heil in 'proefklassen'. Dit zijn klassen die gevormd worden door leerlingen die mogelijk kunnen doorstromen naar het middelbaar onderwijs en daarop vooruitlopend een bepaalde periode les krijgen op de manier zoals dat gaat in het middelbaar onderwijs in vakken zoals die voorkomen op het middelbaar onderwijs door docenten van het middelbaar onderwijs om zo op basis van reële observaties te kunnen bepalen wie van de leerlingen geschikt is voor het middelbaar onderwijs en wie niet.

In 1957 maakt psycholoog Kouwer de balans op ten aanzien van geschiktheidsbepalingen op basis van psychologisch onderzoek en bijbehorende instrumenten zoals die dan beschikbaar zijn (Kouwer, 1957). Hij stelt dat psychologisch onderzoek naar geschiktheid in bedrijven redelijk gangbaar is geworden maar dat dat niet geldt voor selectie voor onderwijsdoeleinden. Dit in tegenstelling tot de Verenigde Staten waar psychologische selectie voor scholen, colleges of universiteiten zeer algemeen verbreid is (Kouwer, 1957: 19). Het voorspellen van succes of falen in de toekomst is in de Verenigde Staten heel gebruikelijk geworden. In Nederland is meer aarzeling, ook bij Kouwer zelf: 'De toekomst is onberekenbaar, en menselijke gedragingen liggen vrijwel nooit vast.' (Kouwer, 1957: 25). Hij maakt onderscheid tussen drie functies van tests die in zijn woorden theoretisch 'streng' van elkaar gescheiden behoren te worden: de voorspellende, de beschrijvende en de zelf-ontdekkende functie (Kouwer, 1957: 33). In Nederland ligt in de psychologische praktijk de meeste nadruk op de beschrijvende functie die men wel gebruikt voor het doen van voorspellingen zodat de twee functies, voorspellen en beschrijven, juist niet strikt van elkaar worden gescheiden: 'Een veel voorkomende gewoonte is hier b.v. aan op zuiver subjectief-descriptieve wijze verkregen resultaten een voorspellende waarde te verlenen.' (Kouwer, 1957: 33). Kouwer is niet erg onder de indruk van intelligentietests. Hij vraagt zich af wat die feiten zijn die dergelijke tests zeggen te voorspellen. Een theoretische onderbouwing ontbreekt en hij stelt dat 'er gewacht zal moeten worden totdat op grondiger wijze dan tot nog toe geanalyseerd is wat wij in het dagelijks leven onder 'intelligentie' verstaan, voordat het mogelijk is tests te construeren die dit intuïtieve begrip beter benaderen dan tot nog toe het geval is' (Kouwer, 1957: 69). In de tijd van Kouwer wordt voor de ontwikkeling van intelligentietests gekozen voor een pragmatische benadering; de gemiddelde testscore van een groep die wat betreft klas en leeftijd overeenkomt, wordt als uitgangspunt genomen om nieuwe proefpersonen mee te vergelijken. Zo wordt gesteld dat 'als een score hoger ligt dan het gemiddelde van klasgenoten, dan wordt verwacht dat dit een goede leerling zal zijn; ligt zij lager, dan zal deze leerling ook minder gunstige resultaten in de klas behalen' (Kouwer, 1957: 64). Daarmee is intelligentie die gemeten wordt met een intelligentietest niet een bepaalde 'hoeveelheid' intelligentie maar gebonden aan

de test waarmee zij verkregen is (Kouwer, 1957: 65). Het is een relatief begrip waarbij de score in hoge mate afhankelijk is van de rest van de populatie en met een andere test zou een andere score en een ander oordeel kunnen worden gegeven. Nederland is volgens Kouwer nog niet ver gevorderd met het ontwikkelen van kwalitatief hoogwaardige, voorspellende tests: 'Inderdaad zijn adequaat geijkte voorspellende tests zeldzaam en in ons land nog in het geheel niet beschikbaar.' (Kouwer, 1957: 63).

Hoewel er sprake is van een groeiende, kritische houding ten opzichte van intelligentietests in relatie tot het voorspellen van toekomstige geschiktheid, wordt toch vooral bij psychologische instituten gezocht naar een testinstrumentarium waarmee toekomstige prestaties kunnen worden voorspeld. Zo promoveert Groen in 1967 op een grootscheeps, longitudinaal onderzoek naar 'De voorspelbaarheid van schoolcarrières in het voortgezet onderwijs'. Dit onderzoek is vooral gericht op het achterhalen van factoren die leiden tot succes of falen in schoolcarrières. Maar meer aandacht komt er vooral voor de rol van onderwijs en milieuomstandigheden voor scores op intelligentietests. Eind jaren zestig bereikt de kritiek op intelligentietests en het concept intelligentie een hoogtepunt. Het leunen op uitkomsten van intelligentietests alleen wordt niet langer acceptabel gevonden. Het leidt tot het verleggen van de focus van intelligentietests naar een breder aanbod van testinstrumenten om talenten in kaart te brengen, waar intelligentietests dan wel een onderdeel van kunnen zijn. In combinatie wordt geprobeerd een zo goed mogelijk beeld te krijgen van het kind waarbij het oordeel van een specialist in de vorm van een psycholoog nodig is voor de interpretaties van de verschillende (test)gegevens om te komen tot een oordeel (Haas, 1995: 186-187). Dit maakt dat IQ-tests niet meer prominent in beeld zijn maar ook nooit zijn verdwenen; ze worden onderdeel van grotere testbatterijen.

### 4.3.2 Kennen of kunnen als selectieprincipe

In de loop van de vorige eeuw is het gemeengoed geworden, zeker in de Verenigde Staten, om onderscheid te maken tussen 'intelligentietests' enerzijds en 'achievementtests' of 'schoolvoorderingentoetsen' anderzijds. Theoretisch is hier een scherp onderscheid tussen: het eerste type test stelt vast wat een leerling *kán* bereiken en het tweede type is gericht op het vaststellen van wat een leerlingen *héeft* bereikt. De in die tijd gangbare discussie over de vraag of voorbereiding op een intelligentietest mogelijk is, is volgens Stellwag 'niet een theoretisch wetenschappelijke vraag maar ook een ideologisch principe' en zij stelt de vraag retorisch: 'Is het inderdaad mogelijk van alle school- en levenservaring op een dusdanige wijze te abstraheren, dat men de mogelijkheden die in het kind zèlf gelegen zijn, onberoerd door uiterlijke omstandigheden, levenslot en levensloop, op ieder willekeurig tijdstip met objectieve exactheid uit kan pellen, en wel zo dat men zich ook aan enkele voorspellingen omtrent de toekomst kan wagen? Kortom, kan men het ene kind op grond van zijn prestaties in de verstandtest wèl, het andere niet de kans geven?' (Stellwag, 1955:137).

Stellwag wijst hier op een belangrijk probleem, namelijk of prestaties gebaseerd op verleden een goede voorspelling kunnen geven van de toekomst. Het testresultaat gaat namelijk vooraf aan het toekomstig gedrag. Eerst komt het vaststellen van potenties en dan ontstaat zicht op het daadwerkelijke presteren. Dit maakt volgens Hanson de kans klein dat iemand alsnog faalt en het maakt het vrijwel onmogelijk om te ervaren of iemand die op basis van een test is afgewezen het alsnog wel goed zou hebben gedaan (Hanson, 1993: 291). Voor het effect op het individu is

het dan ook niet van belang of er gebruik wordt gemaakt van een potentiëmeting of een prestatie-meting. Daar komt bij dat het onderscheid in de praktijk lastig te maken is. Ook Stellwag stelt dat er geen feitelijk onderscheid tussen intelligentietests en achievementtests mogelijk is. Ze stelt dat in beide groepen van tests 'gelijksortige procedures te vinden zijn voor de vaststelling van de definitieve redactie, gelijke of gelijksoortige methodes om het statistisch gemiddelde of de "norm" te vinden, en dezelfde differentiërende principes: oplopende moeilijkheid of snelheid. In beide worden vragen gesteld in dezelfde vorm: alternative choice, multiple choice, matching, ranging, completion etc. Nu zou ons dit nog niet tot identiteit doen besluiten, zo de tests, wat inhoud betreft, dan ook maar genoeg uiteenliepen. Het merkwaardige is, dat zij dit niet doen.' (Stellwag, 1955: 137). Zij ziet met name overeenkomsten tussen de kenmerken van beide testvormen die ook nog eens inhoudelijk uitwisselbaar zijn in de praktijk: 'Niet minder merkwaardig is dat sommige van de tests die oorspronkelijk als scholastic-achievement-test gebruikt zijn, gaandeweg in het genre intelligentietest zijn opgenomen en omgekeerd.' (Stellwag, 1955: 138). Ze haalt de Engelse toetsontwikkelaar Ballard aan die al in 1920 stelt 'dat het onmogelijk is een test te bedenken die geschiktheid (ability) onderzoekt en dan niet tevens ook kennis onderzoekt' (Stellwag, 1955: 138). Samenvattend komt Stellwag tot de conclusie dat 'de intelligentie- groepen schooltests nòch in vorm nòch in inhoud zodanig verschillen, dat ze als twee verschillende categorieën kunnen worden beschouwd. Hiermee is de vooropstelling, dat de intelligentietest natuurlijke intelligentie "meet" en ieder kind eenzelfde kans heeft, op zijn minst twijfelachtig geworden.' (Stellwag, 1955: 144). Daarnaast wijst Stellwag op het 'terugslageffect' van toetsen op het onderwijs en is daar beducht voor: 'We stellen inhoud niet gelijk met oppervlakkige kennis, maar we geloven dat het de taak van het onderwijs is, dat voedsel te geven waardoor de geest zich in omvang en genuanceerdheid kan ontwikkelen.' (Stellwag: 144). Zij gaat in op grootschalig onderzoek in Engeland dat ten tijde van haar studie plaatsvindt, over het effect van 'drillen' op het behalen van scores op intelligentietests. De eerste conclusies wijzen uit dat het gemiddelde IQ omhoog gaat na systematische oefening (Stellwag, 1955: 158). 'Wanneer men dit alles leest wordt men zeer weemoedig gestemd. Is dit nu het resultaat van ongeveer een halve eeuw rekenen en scoren?' (Stellwag, 1955: 159-160). Vervolgens haalt ze Kohnstamm aan die 'bijna 30 jaar geleden al als zijn mening uitgesproken heeft, dat de intelligentie het best beoordeeld kan worden op grond van opgaven waar zij ex officio ingeschakeld wordt en dat is o.a. het doodgewone schoolwerk' (Stellwag, 1955: 160). En met deze conclusie maakt zij duidelijk dat zij selectie op basis van het voorspellen van geschiktheid niet mogelijk acht.

#### 4.4 Selectie en groei van onderwijsdeelname

De periode van de wederopbouw en de democratiseringsgolf hebben in Nederland gevolgen voor het denken over de inrichting van de samenleving, de rol van beleid daarbij en ook de mogelijkheden die worden gezien voor een actieve bijdrage van onderwijs aan het veranderen van die samenleving. Waar de gebeurtenissen in de Tweede Wereldoorlog in veel landen aanleiding geven voor bezinning op een grotere verbondenheid tussen opvoeding en onderwijs en verbroedering van alle burgers met elkaar om de mensheid te behoeden voor herhaling van de grote menselijke tragedie, is er in Nederland vooral aandacht voor een zo snel en efficiënt mogelijke inzet van arbeidskrachten om Nederland weer op te bouwen en om als land mee te kunnen in de vaart der volkeren. De leerplicht is in 1950 uitgebreid tot 8 jaar wat betekent dat de 'babyboom' van na de Tweede Wereldoorlog in de jaren zestig druk uitoefent op het onderwijsstelsel want voor al die kinderen moet na de lagere school plek worden gecreëerd. En een passende plek zo-

dat efficiënte opleidingstrajecten mogelijk worden. Daarmee krijgen selectie en zorgen daarover in de naoorlogse periode in Nederland naast zorgen over de uitval nog een ander belang: het zodanig selecteren dat er sprake is van efficiënte en ook effectieve onderwijsstrajecten waarmee met zo beperkt mogelijke kosten de arbeidsmarkt zo goed mogelijk bediend kan worden.

#### 4.4.1 Zoeken naar talent

Verborg en met name de zoektocht daarnaar is na de Tweede Wereldoorlog urgent geworden in de ogen van velen omdat de overtuiging heerst dat alle talent zo goed mogelijk moet worden benut om mee te kunnen in de vaart der volkeren. Het gevoel dat de westerse wereld achterblijft bij de Sovjet-Unie speelt daarbij een grote rol. Als Nederland een rol wil spelen op het wereldtoneel dan is actie geboden. Voor onderwijs wordt een grote rol daarbij gezien, ondersteund door de 'human capital-theorie' die stelt dat onderwijs even belangrijk is voor economische groei als investeringen in industrie of machines. Belangrijk uitgangspunt van de theorie is dat elk extra jaar scholing als een waardevolle investering moet worden gezien in individuele toekomstige productiviteit en daarmee als een investering die zich terugbetaalt op termijn. Deze theorie neemt in de jaren zestig aan belang toe en zet de discussie over selectie en uitval weer op scherp. Socioloog Van Kemenade vat het denken over onderwijs en samenleving in die tijd als volgt samen: 'Onderwijs werd herontdekt als de motor van sociale en economische vooruitgang en de human capital-theorie legde de wetenschappelijke basis voor de opvatting dat verhoging van het scholingsniveau als een essentiële bron voor economische groei moest worden beschouwd.' (Van Kemenade, 1981: 130).

#### OECD

Het sturen op economische groei in combinatie met onderwijs wordt gestimuleerd door de in 1948 opgerichte 'Organisation for Economic Co-operation and Development (OECD) waar Nederland direct vanaf de start lid van is. De OECD heeft onderwijs en onderwijsbeleid vanaf het begin geplaatst in het licht van het streven naar economische groei en ontwikkeling. In artikel 8 van de Conventie is afgesproken dat 'the Contracting Parties will make the fullest and most effective use of their available manpower' (Papadopoulos, 1994: 22). De Verenigde Staten zijn ongerust over de 'technological gap' tussen hen en de Russen. In 1958 komt de OECD met een programma gericht op het verbeteren van de kwaliteit van onderwijs en de groei van onderwijsdeelname gericht op het verkleinen van die 'gap'. Deelnemende landen worden gevraagd hiervoor plannen op te stellen en zich daarbij te baseren op data die voor de langere termijn de behoefte aan wetenschappelijk en technisch arbeidspotentieel inschatten. Hiermee doet onderwijsplanning zijn intrede, wat zich zal consolideren in de jaren daarna. Tevens wordt van deelnemende landen verwacht dat zij in hun land meer aandacht vragen voor de belangen voor de ontwikkeling van wetenschap en techniek. Deelnemende landen worden geacht hun mogelijkheden voor wetenschappelijk en hoger technisch onderwijs uit te breiden en de toegankelijkheid van dat onderwijs te verbeteren zodat alle talent de kans krijgt deel te nemen aan onderwijs gericht op wetenschap en techniek (Papadopoulos, 1994: 25). Met de uitvoering van 'surveys' probeert de OECD de ontwikkeling in de diverse landen te monitoren. Lastig daarbij is dat landen hun eigen vormen van rapportage kennen gebaseerd op eigen data: 'The major problem was the lack of satisfactory international standards of education and manpower classifications and statistics.' (Papadopoulos, 1994: 27). De OECD werkt hard om te komen tot standaarden op basis waarvan vergelijken beter mogelijk wordt en competitie, ook binnen de OECD wordt aangemoedigd:

'Thenceforth, in the minds of many Europeans, attention shifted to catching up with their North American Partners in the new competitive economic environment, stimulated by the liberalisation of trade. And this [...] became part of the hidden, and at times, explicit, agenda of the education and science policies of the more advanced among the European OECD countries.' (Papadopoulos, 1994: 28). De invloed van de OECD gaat tot op het niveau van het curriculum. Met name voor wis- en natuurkunde streeft de OECD naar modernisering van curricula volgens Amerikaans model. Onderzoek van de OECD in 1959 naar de curricula voor wiskunde van deelnemende landen heeft namelijk laten zien dat de bestaande curricula niet adequaat zijn om moderne wiskunde te kunnen onderwijzen. Dit zelfde patroon herhaalt zich voor natuurkunde, scheikunde en biologie (Papadopoulos, 1994: 31). De aandacht die sociologen vanuit diverse landen vragen voor sociale barrières die onderwijsdeelname in de weg staan, wordt door de OECD opgepakt als een extra uitdaging die landen hebben om alle beschikbare talent op te sporen om de kloof te kunnen dichten.

### **Rapport Van Heek**

In 1968 verschijnt het rapport met de bij de tijdgeest passende titel 'Verborgten Talent'. Het bevat de uitkomsten van een grootschalig onderzoek van de Universiteit Leiden onder leiding van de socioloog Van Heek naar 'welke maatschappelijke krachten genoemd verschijnsel beïnvloeden' (Van Heek, 1968: V). Het verschijnsel waarop gedeut wordt is het feit dat 'een jongen, afkomstig uit het milieu der vrije beroepen en hogere employés, gemiddeld een minstens vijftienmaal grotere kans maakt tot het v.h.m.o. te worden toegelaten dan een leeftijdsgenoot uit het milieu der niet-agrarische handarbeiders' (Van Heek, 1968: V). Met zijn grootschalige onderzoek wil Van Heek de vraag beantwoorden of er in Nederland daadwerkelijk sprake is van verborgen talent en hoe groot dit potentieel dan is. Want in die tijd is Nederland zich ervan bewust dat het bestaande schoolstelsel wellicht kinderen niet selecteert op basis van achtergrondkenmerken die wellicht wél de kwaliteiten hebben in het licht van wetenschapsontwikkeling en internationale concurrentie. Van Heek vindt zeker wel verschillen in deelname aan onderwijs voor verschillende milieus, maar die kunnen volgens hem worden verklaard uit verschillen in geschiktheid. Van Heek concludeert dat er niet of nauwelijks talent verloren gaat. Bij een nadere analyse van dezelfde data komt onderwijssocioloog Van Kemenade tot een andere conclusie. Als hij de leerlingen met elkaar vergelijkt die eenzelfde prestatiescore hebben komt hij tot de conclusie dat 'bij leerlingen in de zesde klas van het basisonderwijs, die vrijwel gelijke schoolprestaties leveren, milieufactoren de doorstroming naar het V.H.M.O. sterk beïnvloeden ten nadele van kinderen uit de 'lagere' sociale milieus' (Van Kemenade & Kropman, 1972: 227). Volgens Van Kemenade is er dus bij nadere beschouwing wel degelijk sprake van verborgen talent.

De human capital-theorie zorgt ervoor dat onderwijs wordt gezien als een relevante en op termijn renderende bron van investering. Aan onderwijs wordt economisch nut en noodzaak toegedicht om Nederland via onderwijs mee te kunnen laten spelen op het hoogste internationale niveau. Zoals Idenburg het formuleert: 'Ons land is betrokken in de grote competitie der volken, willen we ons daarin handhaven dan mag geen begaafdheid verloren gaan' (De Vries, 1993: 35). Tot ver in de jaren zestig wordt 'de massale uitbreiding van het vo algemeen toegejuicht in verband met de stijgende economische vraag naar gekwalificeerde arbeidskrachten' (Matthijssen, 1982: 12). Concreet betekent dit dat de opvatting breed gedragen wordt dat meisjes en kinderen met talenten afkomstig uit de 'lagere standen', kansen moeten krijgen in het onderwijs om te

worden toegelaten tot vormen van onderwijs die voorheen waren voorbehouden aan de 'hogere' standen, mits zij daarvoor geschikt zijn. Nederland streeft daarbij naar een betere benutting van het beschikbare potentieel. En zo ligt er bij het onderwijs niet alleen de druk om voldoende, capabele arbeidskrachten te leveren om mee te kunnen op internationaal niveau, maar wordt onderwijs ook de plek waar die selectie en allocatie moet plaatsvinden.

#### 4.4.2 Onderwijs als kans

Het opsporen van verborgen talent en de juiste ontwikkeling daarvan zodat de arbeidsmarkt mee kan in de wereldwijde economische groei die de naoorlogse periode kenmerkt staat centraal in het onderwijsbeleid. De waarde die in die tijd wordt toegekend aan het economisch en maatschappelijk belang van onderwijs maakt dat kritisch gekeken wordt naar tegenstanders die andere belangen voor onderwijs voor ogen hebben zoals sociologen en pedagogen. Bijvoorbeeld Idenburg waarschuwt voor te veel aandacht en ruimte voor de individuele ontplooiing en talentontwikkeling gezien vanuit het belang van het individu zoals hij dat in de discussies over het doel van onderwijs waarneemt, met name in de discussie over kansengelijkheid. Hij karakteriseert die aandacht voor een individualistische invulling van onderwijs als een nieuwe vorm van meer traditioneel denken in termen van standen dat 100 jaar later weer terugkeert. In zijn optiek verhoudt deze manier van denken zich lastig tot een samenleving en een economie die moeten worden opgebouwd met de daarbij best passende mensen: 'Wij ontdekken dat het traditionele ideaal van gelijkheid van kansen toch wel duidelijk individualistische trekken draagt. Het is niet voor niets aan de gedachtewereld der bourgeoisie ontsproten. Wij zullen het recht van de enkeling niet langer mogen beschouwen zonder daarnevens de behoefte der gemeenschap in 't geding te brengen, welke uitgaat naar de best-mogelijke bezetting van en rekrutering voor de meest verantwoordelijke posten in de diverse beroepsgroepen.' (Idenburg, 1960: 182). Idenburg wijst nadrukkelijk op het belang van het zien van een balans tussen belangen omdat aandacht voor het individuele recht op ontplooiing groeit. Dit komt dit keer niet uit de bourgeois-hoek zoals in de eeuw daarvoor, maar vindt nu vooral aanhangers ter linkerzijde van het politieke spectrum en meer aan de onderkant van de samenleving, waarbij de gedachte is dat onderwijs de kansen op stijgingen op de maatschappelijke ladder mogelijk kan maken. Onderwijs wordt gezien als weg naar een betere positie in de samenleving: 'Er heeft zich een jacht naar maatschappelijke opstijging merkbaar gemaakt van dat deel van ons volk dat gegrepen is door de bourgeois-leus: 'Kennis is macht'. (Idenburg, 1960: 509). Idenburg vat de kritiek op de gevolgen van het nastreven van het ideaal van de kansengelijkheid, en dan met name van de individuele benadering van onderwijssucces, als volgt samen: 'Daarnevens verlangt het ideaal van de gelijkheid van kansen in zijn gangbare strekking nadrukkelijk om aanvulling. Het begrip vertoont een noodlottige eenzijdigheid, zolang het alleen maar gezien wordt vanuit het recht van de enkeling. Deze visie veronderstelt namelijk een maatschappij, welke niet meer is dan het operatieterrein voor individuele opklimmingsdrift en begeerte naar invloed en macht.' (Idenburg, 1960: 513). De individualistische insteek van onderwijs staat in de optiek van critici een meer maatschappelijke visie op onderwijs als vorm van een samenleving die je als gemeenschap bouwt, in de weg: 'Wanneer wij van daaruit kunnen geraken tot het besef, dat wij allen leden zijn van een grote gemeenschap, waarin wij ieder naar onze gaven en naar de behoeften der maatschappij onze taak hebben te vervullen, ontstaat een andere samenleving dan die, welke het individualistisch begrip van de "gelijkheid van kansen" ons voor ogen stelt.' (Idenburg, 1960: 513). Een kritische houding ten opzichte van de ontwikkelingen in de samenleving wat betreft onderwijs vindt niet

veel weerklank in het gangbare publieke debat: de politieke partijen ter linkerkant van het politieke spectrum omarmen het 'bourgeois'-denken van de individuele ontplooiing als belangrijke voorwaarde voor het door hen geambieerde 'opwaarts streven'.

#### 4.4.3 Zorg over onderwijsdeelname

De bundeling van krachten, ter linkerkant ingegeven vanuit een wens tot stijgen op de maatschappelijke ladder via onderwijs en ter rechterkant ingegeven vanuit de human capital- gedachte, leidt tot een breed draagvlak om onderwijsdeelname te vergroten en beter toegankelijk te maken, wat ook zichtbaar wordt in de vergrote mogelijkheden om door te stromen naar vormen van hoger onderwijs. Dit laatste wordt volgens De Rooy 'in progressieve kringen voorgesteld als een belangrijk mechanisme in de verdere emancipatie van de arbeidersklasse' (De Rooy, 2018: 176). Hiermee neemt de druk op de onderwijsvoorzieningen toe en 'openbaart zich dit streven in een stormloop op de instellingen van onderwijs. Het schoolwezen is de maatschappelijke lift bij uitnemendheid en deze lift kan de congestie nauwelijks meer verwerken.' (Idenburg, 1960: 159). Hoe logisch en voorspelbaar de groei ook is gegeven het beleid, hij zorgt voor bezorgdheid over het in goede banen kunnen leiden van de aantallen en over de financiële en economische kant ervan. Want hoe vinden al die leerlingen hun plek? En hoe kunnen de kosten van het onderwijs daarbij in de hand worden gehouden? De Prioriteitennota van het Ministerie van Onderwijs en Wetenschappen uit 1968 ademt die zorg. De nota heeft zijn naam ontleend aan de prioriteiten die moeten worden toegekend aan alle uitdagingen waar het onderwijs voor staat en met name de financiële gevolgen. Want naast de groei van de instroom is er nog altijd het probleem van de uitval: 'In het v.h.m.o. behaalde slechts 47 procent van de leerlingen het diploma en van de geslaagden deed de helft er dan nog langer over dan de cursusduur.' (De Rooy, 2018: 181). Dit leidt met name bij politici tot zorgen omdat 'Nederland de internationale concurrentie slechts aankon door veel beter gebruik te maken van het intelligentiepotentieel in de bevolking' (De Rooy, 2018: 176). Zij achten een betere selectie en daarmee verdeling over de schoolsoorten nodig om voor voldoende hoger opgeleiden te kunnen zorgen, enerzijds door de leerlingen te weren voor wie bepaalde vormen van onderwijs te hoog gegrepen zijn en anderzijds door leerlingen toe te laten die voorheen niet doorstroomden naar hogere vormen van middelbaar onderwijs. Afkomst en status zouden zo aan belang inleveren ten gunste van individuele verdiensten en capaciteiten.

In de Prioriteitennota van 1968 maakt het Ministerie van Onderwijs en Wetenschappen onderscheid tussen drie doelstellingen voor het onderwijsbeleid: a) het voldoen aan de sociale vraag naar onderwijs, b) het voorzien in de behoeften van de maatschappij en c) het scheppen van een evenwicht tussen de sociale vraag en de behoeften van de maatschappij (Ministerie van Onderwijs en Wetenschappen, 1968: 2). Uit de uitwerking van wat moet worden verstaan onder de 'sociale' vraag wordt duidelijk dat het hierbij vooral gaat over de regulering van nieuwe fenomenen: een groeiend aantal leerlingen, langere deelname aan het onderwijs en een inhoudelijke en methodische aanpassing van het onderwijs gezien de verandering van deelnemers. De overheid houdt zich nog altijd vooral bezig met het in goede banen leiden van het onderwijs. Het beleid van het opheffen van belemmeringen, het distributieve onderwijsbeleid, gaat echter naar de maatstaven van de jaren zestig niet ver genoeg. Er wordt meer actief handelen van de overheid verwacht om kansen te bieden aan allen voor het volgen van onderwijs. Minister van Onderwijs en Wetenschappen, Veringa, schetst in zijn nota 'Democratisering van het onderwijs' in 1969 dat een grote toename van leerlingen in het onderwijs past bij een verdere democratisering van het



onderwijs. Hierin komt een overheid naar voren die onderwijs en samenleving als samenhangend geheel ziet. Veringa stelt daarin dat 'onderwijs in volstrekt isolement van de samenleving geen zinvol bestaan kan leiden'. Volgens Veringa is duidelijk een koerswijziging in het denken over de rol van onderwijs in de samenleving waarneembaar: 'Het verschil in deelneming per milieu aan de verschillende vormen van voortgezet onderwijs werd aanvankelijk niet opgemerkt, of althans niet als een probleem gezien. Men aanvaardde het als een vanzelfsprekende zaak.' (Ministerie van Onderwijs en Wetenschappen, 1969: 8). Voor Veringa is duidelijk dat onderwijs 'in beginsel een dienstverlenende functie in de maatschappij moet hebben' (Ministerie van Onderwijs en Wetenschappen, 1969: 7), daaraan toevoegend dat 'over de aard van deze dienstverlening het oordeel aan verandering onderhevig is' (Ministerie van Onderwijs en Wetenschappen, 1969: 7). Hij stelt dat 'aanvankelijk de opvatting overheerste dat onderwijs een passief conserverende rol moet spelen'. Onderwijs 'heeft in deze opvatting als voornaamste taak kritisch evaluerend te behoeden wat voor de samenleving van waarde is en dit aan de jongere generaties over te dragen' (Ministerie van Onderwijs en Wetenschappen, 1969: 7). De nota laat zien dat er beweging is in de conserverende rol die men voor onderwijs voor ogen heeft en dat men onderwijs veel meer begint te zien als de plaats waar veranderingen beginnen: 'Vooral de laatste tijd wint echter de gedachte veld dat aan het onderwijs een veel actievere rol dient toe te vallen. Het moet voorop gaan in de vernieuwing van de samenleving, en daarvan als het ware een dynamisch centrum zijn' (Ministerie van Onderwijs en Wetenschappen, 1969: 7). De nota van Veringa ademt in die zin een andere sfeer uit dan de nota's van de decennia daarvoor, duidelijk beïnvloed door de democratiseringsgolf van de jaren zestig: 'Onderwijs als instrument van vorming van de gehele bevolking. Het is langs die weg een sleutelpositie gaan innemen in het proces van de toewijzing van sociale en economische posities en het is van fundamentele betekenis geworden voor de ontwikkeling van de samenleving als geheel.' (Ministerie van Onderwijs en Wetenschappen, 1969: 7). Veringa uit zijn zorg vooral op het punt van vraag en aanbod: 'De vraag rijst, in hoeverre zij ook in overeenstemming zijn met de toekomstige vraag naar de verschillende soorten gekwalificeerde menselijke arbeid. Het gaat hier om de confrontatie van de vraag naar onderwijs, de 'social demand' en de 'economic demand'. (Ministerie van Onderwijs en Wetenschappen, 1969: 30). Alle aandacht voor maatschappelijke ontwikkelingen ten spijt blijft het beleid vooral gericht op het opheffen van belemmeringen, zodat enerzijds voldaan kan worden aan de behoefte van de arbeidsmarkt, namelijk een groei van beter opgeleid personeel, en anderzijds manifesteert zich een overheid die meebeweegt op de democratiseringsgolf en die vanuit sociaal perspectief wijst op het belang van eerlijke (onderwijs)kansen voor allen. De nota spreekt ook zorgen uit over de haalbaarheid en de betaalbaarheid van een grotere toegankelijkheid: 'Hogere vormen van voortgezet onderwijs hebben een langere cursusduur en vragen hogeregraadsdocenten. Een stijgende deelneming aan deze onderwijsvormen zou alleen om die redenen reeds hogere uitgaven vergen.' (Ministerie van Onderwijs en Wetenschappen, 1969: 29). Veringa ziet dit als een noodzakelijk kwaad waar, de human capital-theorie indachtig, genoeg baten tegenover zullen komen te staan: 'Het moet mogelijk zijn een deel daarvan op te vangen door verbetering van het rendement, dat wil zeggen door een snellere doorstroming en vermindering van het aantal uitvallers.' (Ministerie van Onderwijs en Wetenschappen, 1969: 30).

Van Veen, de opvolger van Veringa, kan niet om de gevolgen van de groei van de deelname aan het onderwijs heen, een groei niet alleen ingegeven door economische behoeften maar vooral ook door individuele ontplooiingswensen. Hij slaat met zijn nota een andere weg in. Dit is nodig,

zo stelt hij, omdat er 'een ontwikkeling heeft plaatsgevonden van middelbaar onderwijs voor een minderheid naar voortgezet onderwijs voor allen' (Ministerie van Onderwijs en Wetenschappen, 1972: 3). Hij schetst de ontwikkeling van de veranderingen in doelstellingen voor onderwijs van 'bijdrage aan de economische en maatschappelijke ontwikkeling' naar 'een belangrijk middel voor sociale stijging en inkomensverbetering' (Ministerie van Onderwijs en Wetenschappen, 1972: 3). Van Veen nuanceert de doelstelling van gelijke ontplooiingskansen voor iedereen. Het doel van zijn beleid is 'de jonge mens via het onderwijs, dat wordt gegeven op de onder hun verantwoordelijkheid vallende scholen, de mogelijkheden te bieden, die plaats in de samenleving te vinden, die past bij zijn mogelijkheden, deze noch te boven gaat, noch onbenut laat, een plaats, die hem de kans geeft de harmonie in zijn leven te vinden en te bewaren' (Ministerie van Onderwijs en Wetenschappen, 1972: 3). Hij richt zich op ontplooiing binnen economisch aanvaardbare kaders: 'Onderwijs voor allen betekent niet het absolute primaat van de wensen van het individu; het vereist een behoedzame afweging van de belangen en rechten van het individu tegenover die van de gemeenschap.' (Ministerie van Onderwijs en Wetenschappen, 1972: 3). Daarbij spelen kosten een belangrijke rol. De leerling kan volgens de nota niet zomaar 'recht doen gelden op het volgen van elke door hem gewenste opleiding, ongeacht de kosten die hieraan voor de gemeenschap zijn verbonden en ongeacht de eventuele beperktheid van de behoefte van de gemeenschap aan afgestudeerden van zulk een opleiding' (Ministerie van Onderwijs en Wetenschappen, 1972: 3). Van Veen stelt voor om meer planmatig te kijken naar het onderwijsstelsel: 'Wij hebben reeds opgemerkt, dat het overwegingen van doelmatigheid zijn die in belangrijke mate ten grondslag liggen aan onze overtuiging dat wij geleidelijk aan dienen te geraken tot een stelsel van onderwijsplanning.' (Ministerie van Onderwijs en Wetenschappen, 1972: 8). Deze visie op het doel van onderwijs verhoudt zich lastig tot de democratiseringsgedachte die juist de ruimte voor het individu, ieder individu, in het onderwijs wil creëren. Dit vereist een lossere relatie tussen onderwijs en arbeidsmarkt.

## 4.5 Stelselveranderingen

De veranderende toestroom en ambities van en voor het onderwijs zoals die zich aftekent in de jaren vijftig plaatst onderwijs in een ander daglicht. Dit leidt tot een kritisch beschouwen van het bestaande onderwijsstelsel en het nadenken over een nieuw, beter bij de tijd passend stelsel. De Rooy onderscheidt drie groepen die streven naar een wezenlijke verandering van het onderwijs. De eerste groep bestaat volgens hem uit beleidsmakers die het onderwijs (na de wet van Thorbecke) zien als een 'onoverzichtelijke lappendeken' en wijzen op het belang van een update. De tweede groep bestaat uit pedagogen en onderwijskundigen die zoeken naar een betere aansluiting tussen lager en middelbaar onderwijs, enerzijds om de overgang vloeiender te laten verlopen en anderzijds om het keuzemoment uit te stellen. De derde groep bestaat volgens De Rooy uit politici die zorgen hebben over de aansluiting van Nederland op internationale concurrentie en die een betere verdeling willen van leerlingen over de schoolsoorten zodat uitval en zittenblijven teruggedrongen kunnen worden en kinderen uit de lagere milieus meer kansen krijgen om hogere vormen van onderwijs te kunnen volgen (De Rooy, 2018: 175).

### 4.5.1 Mammoetwet

Uiteindelijk lukt het minister Cals om in 1963 een dan honderd jaar oude onderwijswet te vervangen door een nieuwe. Vanwege de lange en zware weg naar de totstandkoming ervan, wordt de nieuwe onderwijswet 'Mammoetwet' genoemd in de wandelgangen. Met het nieuwe onder-

wijsstelsel vindt een belangrijke verandering plaats ten aanzien van selectie. De gedachte is dat de invoering van een 'brugklas' de mogelijkheid biedt de selectie niet te laten plaatsvinden aan het einde van de lagere school maar leerlingen de mogelijkheid te bieden om het eerste jaar van het voortgezet onderwijs te gebruiken om te laten zien wat hun capaciteiten zijn. Hiermee komt de selectie onder de verantwoordelijkheid van het voortgezet onderwijs. Deen is in 1969 hoopvol gestemd over de aanpak van het probleem van de selectie en stelt: 'De eigenlijke determinatie, en in samenhang daarmee de eigenlijke selectie, voltrekt zich in het brugjaar van de ontvangende school, die daarmee het karakter krijgt van een eenjarige proefklasse. Principieel is echter, dat het hier geen selecterende, maar een determinerende proefklasse betreft: er volgen geen afwijzingen, maar verschuivingen naar een beter passende plaats.' (Deen, 1969: 82). Deen wijst erop dat die selectie een tijdelijk karakter heeft: 'Definitieve beslissingen behoeven niet te worden geveld, ze zijn zelfs strijdig met de brugjaargedachte. Slechts een "grove zifting" is – naar het woord van Stellwag – vereist; de redelijke waarschijnlijkheid, dat de leerling in een van de varianten van mavo, havo en vwo een plaats kan vinden, moet worden vastgesteld.' (Deen, 1969: 81). Idenburg is minder positief. In 1960 stelt hij dat 'het standenonderwijs zoals destijds door Thorbecke bedacht plaats maakt voor een stelsel dat nog steeds gericht is op differentiatie waarbij de belangrijkste verandering wellicht is dat er sprake zal zijn van een brugklas die voor vwo, havo en mavo gelijk zal zijn zodat verandering van school zonder tijdsverlies mogelijk is' (Idenburg, 1960: 364). De vorming van scholengemeenschappen moet overstappen binnen één school makkelijker mogelijk maken. Die vorming van scholengemeenschappen blijkt in de praktijk weerbarstig en scholengemeenschappen met brede brugklassen komen maar beperkt van de grond. In 1969 is ongeveer 15% van het scholenbestand een scholengemeenschap (Schüssler, 2006:19).

Uitstel van selectie blijkt beperkt van waarde, want de keuze voor beroepsonderwijs of algemeen vormend onderwijs moet nog altijd aan het einde van het lager onderwijs plaatsvinden. Omdat de keuze voor avo of beroepsonderwijs blijft bij het lager onderwijs, is het vraagstuk van selectie met de komst van het nieuwe onderwijsstelsel niet van de agenda verdwenen. Een vorm van selectie aan het eind van het lager onderwijs is nodig om leerlingen toe te delen, richting beroepsonderwijs aan de ene kant en algemeen vormend onderwijs aan de andere kant. Daar komt bij dat aangezien brede scholengemeenschappen maar moeizaam van de grond komen en brugklassen in hoge mate categoriaal worden ingevuld, er binnen de meeste scholen voor voortgezet onderwijs de wens is onderscheid te maken voor plaatsing ofwel in een brugklas voor mavo/havo ofwel in een voor havo/vwo. Zo blijft de selectie en advisering door het lager onderwijs van belang. Daarnaast blijven ondanks het nieuwe stelsel de categorale gymnasia bestaan die op de oude voet verder kunnen en hun eigen vormen van selectie hanteren, al dan niet op basis van het advies van de lagere school. Die sterke gelaagdheid van het stelsel laat de sporen zien van de standenmaatschappij die met de komst van de Mammoetwet weliswaar op het oog de nodige aanpassingen laat zien maar in de praktijk niet zo ingrijpend is als de naam suggereert. Zoals De Rooy stelt: 'Zo was nu wel in één wet het algemeen voortgezet en het beroepsonderwijs geregeld, maar in feite was dit onderscheid uit de negentiende eeuw gehandhaafd: een leerling kon niet uit het beroepsonderwijs naar het algemeen vormend onderwijs overstappen' en 'ondanks alle veranderingen werd het klassieke onderscheid tussen 'hoog' en 'laag' onderwijs behouden, zij het enigszins versluierd.' (De Rooy, 2018: 180-183).

Selectie aan het eind van het basisonderwijs is dus met het nieuwe onderwijsstelsel niet verdwenen. Integendeel, met de nieuwe wet is niet alleen selectie aan het eind van het basisonderwijs voor havo/vwo (voorheen v.h.m.o.) nog altijd aan de orde maar is de selectie een plaatsingsvraagstuk geworden voor alle kinderen aan het eind van het basisonderwijs: ieder kind krijgt een naar inzicht van de lagere school best passende plek toegewezen voor vervolgonderwijs. Voor leerlingen die doorgaan naar het algemeen vormend onderwijs geldt daarbij formeel dat in de brugklas een definitieve keuze wordt gemaakt. Schlüssler stelt dat met de komst van de Mammoetwet er in de praktijk niet veel verandert aan het vraagstuk van selectie. Volgens Schüssler komt dat omdat 'een schools en selectief jaarklassensysteem zich niet eenvoudig laat wegpoetsen. Dus wordt de ingewikkelde toelatingsprocedure al vanaf het begin het alibi voor vier verschillende soorten brugklas.' (Schüssler, 2006: 19). De met veel moeite gerealiseerde wijziging van het onderwijsstelsel brengt geen rust: 'De strijd tegen het overeind gebleven standonderwijs gaat onverminderd verder.' (Schlüssler, 2006: 19). De nieuwe wet is weliswaar een knap staaltje politiek, maar blijkt geen extreme veranderingen in petto te hebben. De Rooy bestempelt de stelselherziening als 'een ingewikkeld compromis: enerzijds waren verschillende soorten opleidingen nu wel 'ineengeschakeld', maar tegelijk was de categoriale basis, de afzonderlijke schoolsoorten, behouden.' (De Rooy, 2018: 180). Daarmee is er nog altijd sprake van een stelsel waar op jonge leeftijd, namelijk aan het einde van het primair onderwijs, belangrijke keuzes worden gemaakt en van een stelsel dat blijft gekenmerkt door duidelijk te onderscheiden drempels in de overstap van het ene onderwijs naar het andere, waar met het nieuwe stelsel geen oplossingen voor worden geboden: 'De aansluiting tussen 'lagere' en 'hogere' opleidingen blijft een probleem en zal de verticale doorstroming blijven belemmeren.' (De Rooy, 2018: 180).

#### 4.5.2 De Middenschool

De 'Mammoetwet' wordt niet gezien als oplossing in de strijd tegen schooluitval en zittenblijven en wordt bij invoering als 'achterhaald' beschouwd nog voordat de wet volledig zijn beslag heeft gekregen. Parallel aan de invoering ontstaat steeds meer druk te komen tot een onderwijsstelsel met een lange, gemeenschappelijke periode, een comprehensieve vorm van onderwijs zoals in veel andere Europese landen ontstaat. Van diverse kanten wordt gewezen op de onaantrekkelijkheid van het bestaande onderwijs en wordt aangedrongen op een andere kijk op onderwijs. De tijd rijpt voor onderwijs 'dat gericht is op bredere en persoonlijke ontplooiing' (Schlüssler, 2006: 20). Van links tot rechts ziet de politiek de Mammoetwet als een eerste, maar niet voldoende stap om het onderwijsbestel te vormen naar een bij de nieuwe tijd en samenleving passende vorm van onderwijs. Het kabinet-Biesheuvel (1971-1973) gaat met het idee voor een middenschool aan de slag: 'Minister van Veen (CHU) en staatssecretaris Schelfhout (KVP) laten er geen misverstand over bestaan: de middenschool is de oplossing voor de kwalen van de Mammoetwet. Van Veen ... kondigt experimenten aan en geeft een vergezicht op de wijziging van het onderwijsstelsel. Doel is de integratie van schoolvormen waardoor milieufactoren een minder grote rol gaan spelen bij de school- en beroepskeuze.' (Schlüssler, 2006: 23). Is de Mammoetwet vooral nog een ingreep in de structuur van het onderwijs, de plannen voor de Middenschool passen bij een overheid die verder gaat en zich mengt 'in de leerlingselectie op school en in de inhoudelijke doelstellingen van het onderwijs' (Mellink, 2014: 155). In eerste instantie lijkt de Middenschool een oplossing voor twee zaken: de selectie kan worden uitgesteld omdat er langer een gemeenschappelijke onderwijsperiode is en door die latere selectie kan het onderwijs beter aansluiten op de talenten en capaciteiten die kinderen daadwerkelijk hebben laten zien, wat de

uitval zal verminderen. In de Contourennota van 1975 worden de plannen voor een zogenaamde 'middenschool' door minister Van Kemenade nader uit de doeken gedaan en gepositioneerd in het onderwijsdebat van die tijd. De beoogde middenschool 'kent geen categoriale opbouw en geen scheiding tussen algemeen vormend onderwijs en beroepsonderwijs' en beslaat de periode van 12- 16 jaar, kent geen leerstof-jaarklassen en leerlingen worden niet naar 'cognitieve intelligentieniveaus ingedeeld' (Ministerie van Onderwijs en Wetenschappen, 1975: 25). Het beter aansluiten op de persoonlijke ontplooiing ligt dicht bij het aansluiten op persoonlijke kwaliteiten en talenten. Daarmee krijgt de middenschool twee ambities die op gespannen voet met elkaar kunnen komen te staan, maar die in eerste instantie ook zorgen voor een breed politiek draagvlak: enerzijds is de middenschool bedoeld om uitstel van selectie te kunnen realiseren waarmee kansengelijkheid wordt gediend en anderzijds wordt aan de middenschool een doelmatiger selectiebeleid toebedeeld waardoor onderwijs beter aansluit bij de interesse en talenten van leerlingen. 'Eerlijke onderwijskansen en ruimte voor individuele ontplooiingsmogelijkheden gingen daardoor hand in hand.' (Mellink, 2014: 140).

#### 4.6 Selectie en tegenstrijdige beleidsambities

De Vijlder wijst in zijn onderzoek naar de relatie tussen natiestaat en onderwijs op twee elementen die in de planbare samenleving van na de periode van de wederopbouw van bijzonder belang zijn: het Keynesiaanse denken over economische sturing door de overheid en het geloof in vormen van social engineering voor de aanpak van vraagstukken als sociale ongelijkheid. De overheid ziet zichzelf, in lijn met het denken over economische sturing, als een belangrijke factor in het sturen van wat er in onderwijs moet gebeuren en ziet onderwijs als een van de manieren waarop gestuurd kan worden om de samenleving anders in te richten. De manpowerbenadering betekent voor het onderwijs meer concreet dat 'een onderwijssysteem een output moet kunnen genereren, die bijna perfect aansluit op de behoeften van het arbeidsbestel' (De Vijlder, 1996: 112/113). Maar daarmee zijn de 'oude' waarden waar onderwijs zich traditioneel op heeft gericht, niet gelijk afgeschreven. Er is eerder sprake van dat die in de nieuwe waarden worden opgenomen. Zoals De Vries het stelt: 'De motieven van natievorming, groepsbinding en disciplineren zijn echter niet verdwenen: ze zijn bij het economische motief ondergedoken.' (De Vries, 1993: 37). Net als ook nieuwe waarden, vanuit de democratiseringsgolf gericht op het vergroten van individuele mogelijkheden om als individu de plek in onderwijs te vinden die het beste past, worden ondergebracht in een streven naar onderwijs met een maximale output in termen van aansluiting op de arbeidsmarkt. Onderwijs heeft de selecterende functie opgepakt, wat ervoor heeft gezorgd dat onderwijs 'een essentieel instrument van maatschappelijke verdeling is geworden' (Van Kemenade, 1981: 120). Onderwijs heeft daarbij niet kunnen zorgen voor een andere maatschappelijke verdeling maar voor 'een bestending van de bestaande ongelijkheid' (Van Kemenade, 1981: 122).

##### 4.6.1 Uitstel van selectie

Uitgangspunt van de Contourennota van 1975 is dat onderwijs moet worden gezien als de plaats waar nieuwe generaties worden klaargestoomd voor deelname aan de maatschappij waar straks 'een eigen bijdrage, betrokkenheid en medezeggenschap worden verwacht' (Ministerie O&W, 1975: 12). Daarmee moet onderwijs volgens de Contourennota gericht zijn op 'een volwaardige persoonlijke ontplooiing en een goede maatschappelijke voorbereiding' (Ministerie O&W, 1975: 12). En met dat laatste wordt dan vooral maatschappelijke weerbaarheid of sociale bewustwor-

ding bedoeld. Leerlingen moeten mondig en maatschappelijk weerbaar worden gemaakt. Het verlengen van de schoolperiode beschouwt de Contourennota als een 'traditionele onderwijsstrategie' die niet langer past. Het werkend bestaan zal zo veranderen dat een voorbereiding daarop in het onderwijs niet erg zinvol meer is. Niet een verdere verlenging van de initiële onderwijsperiode is nodig, maar juist een bekorting, aangevuld met 'levenslang verder leren' (Ministerie O&W, 1975: 14). En tegelijkertijd moet die initiële fase aantrekkelijker worden om uitval en zittenblijven tegen te gaan: 'Het onderwijsleerproces zal zich naar de mogelijkheden van de leerlingen moeten richten, zowel naar niveau als naar tempo.' (Ministerie O&W, 1975: 49). Daar hoort bij de 'noodzaak het onderwijs meer te doen aansluiten bij de individuele verschillen in belangstelling, begaafdheid, sociale ervaring en ontwikkelingsfase' (Ministerie O&W, 1975: 15). Hierbij past volgens de nota het leerstof-jaarklassensysteem niet goed meer: 'In zo'n opzet kan met individuele verschillen weinig rekening worden gehouden.' (Ministerie O&W, 1975: 15). Het gaat er volgens de nota niet langer om de leerlingen met uniforme schoolstandaarden te vergelijken maar 'vast te stellen waartoe de leerlingen in staat zijn en hen naar een verwezenlijking van hun mogelijkheden te begeleiden' (Ministerie O&W, 1975: 49). Van Kemenade wil daarbij nadrukkelijk streven naar wat hij noemt 'interne differentiatie', waarmee hij bedoelt dat leerlingen les krijgen op het niveau en in het tempo dat bij hen past, maar hij ziet tegelijkertijd ook het gevaar van 'streaming', het opdelen in homogene groepen, dat daarbij op de loer ligt: 'In hoeverre zich in de loop van de middenschool 'setting' en 'streaming' zullen gaan voordoen, blijft de vraag. In ieder geval moet gepoogd worden deze verschijnselen zo lang mogelijk uit te stellen.' (Ministerie O&W, 1975: 53). Op dit punt hoopt Van Kemenade dat experimenten nadere inzichten geven.

Tegelijkertijd wijst de nota ook op de grenzen die aan de individualisering moeten worden gesteld vanuit maatschappelijk en financieel belang. Dat moet onder andere worden geborgd met verantwoorde leerplanontwikkeling gekoppeld aan toetsing, al dan niet gericht op selectie. Want hoewel de nota pleit voor diagnostische toetsing tijdens het onderwijs en uitstel van selectie tot na het funderend onderwijs, stelt de nota tegelijkertijd dat 'ook tijdens de programma's selectie zal worden toegepast, waarbij studenten worden afgewezen of verwezen naar andere programma's. Aan het eind van programma's zullen eindtoetsingen plaatsvinden, vergelijkbaar met de huidige eindexamens, die recht geven op diploma's.' (Ministerie O&W, 1975: 17).

#### 4.6.2 Onderwijs en arbeidsmarkt

De economische inslag van het onderwijsbeleid wordt door de overheid in de jaren '50 en '60 vooral geoperationaliseerd door financiële belemmeringen om onderwijs te volgen op te heffen. Door het weghalen van met name financiële drempels wordt onderwijs voor iedereen even toegankelijk. Het beleid gericht op het opheffen van barrières leidt in de praktijk niet tot een verschuiving in de deelname van diverse sociale groepen aan de verschillende onderwijssoorten. Er ontstaat druk op een actievare rol van de overheid. In 1970 publiceert Idenburg een artikel in *Paedagogische Studiën* waarin hij uiteenzet wat er volgens hem komt kijken bij een overheid die stuurt op de inrichting van de samenleving met behulp van onderwijs. Hij werkt uit wat er nodig is om te komen tot een 'constructieve onderwijspolitiek'. Het eerste wat nodig is, is helderheid over de beoogde doelen van onderwijs in de samenleving: 'Het uitgangspunt van de constructieve onderwijspolitiek ligt in de doeleinden, die men met het onderwijssysteem wil bereiken.' (Idenburg, 1970: 5). En ten aanzien daarvan stelt hij vast dat 'het gebrek aan visie als een dikke mist over onze onderwijswereld ligt. Alle vaart dreigt erin vast te lopen' (Idenburg, 1970: 18).

Dit baart zorgen vanuit sociaal maatschappelijk perspectief maar ook vanuit sociaaleconomisch perspectief, 'omdat onderwijs faalt als rekruteringsveld voor al het beschikbare talent, en daardoor niet optimaal functioneert als toeleveringsbedrijf van geschoolde manpower, die voor een expanderende economie en voor de ontwikkeling en instandhouding van het welvaartsniveau in steeds grotere getale vereist is' (Van Kemenade, 1981: 128). Het verhogen van het scholingsniveau wordt gezien als een belangrijke indicator van sociale en economische vooruitgang. De nota 'Over het onderwijsbeleid' uit 1972 van de volgende minister van onderwijs, Van Veen, laat contouren zien van een meer sturende overheid met betrekking tot de rol van onderwijs voor de inrichting van de samenleving. Onderwijs vereist volgens de nota 'een behoedzame afweging van belangen en rechten van het individu tegenover die van de gemeenschap' (Ministerie O&W, 1972:3). Schüssler spreekt over deze nota als een 'geweldig keerpunt' omdat volgens hem voor het eerst duidelijk blijkt dat de minister zich daadwerkelijk intensief met het onderwijs gaat bemoeien (Schüssler, 2006: 23). Er is politiek breed draagvlak ontstaan en vertrouwen gegroeid in de overheid als aanjager van onderwijsvernieuwing. Het 'distributieve beleid' heeft langzaam plaatsgemaakt voor beleid dat getypeerd kan worden als 'constructief onderwijsbeleid'; een overheid die met beleid wil sturen en bijdragen aan verandering van maatschappelijke structuren. Een overheid die 'de maatschappelijke verandering dient' (Meijnen et al., 2000: 32). De discussienota uit 1975 'Contouren van een toekomstig onderwijsbestel' van minister Van Kemenade ademt nog meer de sfeer van een overheid die met onderwijs wil sturen. Het is een nota die meer ruimte voor andere functies van onderwijs wil maken: 'Onderwijs is primair een welzijnsvoorziening, gericht op persoonlijke ontplooiing van mensen en op maatschappelijke voorbereiding van zijn leerlingen in veel ruimere zin dan beroepsoefening alleen.' (Ministerie O&W, 1975: 20). In 1973 stelt de Wiardi Beckman Stichting, het wetenschappelijk instituut van de Partij van de Arbeid, dat 'het onderwijs is gaan functioneren als antwoord op de maatschappelijke behoefte aan geschoolde manpower voor de industriële samenleving' (Wansink, 1992: 80). Manpowerplanning is dan een gangbaar begrip in Europa waarmee wordt gepoogd vraag en aanbod op de arbeidsmarkt in lijn te brengen met elkaar. De Wiardi Beckman Stichting pleit er in 1973 juist voor om 'het onderwijs uit de beperkende binding met de arbeidsmarkt los te maken' (Wansink, 1992: 81). Deze kijk op onderwijs sluit meer aan op de democratiseringsgeest van de jaren zestig. De ontkoppeling van onderwijs en arbeidsmarkt wordt aangevoerd als argument om het onderwijs aantrekkelijker te maken waarmee uitval en zittenblijven kan worden teruggedrongen. Maar die ontkoppeling van onderwijs en arbeidsmarkt is in de praktijk lastig. Ten eerste is daar de zorg over de financiële gevolgen van de toestroom naar onderwijs waarop Van Kemenade een antwoord moet bieden. In alle vormen van onderwijs neemt het aantal deelnemers namelijk substantieel toe: in 20 jaar zijn er 1,3 miljoen leerlingen bijgekomen in het onderwijs (Idenburg, 1970: 4). Dit heeft gevolgen voor de totale uitgaven van de overheid: 'Het aandeel van het departement van Onderwijs in de rijksbegroting was in vijftien jaar gegroeid van 14 tot 21 procent.' (De Rooy, 2018: 216). Hoewel de Contourennota pleit voor een uitstel van selectie tot na het funderend onderwijs, is ook in deze nota duidelijk dat er wel degelijk voor de hoogste vormen van onderwijs selectie nodig is. De kosten voor hoger onderwijs zijn dermate hoog dat doorstroming naar hogere vormen van onderwijs niet voor iedereen kan zijn weggelegd omdat dat onbetaalbaar zou worden. Daarnaast tekent zich in die tijd een teveel aan hoger opgeleiden af. Een dergelijk overschot acht minister Van Kemenade van onderwijs niet gewenst: 'Een sterk groeiende discrepantie tussen onderwijs en arbeidsmarkt frustreert persoonlijke verwachtingen, verzwakt de weerbaarheid van de afgestudeerden op de arbeidsmarkt en kan leiden

tot een onrechtvaardige verdeling van schaarse middelen en onderwijskansen.' (Ministerie O&W, 1975: 20). Het maakt dat Van Kemenade ervoor kiest te investeren in het funderend onderwijs ten koste van het hoger onderwijs; hij pleit voor 'regulering' van de instroom in hogere vormen van vervolgonderwijs. In zijn optiek is er selectie nodig om dit te reguleren. Selectie aan het eind van het funderend onderwijs wordt als een vorm van regulering gezien. Want er is volgens de nota sprake van schaarste: 'Onderwijsvoorzieningen zijn beperkt en mankracht en middelen zijn schaars. En bovendien kan de maatschappelijke behoefte aanleiding geven om het aantal plaatsen binnen een opleiding te beperken.' (Ministerie O&W, 1975: 17). Deze meer economisch en maatschappelijke kijk op onderwijs past bij de manier van denken waarmee Nederland zich ook internationaal weet te positioneren onder andere door het werk van de Nederlandse econoom Tinbergen. Tinbergen introduceert in de jaren zeventig het begrip 'race' tussen technologie en onderwijs. De vraag-georiënteerde onderwijsplanning en de human capital georiënteerde aanbodtheorie weet hij te verenigen in modellen waarmee prijzen van onderwijstrajecten kunnen worden bepaald (Heckman, 2018: 4).

De economisch, instrumentele benadering blijft dominant en er wordt weinig ruimte gecreëerd voor een andere, meer vormingsgerichte benadering van onderwijs. Dit wordt versterkt doordat uit onderzoek blijkt dat de verwachtingen vanuit de human capital-theorie dat uitgaven van onderwijs wel zullen worden terugverdiend in de vorm van welvaartsstijging moeten worden getemperd omdat het 'rendement' daalt als het aanbod van opgeleiden groter is dan de opnamecapaciteit en wanneer bovendien de directe 'produktiviteitswaarde' van dat aanbod geringer wordt (Ministerie O&W, 1975: 22). Dit is onder andere gebaseerd op een nieuw fenomeen dat zorgen baart waaraan Van Kemenade niet voorbij kan gaan: 'Het tekort aan hooggeschoolde arbeidskrachten uit de jaren vijftig en zestig dreigt om te slaan in een overschot.' (Wansink, 1992: 83). Het aantal hoger opgeleiden groeit begin jaren zeventig harder dan de arbeidsmarkt kan bedienen: 'Wij bevinden ons in de paradoxale situatie dat de aanvankelijk door de arbeidsmarkt, casu quo economische productie, opgeroepen en gestimuleerde vraag naar meer en hogere scholing zichzelf voorbij schijnt te streven en ... tot 'overproductie' gaat leiden.' (Ministerie O&W, 1975: 19). Het noopt tot keuzes. Van Kemenade kiest voor 'het beperken van de uitgaven voor het hoger onderwijs ten gunste van het onderwijs aan de leeftijdscategorie tot negentien jaar' (De Rooy, 2018: 217). Het automatisme waarmee leerlingen na het voortgezet onderwijs doormoeten naar hoger vervolgonderwijs 'moest worden afgeremd, studenten moesten begrijpen dat hun 'zeldzaamheidswaarde' sterk was afgenomen' (De Rooy, 2018: 217). Het verschijnsel van 'economische overkwalificatie' maakt dat volgens Matthijssen het overheidsbeleid tegenstrijdig wordt: 'Het blijft enerzijds gericht op het stimuleren van voortgezet onderwijs, terwijl anderzijds de toegang tot het hoger onderwijs kunstmatig wordt afgeremd.' (Matthijssen, 1982: 12). 'De met succes gestimuleerde onderwijsvraag wordt nu aan de top van het onderwijssysteem ingedamd op grond van het economisch motief van overproductie van hooggeschoolden.' (Matthijssen, 1982: 148). Matthijssen spreekt in die context over onderwijs als een opleidingspiramide die werkt als een verdeelmechanisme (Matthijssen, 1982: 125). 'Deze opleidingspiramide voorziet in een verdeling van de jeugd over een groot aantal diplomacategorieën, die het niveau van inschaling in de beroepshierarchieën van het arbeidsproces in hoge mate bepalen.' (Matthijssen, 1982: 125).



### 4.6.3 Onderwijs en onderwijsachterstanden

Hoewel het constructieve onderwijsbeleid op steun kan rekenen, verschijnt er steeds meer onderzoek, wereldwijd, dat laat zien dat het veranderen van de samenleving door te sturen op onderwijs taaier is dan op het eerste gezicht lijkt. Vanuit de wetenschap groeit in de jaren '70 de overtuiging dat ongelijkheid niet via het onderwijs op te lossen is en wordt duidelijk welke beperkte mogelijkheden onderwijs heeft ten aanzien van het bevorderen van gelijke kansen. De rol die scholen daarin kunnen vervullen, blijkt beperkter dan men aanvankelijk dacht. Volgens historicus Bram Mellink 'bleek uit sociaalwetenschappelijk onderzoek dat scholen weinig middelen in handen hadden om de ongelijkheid in de samenleving terug te dringen' (Mellink, 2014: 153). Veel sociologisch onderzoek wijst uit dat achterstanden zich al in belangrijke mate hebben gevormd voordat kinderen in het onderwijs instromen en dat het voor onderwijs lastig is hierin een grote verandering te voltrekken: 'Volgens sociologen ontstonden onderwijsachterstanden namelijk niet zozeer door slecht onderwijs, maar werden zij in de eerste plaats beïnvloed door de sociale milieus waarin kinderen opgroeiden.' (Mellink, 2014: 153). In de nota van Veringa uit 1969 wordt nog gesproken van het belang dat moet worden gehecht aan het veranderen van de school zelf: 'Het onderwijssysteem was tot dusverre niet bereid of bij machte zichzelf fundamenteel te veranderen. Het helpt de kinderen uit milieus die weinig tot de ontwikkeling van de schoolgeschiktheid bijdragen onvoldoende bij hun aanpassing aan zijn eisen en zijn klimaat. Het levert derhalve een te geringe bijdrage aan de opheffing of verzachting van wat Matthijssen noemt de discontinuïteit tussen de opvoedingsmilieus gezin en school.' (Veringa, 1969: 18). Belangrijk in dit verband is het boek 'Schooling in capitalist America' van de Amerikaanse sociologen Bowles en Gintis uit 1974, waarin zij beweren dat onderwijs sociale ongelijkheid niet tegengaat maar juist reproduceert. Zij stellen dat verschillen in het aantal jaren dat scholing is gevolgd al jaren afneemt maar dat dit niet geldt voor de ongelijkheid in inkomens. De econoom Hirsch werkt in zijn boek 'Social Limits to growth' (1977) het onderscheid uit tussen materiele en positionele goederen. Positionele goederen zijn lastig op waarde te schatten want hun waarde is relatief en afhankelijk van hoeveel anderen ervan hebben. Positionele goederen krijgen hun waarde door schaarste. Zou het beoogde goed wijder verspreid zijn, dan verliest het zijn waarde als bepaler van een hoge status. Dit is wat in onderwijs gebeurt. Ook onderwijs is te beschouwen als een positioneel goed (Ultee, 1980). Het gaat niet om de hoeveelheid onderwijs die iemand heeft gevolgd maar om hoeveel meer of minder onderwijs iemand heeft gevolgd in vergelijking met de anderen met wie hij de concurrentie aan moet gaan op de arbeidsmarkt. Hoe meer mensen met een bepaald diploma op een bepaald niveau de arbeidsmarkt betreden, hoe minder de waarde ervan zal zijn. Dit mobiliseert krachten om hier weerstand aan te bieden. De onderwijsongelijkheid blijkt terugkijkend volgens Dronkers niet te zijn afgenomen als gekeken wordt naar de relatieve afstand tussen het onderwijsniveau van de hogere en lagere strata: die is onveranderd groot gebleven (Dronkers, 2007:65). Hij verklaart dit met de 'maximal maintained inequality-these' die inhoudt dat een grotere deelname door de lagere strata aan een bepaald onderwijsniveau of -type alleen wordt 'toegestaan' als de deelname door de hogere strata aan dat onderwijstype of -niveau al maximaal is en een hoger en nog exclusiever onderwijsniveau of -type beschikbaar is (Dronkers, 2007: 66). Diploma's vormen een snelle, goedkope manier om zicht te krijgen op kenmerken van nieuw personeel en geven een indicatie van investeringen die nog nodig zijn om iemand productief in te kunnen zetten. Het is een handige voorsortering die door werkgevers verder verfijnd kan worden. Zodra hogere onderwijsresultaten zichtbaar leiden tot hogere arbeidsmarktposities, ontstaat druk bij de lagere statusgroepen om door onderwijs hogerop te komen. Op het moment

dat dit gebeurt, is er sprake van diploma-inflatie wat leidt tot het volgen van nog meer onderwijs en het vergaren van meer diploma's om de status te behouden. Het gemiddelde opleidingsniveau stijgt maar de productiviteit neemt niet toe. Meer onderwijsdeelname leidt niet zomaar tot meer productiviteit en economische groei maar veeleer tot diploma-inflatie, overscholing en verdringing op de arbeidsmarkt (Eidhof et al., 2016: 138).

Van Kemenade pleit ervoor een rem te zetten op de gedachte dat onderwijs te beschouwen is 'als een ladder tot meer prestige en inkomen' (Van Kemenade, 1975: 21). Hij spreekt over 'gelijke kansen' als 'de reële mogelijkheid van alle leerlingen tot ontplooiing' (Ministerie O&W, 1975: 47). Hij nuanceert het streven naar gelijke kansen door te stellen dat dit streven 'een groot maatschappelijk gevaar in zich bergt. Het gevaar van een toenemende sociale ongelijkheid tussen mensen met een verschillende begaafdheid op opleidingsniveau. Het loopt het risico een nieuwe en wellicht nog scherpere sociale tegenstelling op te roepen.' (Ministerie O&W, 1975: 12). Dit besef leidt er bij Van Kemenade toe dat hij erop aandringt dat onderwijsbeleid zich niet moet 'beperken tot het scheppen van gelijke kansen op deelname aan onderwijsvoorzieningen' (Ministerie O&W, 1975: 12). Van Kemenade stelt dat onderwijsbeleid 'zich daarenboven ook moet richten op het verbeteren van de ontplooiingsmogelijkheden voor verschillende begaafdheden en vooral op het verbeteren van de mondigheid en de maatschappelijke weerbaarheid van allen, ongeacht hun begaafdheid (1975: 12). Met name dit punt blijkt volgens De Rooy, achteraf, 'de achilleshiel' voor de discussie over de middenschool (De Rooy, 2018: 218). Het debat belandt daarmee volgens De Rooy midden in de politieke arena met de eerder beschreven gevolgen van dien.

Van Kemenade zoekt naar een balans tussen willen en kunnen: 'Onderwijs kan weliswaar niet op eigen kracht nieuwe maatschappelijke verhoudingen en meer gelijkwaardigheid in de samenleving bewerkstelligen, maar kan er wel een belangrijke bijdrage toe leveren.' (Ministerie O&W, 1975: 8). Matthijssen stelt dat de standpunten met betrekking tot de sociale gelijkheid van onderwijskansen vanaf die tijd zo ongeveer variëren tussen 'het móet wel, maar het kán niet' en 'het kán niet, maar het moet tóch' (Matthijssen, 1982:155). Daar komt bij dat duidelijk wordt dat de groeiende expansie van het onderwijssysteem een nieuwe ontwikkeling oproept die deze expansie op haar beurt verder zal stimuleren. Namelijk het feit dat kinderen tenminste het opleidingsniveau van hun ouders zullen nastreven of liever nog, zullen willen overtreffen. Kinderen zullen hun relatieve sociaal culturele voorsprong niet willen opgeven en deze willen behouden waardoor 'de drang tot het volgen van langer en beter onderwijs steeds sterker wordt' (Van Hoof & Dronkers, 1980: 91). Dit sluit aan bij het door Boudon in 1973 geschetste dilemma. Namelijk dat de sociaal-culturele en materiële rentabiliteit van het volgen van meer onderwijs voor het individu afneemt naarmate een steeds groter deel van zijn leeftijdsgenoten hetzelfde doet (Van Hoof & Dronkers, 1980: 91). Het leidt tot een verdere differentiatie van het onderwijsaanbod en een beperking van de toegankelijkheid.

Van Dyck concludeert in een studie in opdracht van de Onderwijsraad over de toegankelijkheid van het Nederlandse onderwijs dat de toewijzing van leerlingen aan de verschillende schoolsoorten is bedoeld om iedere leerling een passende plek in het onderwijs en daarmee in de samenleving te bieden, maar zeker niet om alle leerlingen alle ruimte te bieden hun eigen weg te zoeken in het onderwijs (Van Dyck, 1997:21). In zijn Commentaar op de nota 'Contouren van een

toekomstig onderwijsbestel' geeft de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) hieraan uitdrukking door te stellen dat het 'onderwijs niet het vermogen heeft veel te doen aan sociale ongelijkheid. De school vormt de samenleving niet, maar volgt' (De Rooy, 2018: 220). Van Kemenade stelt in 1981 terugkijkend vast dat 'de verwachting van velen dat dit ook tot een verandering in de maatschappelijke verdeling, tot een herverdeling van mogelijkheden over de hele bevolking en tot een meer gelijkwaardige samenleving zou leiden, echter slechts in zeer beperkte mate gerealiseerd is' (Van Kemenade, 1981: 120). De beperkte mogelijkheden voor onderwijs om een wezenlijke verandering in de samenleving teweeg te brengen, valt goed bij hen die zich zorgen maken over de kosten die meer en langere scholing, al dan niet aangevuld met speciale programma's voor steeds meer leerlingen, met zich brengen. Deze bezorgdheid is niet nieuw maar de onderzoeksuitkomsten geven wel nieuwe munitie.

Met de komst van nieuwe doelgroepen in het onderwijs, komen achterstand en achterstandsbeleid in een nieuw daglicht te staan. Aanvankelijk is er apart beleid voor de inpassing van andersstalige leerlingen maar in 1985 worden de beleidslijnen voor autochtone achterstandsleerlingen en het tot dan toe gevoerde beleid ten aanzien van allochtone leerlingen samengebracht. Het gezamenlijke beleidsprogramma gaat verder onder de naam 'OVV' (onderwijsvoorrrangsbeleid). In die periode ligt de uitvoering van tal van projecten op lokaal niveau. Met het OVV komt er een formele, stabiele basis onder de diverse initiatieven en ontstaat ook behoefte om de effecten van de diverse initiatieven te monitoren. Voor het meten van resultaten komt een landelijk longitudinaal onderzoek van de grond, Landelijke Evaluatie Onderwijsvoorrrang (LEO) dat later overgaat in de PRIMA cohort onderzoeken, bedoeld als informatiebron voor het volgen van ontwikkelingen, zowel wat betreft de prestaties van de leerlingen als wat betreft de inspanningen van de scholen. Het inzetten van compensatieprogramma's en andere initiatieven blijkt in de praktijk niet tot indrukwekkende resultaten te leiden. Wansink refereert aan het Rotterdamse project Onderwijs en Sociaal Milieu als 'het duurste en grootste compensatieprogramma van Nederland' waaruit in 1986 zelfs een negatief effect is gebleken op de kansengelijkheid (Wansink, 1992:75).

In de praktijk is sprake geweest van beleid dat met de combinatie van het streven naar kansengelijkheid en selectie als contradictoir kan worden gekarakteriseerd. Leune spreekt van een 'tegenstrijdige beleidsambitie' als hij het heeft over 'het gelijktijdige streven naar het versterken van de selectiefunctie van het onderwijs en het willen vergroten van de onderwijskansen van leerlingen uit de maatschappelijke achterhoede' (Leune, 2007: 14). Onderwijs is in Nederland vooral een vraagstuk van verdelen geworden. De koek moet verdeeld worden onder meer mensen: meer mensen hebben minder kans op een plaats in hoger of universitair onderwijs waarmee de aantrekkelijkheid van dit deel van de koek is vergroot (zie verder hoofdstuk 7).

#### 4.7 Selectie en efficiency

Het aanvankelijke vertrouwen in de rol die onderwijs, naast het selecteren van talent, kan nemen in het ontwikkelen van talenten kalft in de jaren zeventig en tachtig verder af. De teleurstellende resultaten van het achterstandenbeleid in combinatie met druk op de overheidsfinanciën en andere politieke winden zorgen voor een veranderende opvatting over de taak van de overheid en de rol van de school. Een overgang naar een meer terugtrekkende overheid ontrolt zich: 'Het ideaal van een 'constructief onderwijsbeleid' kwam in de knoop met zichzelf.' (De Rooy, 2018: 224). Het 'geloof' in de creatie van kansengelijkheid door een sturende overheid in combinatie met een

afnemend vertrouwen in de rol die onderwijs kan vervullen in het vormen van een meer egalitaire samenleving, leidt tot meer nadruk op de rol van het individu. De pedagoog Van Gelder wees er al begin jaren zeventig op dat 'opvoeding tot zelfontplooiing tot scherpe concurrentie tussen leerlingen kon leiden' (Mellink, 2014: 177). De sterke nadruk van de middenschool op individuele zelfontplooiing heeft bestaande ongelijkheid volgens Mellink 'slechts in nieuwe vormen gegoten, dus bestendig' (Mellink, 2014: 177). Het geloof in aangestuurde, structurele emancipatie van leerlingen via het onderwijs kalft af en raakt binnen enkele jaren omstreken (Mellink, 2014: 178). Met name de VVD verzet zich tegen 'staatsgestuurde onderwijshervormingen' daarbij gesteund door protestante en katholieke besturenorganisaties en het Nederlands Genootschap van Leraars (Mellink, 2014: 178). Begin jaren tachtig staan links en rechts volgens Mellink samen in hun afkeer van overheidsinmenging in processen van emancipatie (Mellink, 2014: 178). Het verdwijnen van het brede vertrouwen in politiek gestuurde onderwijshervormingen leidde ertoe dat 'het geloof in het vrije, zelfstandige individu juist werd versterkt' (Mellink, 2014: 151). Het is meer en meer aan het individu om zijn kansen te pakken bij de verdeling van de maatschappelijke posities, te beginnen bij de onderwijsdiploma's met een overheid en een school die hier voor zichzelf minder en minder een egaliserende rol weggelegd zien. Individualisering wil vooral zeggen dat het individu wordt teruggeworpen op zichzelf en zich niet ondersteund weet door een overheid die corrigeert op eerder ontstane ongelijkheid of een school die zich verantwoordelijk voelt voor het realiseren van gelijke uitkomsten. En uiteindelijk ook niet door een herziening van het onderwijsstelsel, want de middenschool komt er bij nader inzien toch niet. Mellink wijst erop dat met deze ontwikkeling emancipatie een persoonlijke verantwoordelijkheid is geworden die daarmee 'op een vreemde manier individueel 'meetbaar' wordt' (Mellink, 2014: 179). 'Emancipatie werd een maatstaf voor volwaardige deelname aan de maatschappij en een opdracht aan het individu. Want mensen konden als persoon wel of niet geëmancipeerd zijn en wie zijn onderwijskansen aangreep, behoorde tot de geëmancipeerden, wie dat naliet en zijn zelfontplooiingsplicht verzuimde, moest zelf op de blaren zitten.' (Mellink, 2014: 167-8).

### ***Hervormingen en bezuinigingen***

Vanaf de tweede helft van de jaren zeventig staat onderwijs, net als andere sectoren, onder druk van grootschalige bezuinigingen die moesten worden doorgevoerd. De bezuinigingen gingen gepaard met een andere visie op beleid en de aansturing ervan. Meer ruimte voor de markt en een overheid op afstand karakteriseren het beleid. Om de overheid te helpen zich meer bedrijfsmatig op te stellen wordt op het ministerie van onderwijs adviesbureau McKinsey ingeschakeld. De 'beleidscyclus' wordt het centrale concept en zorgt voor een departementale onderverdeling in beleidsontwikkeling, - uitvoering en toezicht. Onderwijs werd meer benaderd als een proces van 'input' en 'output' dat geëvalueerd moest worden in termen van kwaliteiten, uitgaande van opbrengsten. Het algemene denken over het vergroten van de efficiency leidde in de jaren tachtig tevens tot hervormingen gericht op schaalvergroting. In de jaren tachtig werd het kleuter- en lager onderwijs samengevoegd tot 'basisonderwijs' en werd het aantal scholen van ongeveer 17.000 teruggebracht naar ongeveer 8500 scholen. De leerplicht werd met een jaar vroegerd met als motto 'vijf jaar moet, vier jaar mag' (De Rooy, 2018: 250). En kwaliteit werd de focus van beleid. Minister Pais kwam in 1981 met een nota met als titel 'kwaliteit van onderwijs', een nota die volgens Mertens in zijn benadering 'sterk verbonden was met de empirisch-analytische wetenschapsbenadering: evalueren is "meten"; meten vindt plaats op basis van operationele definities van kwaliteit. Voor die metingen dienden instrumenten ontwikkeld te worden die

valide en betrouwbaar zijn voor de te onderscheiden systeemniveaus die het onderwijsstelsel kent: macro, meso en micro.' (Mertens, 2014: 13).

Voor het onderwijsstelsel was er in de jaren tachtig nog het 'probleem' van de middenschool. De WRR probeert in 1986 de gepolariseerde discussie te beëindigen door de aandacht te verschuiven van structuur naar inhoud van het onderwijs: de eerste jaren voortgezet onderwijs krijgen de vorm van 'basisvorming' waarin op elk schooltype een pakket van veertien vakken zou moeten worden gegeven. Volgens De Rooy kwam de wet erdoorheen 'omdat iedereen van het hele probleem af wilde zijn' (De Rooy, 2018: 251). De basisvorming kent nog een aantal wijzigingen en wordt in 2003 omgedoopt tot 'onderbouw' en daarmee in feite afgeschaft.

Voor grote delen van de bevolking is het de wens om hun kinderen via het onderwijs te laten stijgen op de maatschappelijke ladder. Voor wisselende regeringen is het een voortdurend punt van aandacht geweest om de arbeidsmarkt te laten bestaan uit voldoende, lager én hoger gekwalificeerd potentieel. Dit leidt in de jaren tachtig tot zorg over de aantrekkelijkheid van het beroepsonderwijs omdat het aandeel leerlingen dat dan doorstroomt naar het lbo (lager beroepsonderwijs) in 1982 gedaald is tot 35% van de jongens en 25% van de meisjes ten opzichte van respectievelijk 40% en 39% in 1970 terwijl de toestroom naar het avo (mavo, havo/vwo) in die periode sterk is gegroeid zoals de WRR in 1986 meldt in het rapport over de basisvorming (WRR, 1986: 54). De WRR stelt dat 'het huidige lbo eigenlijk vlees noch vis is. Het is algemeen zonder de status van een algemeen vormend schooltype te genieten; het is beroepsgericht zonder een daadwerkelijke beroepsopleiding te zijn.' (WRR, 1986:55). De Raad wijst op het ontstaan van een tweedeling in het voortgezet onderwijs: 'De meeste scholengemeenschappen richten zich hetzij op vwo/havo/mavo, hetzij op mavo/lbo.' (WRR, 1986: 226). Dit is een ontwikkeling die de Raad ongewenst vindt.

In de loop van de jaren negentig doet zich een nieuw probleem voor: Het aantal leerlingen loopt terug van ruim 1 miljoen in 1987 naar 866.000 in 1992. Daarbinnen loopt het aandeel voorbereidend beroepsonderwijs, vbo, terug van 21% naar 18% (Dijsselbloem, 2008). Dit leidt tot een roep om versterking en verbetering van het beroepsonderwijs. Staatssecretaris Netelembos werkt de plannen uit voor het voorbereidend middelbaar beroepsonderwijs, het vmbo, een samenvoeging van het voorbereidend beroepsonderwijs (vbo), het lager beroepsonderwijs (lbo), het middelbaar algemeen voortgezet onderwijs (mavo) en sommige vormen van voortgezet speciaal onderwijs. In 1999 voert het kabinet-Kok II de onderwijsvorm daadwerkelijk in, onder het VVD-ministerschap van Hermans. Aan de voorziene, sterk groeiende vraag naar hoger opgeleiden eind jaren negentig en begin deze eeuw wordt voorzien door de route van vmbo-mbo-hbo te versterken en perspectief te bieden tot en met een hbo-master zoals onderwijsminister Hermans en staatssecretaris Adelmund dat uitwerken in hun nota 'Grenzeloos leren' (Ministerie OCW, 2001). Zonder stelselwijziging heeft er daarmee een vernieuwing plaatsgevonden die volgens Bronneman-Helmers belangrijk is omdat ze verandering teweegbracht in de positie van verschillende schooltypen en leerlingstromen. Als gevolg van de samenvoeging van de bovenbouw van het vbo (leerjaar 3 en 4) en het middelbaar algemeen voortgezet onderwijs (mavo) ontstond feitelijk een tweedeling tussen vmbo aan de ene kant en havo/vwo aan de andere kant (Bronneman-Helmers, 2011: 72). Dit werd volgens haar nog verder versterkt door de inperking van mogelijkheden om door te stromen van mavo naar havo. Het zogenaamde 'stapelen' van

opleidingen is daarmee sterk teruggelopen (Bronneman-Helmers, 2011: 273). Op die manier is het vmbo gepositioneerd als de voorbereidende opleiding voor het middelbaar beroepsonderwijs en havo/vwo als voorbereidende opleidingen voor respectievelijk het hoger beroepsonderwijs en het wetenschappelijk onderwijs (Bronneman-Helmers, 2011: 73).

In haar analyse van het besluitvormingsproces met betrekking tot (vroeg) selectie (1990-2010) komt Bronneman-Helmers met een aantal aspecten die het verloop hebben bepaald. Ten eerste is er politiek onvoldoende draagvlak geweest om daadwerkelijk wat te doen aan de vroege selectie. Met name de VVD heeft geen problemen met selectie, eerder het tegendeel, want selectie leidt tot 'selection des meilleurs'. Hoewel het CDA aanvankelijk op de lijn van uitstel van selectie zit, verandert dat in de loop van de tijd. Het CDA legt meer prioriteit bij de zorg voor individuele leerlingen. Van de belangenverenigingen is het Nederlands Genootschap voor Leraren een tegenstander van uitstel. Dit genootschap stelt dat zij met homogene groepen hun opdracht het beste kunnen realiseren. Organisaties van schoolbesturen stellen zich volgens Bronneman-Helmers vooral strategisch en niet zozeer ideologisch op. Dit geldt ook voor schoolleiders. Zeker de invoering van de basisvorming is vooral pragmatisch opgepakt en zoveel mogelijk beleidsarm ingevoerd. Bronneman-Helmers ziet bij ouders een verschil in belangen. Hoogopgeleide ouders hebben in de regel een voorkeur voor categoriale scholen en weten hun weg wel te vinden. Dit geldt niet voor lager opgeleide ouders die daarbij niet goed vertegenwoordigd zijn in besluitvormingsprocessen. Vooral ouders van kinderen die wel voldoende capaciteiten hebben maar dat niet vertaald zien in schoolprestaties door bijvoorbeeld taalachterstanden of tragere ontwikkeling, hebben belang bij uitstel (Bronneman-Helmers, 2011: 281). Pleidooien voor uitstel van selectie maken veelal gebruik van onderzoeken die laten zien dat onderwijsprestaties van leerlingen uit lagere sociale milieus beter zijn in landen die later selecteren. Dit is ook waar de OECD op wijst; vroege selectie heeft negatieve gevolgen voor kinderen uit lagere sociale milieus en allochtone milieus. Vergelijkingen met andere onderwijsstelsels die wel hebben gekozen voor uitstel van selectie stuiten echter op de nodige problemen vanwege de verschillen in maatschappelijke context die mogelijk ook of wellicht meer een rol van betekenis spelen in eventuele prestatieverschillen. Bronneman-Helmers stelt dat het beleid gericht op uitstel van studie- en beroepskeuze en daarmee van selectie een drietal averechtse effecten heeft gehad. Als eerste stelt zij dat het streven naar uitstel de selectie verder verscherpt heeft door de tweedeling die sterker is geworden tussen vmbo aan de ene kant en havo/vwo aan de andere kant. Door de scores op de eindtoetsen mede te gebruiken voor het aantonen van rendement is de selectie met behulp van deze test verder verscherpt. Als tweede punt van versterking noemt Bronneman-Helmers de vorming van vakcolleges die het nodig maakten op 12-jarige leeftijd een beslissing te nemen omdat die colleges al in de onderbouw starten. Ten slotte wijst Bronneman-Helmers op de schaalvergroting die ertoe heeft geleid dat scholen de verschillende schoolsoorten in aparte vestigingen gingen onderbrengen. Met name de beroepsgerichte leerwegen zijn daarbij in aparte vestigingen ondergebracht wat de segregatie naar sociaal milieu heeft doen toenemen (Bronneman-Helmers, 2011: 284-5).

Van uitstel van selectie zoals aanvankelijk beoogd met de middenschool, is geen sprake meer. En door de aangebrachte tweedeling tussen avo- en beroepsonderwijs is er feitelijk een groter belang komen te liggen bij de selectie aan het eind van het basisonderwijs want daar moet nu het onderscheid worden aangebracht tussen wie via beroepsonderwijs zijn weg vervolgt, wie via ho-

ger onderwijs en wie direct kans maakt op wetenschappelijk onderwijs. Selectie aan het eind van basisonderwijs zou met de afschaffing van de basisvorming weer op de politieke agenda hebben moeten belanden, omdat het als thema uitsluitend 'geparkeerd' was in afwachting van een meer comprehensieve onderwijsstructuur. In de praktijk heeft selectie in de periode van zoeken naar uitstel ervan in de vorm van een nieuw onderwijsstelsel zelfstandig een weg gevonden in de praktijk van het funderend onderwijs. Ten tijde van de afschaffing van de basisvorming wordt de politieke agenda in de geest van die tijd gedomineerd door het vraagstuk van kwaliteit. Dat heeft gevolgen voor de manier waarop wordt gedacht over opbrengsten van onderwijs en de rol van toetsing daarbij. Dit krijgt meer aandacht en zorgt ervoor dat het onderwerp van selectie meer in de luwte blijft en niet terugkeert op de politieke agenda.

#### 4.8 De eerste vorm van eindtoetsing

Voor de toelating tot het v.h.m.o. is in 1965 een nieuw Koninklijk Besluit (KB) opgesteld, nog geënt op de dan bestaande schoolstructuur, maar uit de memorie van toelichting blijkt dat dit KB bedoeld ook bedoeld is voor de periode na de invoering van de Mammoetwet, want daarin staat 'dat de minister voornemens was de toelating tot het H.A.V.O./V.W.O.-brugjaar op overeenkomstige wijze te regelen, en zo is het inmiddels in de uitvoeringsbesluiten vastgelegd' (Deen, 1969: 84). Zoals van meerdere kanten geadviseerd, heeft het KB van 1965 voor de toelating tot het v.h.m.o. een keuzemenu waarbij één knoop duidelijk is doorgemaakt: alléén een geschiktheidsverklaring van het Hoofd van de lagere school voldoet niet langer voor toelating tot het v.h.m.o. Er is aanvullende informatie nodig en wel in één van de vier volgende vormen volgens het KB:

- Een toelatingsexamen, uit te voeren bij het v.h.m.o., tenminste gaand over Nederlandse taal en rekenen;
- Een proefklasse, verbonden aan de school;
- Een onderzoek naar de kennis en het inzicht van de kandidaat gedurende tenminste zijn laatste leerjaar aan de lagere school;
- Een psychotechnisch onderzoek.

##### 4.8.1 Het Koninklijk Besluit van 1965 en de Amsterdamse Schooltoetsen

Voor de Gemeente Amsterdam is dit Koninklijk Besluit lastig verenigbaar met de in deze gemeente in de jaren zestig gegroeide praktijk. Amsterdam heeft een eigen procedure voor toelating: het hoofd van de lagere school heeft de keus om ofwel direct een geschiktheidsverklaring voor het v.h.m.o. af te geven of bij twijfel een toelatingsexamen te adviseren. De in die tijd geldende landelijke verplichting voor lycea om te werken met een toelatingsexamen omzeilt de gemeente Amsterdam door te besluiten alle lycea-kandidaten te beschouwen als (potentiële) gymnasiumkandidaten (waarvoor tot dan toe geen toelatingsexamen geldt). Hiermee wijkt Amsterdam af van de gangbare praktijk in Nederland waarbij het v.h.m.o. de toelating regelt met een eigen toelatingsexamen. Het KB van '65 zet een streep door de aanpak zoals de gemeente Amsterdam die op dat moment hanteert. De gemeente zoekt een oplossing voor het ontstane probleem zonder terug te hoeven vallen op de herinvoering van een toelatingsexamen, uit te voeren door het v.h.m.o., de eerste optie van het KB. Het werken met proefklassen, de tweede optie van het KB, is op de schaal van Amsterdam niet realistisch. Het werken met proefklassen betekent dat leerlingen gedurende een langere periode moeten worden geplaatst in een gesimuleerde voortgezetonderwijsomgeving, uit te voeren door een samenwerking van lager en middelbaar

onderwijs, en dat in die geconstrueerde context een oordeel wordt geveld over dat functioneren van een leerling. Dit is logistiek en financieel niet uitvoerbaar voor de gemeente Amsterdam. Een psychotechnisch onderzoek, de vierde optie van het KB, stuit op vergelijkbare problemen. Voor dergelijke intelligentieonderzoeken is toestemming van ouders vereist en deze moeten door (test)psychologen verbonden aan psychologische centra een-op-een worden uitgevoerd. Ook deze optie zou de gemeente voor grote, logistieke problemen plaatsen gezien het relatief grote aantal deelnemers aan het v.h.m.o. Feitelijk biedt alleen de derde optie perspectief. De wethouder van onderwijs van Amsterdam, Koets, is dan al een aantal jaar voorzitter van het bestuur van het RITP, het Research Instituut voor Toegepaste Psychologie, in 1957 door de psycholoog A.D. de Groot opgericht. De wethouder wijst De Groot erop dat de uitslag van een schoolvorderingentest valt onder de nieuwe tekst van het KB (Cito, 1993a: 14). In eerste instantie adviseert het RITP gebruik te maken van bestaande intelligentietoetsen, maar stelt tegelijkertijd vast dat bestaande toetsen het nadeel hebben dat de opgaven weliswaar 'succes in het V.H.M.O. vrij goed voorspellen' (Samenwerkende Instituten, 1967: 18) en inhoudelijk voldoende aan lijken te sluiten bij de leerstof in het lager onderwijs, maar dat de geheimhouding van de opgaven<sup>20</sup> op grote problemen zal stuiten (Samenwerkende Instituten, 1967: 19). Bovendien bevatten de bestaande tests 'van het type intelligentietest' (Samenwerkende Instituten, 1967: 19) opgaven die zouden kunnen worden geschaard onder de criteria van een 'psychotechnisch onderzoek' waardoor toestemming van ouders toch nodig zal zijn of op bezwaren van psychologische centra kan stuiten. Dat maakt dat het RITP bij nader overleg toch met het voorstel komt een toets te ontwikkelen met compleet nieuwe opgaven, niet herkenbaar als bestaande 'psychotechnische' opgaven. De tijd dringt en het KB moet worden uitgevoerd. Een nieuwe toets ontwikkelen wordt alleen haalbaar als De Groot hulp krijgt. Die zoekt hij bij het Nutsseminarium dat dan onder leiding staat van de pedagoog Deen. De medewerking van het Nutsseminarium is in het licht van de standpunten ten aanzien van het aansluitingsvraagstuk in het verleden wel bijzonder. Deen onderbouwt de deelname door te stellen dat selectie niet het uitgangspunt is van het te ontwikkelen instrument maar dat de ontwikkelde toetsen meer het karakter hebben van een 'schooleindeonderzoek' (Deen, 1969: 94): 'Dat betekende, dat het onvruchtbare selectiegezichtspunt niet meer de opzet zou behoeven te richten, maar dat de te ontwikkelen schoolprestatietoets zou gaan functioneren in een determinatieproces, dat in het brugjaar wordt voortgezet. In die situatie wordt van het schooleindeonderzoek niet meer gevraagd dan bij te dragen aan de "grove zifting" der leerlingen bij de intrede van het brugjaar.' (Deen, 1969: 94).

De 'Samenwerkende Instituten', zoals het consortium zich noemt, krijgt op 11 maart 1966 van het Gemeentebestuur van Amsterdam per brief de opdracht 'een schooltoets te ontwerpen voor het toetsen van kennis en inzicht van alle leerlingen van de zesde klassen der openbare scholen voor gewoon lager onderwijs en mede te werken bij de uitvoering van het betreffende onderzoek' (Samenwerkende Instituten, 1967: 21). Een onderzoek niet alleen afgenomen bij de leerlingen die zich aanmelden voor het v.h.m.o. maar een onderzoek bij alle leerlingen die dan in de zesde klas zitten, in Amsterdam zo'n 6000 leerlingen in totaal.

---

20] Opgaven uit psychologische testen zijn bedoeld om geheim te blijven. Bij een grootschalig gebruik zoals in deze context voorzien, zou dat niet goed mogelijk zijn. Alle leerlingen, hun onderwijzers en de scholen zelf zouden bij een afname inzicht hebben in de opgaven.



In een artikel in het Schoolblad, het weekblad van de Algemene Bond voor Onderwijzend Personeel (ABOP), gaat De Groot in op mogelijke doelen die in zijn algemeenheid kunnen worden beoogd bij de constructie en toepassing van schooltoetsen en welke van die doeleinden passen bij de Amsterdamse schooltoetsen. Hij benoemt zes mogelijke doeleinden (Samenwerkende Instituten, 1967: 1):

- I Selectie voor het v.h.m.o.
- II Advies voor schoolkeuze aan de ouders
- III Opsporen van 'reserve' aan intellect
- IV Analyse van de stand van het onderwijs
- V Concretisering en ter discussie stellen van de (eind)doelstelling van het l.o.
- VI Gunstige beïnvloeding van het l.o.

Het eerste punt, selectie voor het v.h.m.o., is het doel van de wet waaraan de Gemeente Amsterdam moet voldoen en waarvoor ze te rade gaat bij De Groot. De Groot is over dat doel van de wet heel duidelijk in de verantwoording: hij zou aan de ontwikkeling van de toets niet zijn begonnen als de toets zuiver bedoeld zou zijn geweest voor 'de verbetering van de selectie voor het v.h.m.o.' (Samenwerkende Instituten, 1967: 5). Hij ziet weinig heil in het verbeteren van de selectie en noemt die ambitie 'dubieus' (De Groot, 1966: 4). Hij vat de uitkomsten van onderzoekingen van de afgelopen jaren samen door te stellen dat de huidige selectie dicht bij de grens van 'het haalbare' komt. Hij stelt dat het lastig is om van een leerling aan het eind van het basis-onderwijs de potentie te kunnen vaststellen: 'Wat er bereikbaar is, wordt namelijk sterk beperkt, ten eerste doordat de ontwikkeling van een puber onder invloed van grotendeels onbekende omgevingsfactoren maar heel gedeeltelijk te voorspellen is, ten tweede doordat ons voortgezet onderwijs met rekbare maatstaven werkt' (Samenwerkende Instituten, 1967: 4). Hij bestempelt de eis van het KB dan ook als iets waar 'wij – zeker 'wij in Amsterdam' – niet om hebben gevraagd' (Samenwerkende Instituten 1967: 4). Hij wil de selectie-eis combineren met andere doeleinden die hij belangrijker acht.

Over het tweede punt, een advies voor schoolkeuze aan ouders, is De Groot in de verantwoording heel kort: daar is de toets wat hem betreft niet voor bedoeld: 'Als ouders advies wensen, wenden zij zich gewoonlijk tot een psychologisch bureau of een bureau voor beroepskeuze' (Samenwerkende Instituten, 1967: 5).

Het derde mogelijke doel, het opsporen van 'reserve' aan intellect, is in die tijd een onderwerp waarvoor de nodige aandacht is. De Groot is kritisch over het doel om met de Amsterdamse Schooltoetsen reserve aan intellect te kunnen opsporen. De koppeling van de inhoud van de Schooltoetsen aan schoolstof staat dit doel volgens De Groot enigszins in de weg. Voor het goed kunnen opsporen van reserve aan intellect zijn volgens De Groot meetinstrumenten nodig die niet direct aan leerstof zijn gekoppeld omdat daarmee 'ongeschoold intellect' kan worden opgespoord. Voor de Amsterdamse Schooltoetsen ziet hij vooral de mogelijkheid om leerlingen die 'hoge' scores behalen maar zich niet hebben aangemeld voor het v.h.m.o. te wijzen op de mogelijkheid voor een 'hogere' schoolkeuze maar dit bestempelt hij als 'de, veel kleinere, 'reserve' (Samenwerkende Instituten, 1967: 6).

Waar de Amsterdamse Schooltoetsen vooral heel geschikt voor zijn in de optiek van de ontwikkelaars, is het vierde doel: het kunnen bieden van een 'analyse van de stand van het onderwijs, in een stad, een provincie, het land'. Als hoofdvraag daarbij ziet De Groot: 'in hoeverre hebben de leerlingen van verschillende scholen, aan het eind van een onderwijsperiode, inderdaad geleerd wat zij volgens de onderwijsdoelstellingen en het leerplan geacht worden te hebben geleerd. Om hierover informatie te verkrijgen is het middel bij uitnemendheid een op alle scholen van een bepaald type af te nemen schooltoets – die hierop berekend is' (Samenwerkende Instituten, 1967: 6). De Groot realiseert zich dat het vergelijken van scholen met elkaar op weerstand stuit, sterker, als bedreigend wordt ervaren (Samenwerkende Instituten, 1967: 7). Daar brengt hij het volgende tegenin: 'Mijn mening is, dat wij, in een democratie levend, van de angst, dat de inspectie (of de gemeente, of het Rijk) met zulke feitelijke gegevens 'iets lelijks zal gaan doen' moeten zien af te komen. Zelf weten wij – onderwijzers, leraren, hoogleraren – heel goed, dat de resultaten van ons onderwijs, ..., ook zeer sterk van de omstandigheden en van de aanleg van onze leerlingen afhangen; wij moeten òf van de wantrouwige veronderstelling, dat anderen, met name beleidsinstanties, dit niet zullen begrijpen en verkeerde conclusies zullen trekken.' (Samenwerkende Instituten, 1967: 7). Om aan de gevoelens van onrust tegemoet te komen, worden de resultaten van de toetsen in 1966 niet gedeeld met de Inspectie. Zelfs de deelnemende scholen krijgen hun eigen resultaten niet terug. Alleen het v.h.m.o. krijgt de scores van de leerlingen en dan alleen van de leerlingen die zijn toegelaten. De Groot is het hier niet mee eens: 'Zou de inspectie (afdeling onderwijs) de schooluitslagen ook in de toekomst niet in handen krijgen, dan is de betekenis ervan voor het gemeentelijk beleid uiteraard illusoir' (Samenwerkende Instituten, 1967: 7). De analyse van de stand van het onderwijs vindt hij uiterst belangrijk en 'een schooltoets – mits goed samengesteld – is een voortreffelijk middel, ja, het enige bruikbare middel daartoe' (Samenwerkende Instituten, 1967: 6).

Ook het vijfde doel om te komen tot 'concretisering van doelstellingen van het onderwijs' past bij de ambities van de Amsterdamse Schooltoetsen. De Groot hecht veel waarde aan minimumdoelen voor het lager onderwijs of ten minste overeenstemming over 'de grootste gemene deler' van het onderwijs. Hij vindt dat dat 'volkszaak' moet zijn. 'Dit doel is zeer belangrijk – daarmee zullen de meeste mensen in (en buiten) ons onderwijs het wel eens zijn.' De ontwikkeling van schooltoetsen ziet hij als een belangrijke manier om te komen tot een 'doordachte communis opinio' (Samenwerkende Instituten, 1967:8). Dat er geen doelen zijn, wijt De Groot aan 'angst' bij de tegenstanders: 'Waarschijnlijk zien degenen, die zo denken in hun fantasie de schooltoetsen al in handen van een dictatoriaal regelende of bureaucratisch verstarde overheidsinstantie, waartegen niets te beginnen is: een nachtmerrie à la '1984.' (Samenwerkende Instituten, 1967: 8). Hij wijst op de Verenigde Staten, waar het gebruik van schooltoetsen al verder gevorderd is en waar volgens De Groot schooltoetsen juist hebben geleid tot 'democratische discussies over en beweging in de onderwijsdoelstellingen'. En voegt hij eraan toe: "Verstarring' is wel het laatste, dat men de Amerikanen kan verwijten.' (Samenwerkende Instituten, 1967: 8).

De Groot gebruikt een 'omgekeerde' volgorde van werken: niet eerst overeenstemming bereiken over onderwijsdoelen alvorens aan de slag te gaan met de operationalisatie ervan, maar door aan de slag te gaan met concrete toetsopgaven wil hij komen tot concrete onderwijsdoelen met het idee dat aan de hand van concrete toetsvragen later het gesprek kan komen of 'wij willen, dat de leerlingen dat kunnen'. In het boek Studietoetsen spreekt hij over deze manier van werken

als 'leren weten wat men wil' (De Groot & Van Naerssen, 1969: 25). Het construeren van items 'activeert het doelstellingsgeweten.' 'Via de concretisering en operationalisatie van onderwijsdoelstellingen waartoe de constructie van studietoetsen itemschrijvers en itembeoordelaars min of meer dwingt, komen ook vanzelf andere principiële onderwijsvragen aan de orde.' (De Groot & Van Naerssen, 1969: 26). De Groot voegt daar aan toe dat 'voorwaarde voor die effectiviteit natuurlijk wel is, dat deze zaken voor discussie worden opengelegd' (Samenwerkende Instituten, 1967: 9).

Met discussie over de doelstellingen van onderwijs hangt het zesde doel, 'Beïnvloeding van het lager onderwijs', nauw samen. Onderwijs richt zich naar wat er in een schooltoets aan de orde komt, en zeker als het gaat om een schooltoets met selectiedoeleinden. De kunst is, in de optiek van De Groot, te zorgen dat het gaat om een 'gunstig effect' en niet om 'ongunstige effecten' (Samenwerkende Instituten, 1967: 9). Hij acht dat mogelijk als er goede overeenstemming is over de inhoud van schooltoetsen. Maar hij is zich ervan bewust dat toetsing er ook toe kan leiden dat leerkrachten zich in het onderwijs gaan beperken tot dat wat in de toets aan bod komt of hun leerlingen zelfs gaan trainen voor de toets: 'Het eenvoudigste en belangrijkste middel om een te grote, ongezonde invloed van de toets op het onderwijs te voorkomen, is echter: ervoor te zorgen, dat de afname minder het karakter van een examen krijgt.' (Samenwerkende Instituten, 1967: 10).

De betrokkenen bij de ontwikkeling van de Amsterdamse Schooltoetsen worstelen met de verenigbaarheid van de verschillende doelen van de schooltoetsen. In een artikel in *Pedagogische Studiën* stelt Warries (als medewerker van De Groot bij het RITP betrokken bij de Amsterdamse Schooltoetsen en later directeur van het RITP) dat 'constructie, scoring en analyse van een toets bestemd voor selectie zal leiden tot een toets die a) tamelijk moeilijk is, b) niet specifiek hoeft te zijn voor het genoten onderwijs en c) gemakkelijk een vergelijking van de prestaties van kandidaten toelaat.' Warries wijst daarbij op 5 hypothetische risico's die hij voorziet bij een grootschalig gebruik van dergelijke studietoetsen (Warries, 1971):

1. Men toetst niet wat geleerd is, maar men toetst de aanleg;
2. Men construeert selecterende toetsen terwijl men meent te evalueren of te diagnosticeren;
3. Men diagnosticeert niet, maar groepeerst;
4. Men evalueert niet voor feedback, maar voor beoordeling;
5. Men selecteert te veel omdat het zo gemakkelijk gaat: als men de toetsen maar moeilijk genoeg maakt, is er altijd een caesuur tussen 'slechte' en 'goede' studenten aan te brengen.

Vanaf het begin van de ontwikkeling van de eerste eindtoets weten de ontwikkelaars welke spanningen er zijn tussen de selecterende en de evaluatieve functie van de eindtoets. De Groot maakt melding van gesprekken binnen de Samenwerkende Instituten over de 'combineerbaarheid van de doelstellingen'. Zonder op de inhoud van de discussie in te gaan, volstaat hij in de verantwoording met de conclusie 'dat het mogelijk moet zijn schooltoetsen te ontwikkelen, waarvan de opgaven (items) uitsluitend een beroep doen op leerstof, die tot de 'Grootste Gemene Deler' van het lager-onderwijsprogramma behoort, maar die toch óók voldoende moeilijke vragen bevatten om voor doeleinden van selectie en opsporing van reserve goed genoeg naar boven te differentiëren' (Samenwerkende Instituten, 1967: 10). Hij voegt daaraan toe dat 'dit

dubbel ideaal al enigszins benaderd is [met de toets van 1966] maar dat dit in volgende jaren ongetwijfeld consequenter en beter kan' (Samenwerkende Instituten, 1967: 11).

#### 4.8.2 Opzet van de Amsterdamse Schooltoetsen

De inhoudelijke keuzes die worden gemaakt om te komen tot een inhoudelijke afspiegeling van de grootste gemene deler van het onderwijs, komen meer voort uit 'praktische dan uit principiële overwegingen'. Gegeven het feit dat de toets plaats zou vinden aan het einde van het zesde leerjaar van het lager onderwijs zou de stof uit dat laatste jaar centraal moeten staan. Ervaring met vormen van toetsen aan het eind van het basisonderwijs is er dan genoeg, omdat de meeste scholen al werken met eigen toetsen en daar maakt het consortium gebruik van: 'een viertal Hoofden van lagere scholen heeft waardevol opgavenmateriaal geleverd, dat later bewerkt en gebruikt kon worden'. Er wordt 'besloten de inhoud van de toets dit eerste jaar te beperken tot, in hoofdzaak, Nederlandse Taal en Rekenen (Samenwerkende Instituten, 1967: 33). Enerzijds omdat aangenomen kan worden dat de basisscholen daar wel allemaal (genoeg) aandacht aan besteden. En anderzijds omdat er ervaring is met taal en rekenen als onderdelen van het vaste programma van toelatingsexamens en 'geschiktheidsmetingen' zoals IQ-tests. De Groot gaat in op de mogelijke tegenwerping dat op 'basis-leerstof' gerichte toetsen geen prognostische waarde kunnen hebben ten aanzien van de selectie van de meest geschikte v.h.m.o.-kandidaten. Dit ziet hij 'als een misverstand'. Want, zo stelt hij, 'grondig inzicht in de "gewone" stof is voorbehouden aan de betere leerlingen'. Zowel 'concrete ervaringen in de testresearch als statistische overwegingen wettigen de verwachting, dat met item-series van een gemiddeld niveau een goede - en valide - differentiatie (spreiding) kan worden bereikt, ook "naar boven" (Samenwerkende Instituten, 1967: 33). Volgens De Groot 'is het genoegzaam bekend dat dit soort toetsen [en dan bedoelt hij schoolvorderingentoetsen in lijn met de Amsterdamse Schooltoetsen] een sterk voorspellende waarde heeft ten aanzien van schoolsucces' (Samenwerkende Instituten, 1967: 93). In de verantwoording is te lezen dat er bij de constructie en verwerking van de opgaven gewerkt is met opgaven die een spreiding in moeilijkheidsgraad kennen. Dit is volgens psychometrische uitgangspunten nodig om leerlingen van elkaar te kunnen onderscheiden. Ook moeilijkere items ( $p < 0,50$ ) kunnen daarbij worden beschouwd als een 'goed' item 'mits deze items goed differentiëren tussen goede en slechte leerlingen' (Samenwerkende Instituten, 1967: 123).

De verwachte overeenstemming over beoogde doelen voor taal en rekenen geldt in die tijd niet voor de zaakvakken, zo denken de Samenwerkende Instituten. De Groot ziet 'algemene kennis' wel als een relevante doelstelling om zicht te krijgen op de stand van het onderwijs. Daarom worden daarvan wel onderdelen opgenomen in de schooltoetsen maar die tellen niet mee in de uitkomst van de schooltoetsen.

Omdat De Groot een test ziet als een systematische procedure in het kader van het kunnen vergelijken (in lijn met de definitie van een test zoals Cronbach die beschrijft) moet men, in de woorden van De Groot 'precies weten wát men gaat observeren en vergelijken aan personen, en hòe men dit gaat doen' (Samenwerkende Instituten, 1967: 27). Dit brengt De Groot tot een tweetal testeisen: de eis van uniformiteit ('iedereen dezelfde toets') en de eis van objectiviteit ('iedereen op dezelfde manier beoordeeld'). Er wordt daarom uitsluitend gebruik gemaakt van meerkeuze-opgaven. Hiermee is volgens De Groot een 'objectieve score-bepaling' mogelijk omdat er 'geen interpretaties van de zijde van een beoordelaar' meer nodig zijn (Samenwerkende

Instituten, 1967: 27). Daarnaast ziet De Groot 'ieder meerkeuze-item als een hypothesetoets op zich' waarmee dit soort vragen 'een uniek hulpmiddel zijn bij de analyse en concretisering van onderwijsdoelstellingen' (Samenwerkende Instituten, 1967: 29). Ten slotte wijst hij erop dat meerkeuzevragen zich ook goed lenen voor 'psychometrisch-statistische analyses' en bovendien snel en makkelijk te verwerken zijn.

Het feit dat alle leerlingen van het openbaar onderwijs in Amsterdam deelnemen aan de toetsing en niet alleen de leerlingen voor wie het KB is bedoeld (de leerlingen die zich willen aanmelden voor het v.h.m.o.), acht De Groot van belang voor het kunnen opstellen én gebruiken van percentielscores. Het gebruik van percentielscores 'maakt de scores van alle leerlingen onderling vergelijkbaar, in één rangorde'. Zo kunnen leerlingen die naar het v.h.m.o. gaan, worden vergeleken met 'alle andere zesde-klas-leerlingen'. Het werken met percentielscores maakt 'de prestatie van elke leerling onmiddellijk vergelijkbaar met de prestaties van alle andere onderzochte leerlingen. De afzonderlijke leerling kan dus worden bekeken tegen de achtergrond van de totale groep.' (Samenwerkende Instituten, 1967: 36). Dit is nodig om een betrouwbare percentielscore-schaal te kunnen opstellen (Samenwerkende Instituten, 1967: 31). Een bijkomend voordeel van het werken met percentielscores is volgens De Groot dat dit systeem 'voor niet-testspecialisten gemakkelijk te hanteren en te begrijpen is: de percentielscore geeft eenvoudig de plaats van de leerling(-score) in de groep aan' (Samenwerkende Instituten, 1967: 36).

De eerste Amsterdamse schooltoetsen worden op twee momenten afgenomen: een afname in maart en een in april. Voor iedere afname wordt per leerling de percentielscore bepaald voor taal I en rekenen I en de gecombineerde percentielscore. En voor de tweede afname worden de percentielscores bepaald voor taal II en rekenen II en ook de gecombineerde percentielscore. Zo krijgt iedere leerling feitelijk twee keer een plek in de rangorde, gebaseerd op twee gecombineerde scores voor taal en rekenen. Afgaan op de totaalscores, de optelling van taal en rekenen, is volgens De Groot 'in het algemeen betrouwbaarder als meetuitkomst en prognostisch waardevoller' (Samenwerkende Instituten, 1967: 36). In de verantwoording wijst de Groot er met regelmaat op de nodige 'omzichtigheid in de hantering van de toetsresultaten te (willen) bevorderen'. Er wordt nadrukkelijk het advies gegeven 'toetsresultaten niet te gebruiken voor de afwijzing van een kandidaat, indien de kandidaat anders (bij afwezigheid van de schooltoets) wél geaccepteerd zou zijn.' Geven de twee opgetelde percentielscores verschillende aanwijzingen, dan pleit hij ervoor de leerling 'the benefit of the doubt' te geven. Want, zo stelt hij, 'een hoge score bewijst eerder dat een leerling iets kan, dan een lage score dat hij/zij het niet kan; men maakt een zo groot mogelijk aantal opgaven nu eenmaal niet 'toevallig' goed, terwijl het best 'toevallig' fout kan gaan (indispositie, en dgl.)'. En hij stelt dat er aan de percentielscores niet een te exacte interpretatie kan worden toegekend omdat 'de meetbetrouwbaarheid van deze percentielscores natuurlijk verre van perfect is'. Hij onderstreept nadrukkelijk dat de meting niet meer is dan een indicatie en nadrukkelijk géén geschiktheidsoordeel: 'de Samenwerkende Instituten leveren slechts uitkomsten op de toets. Deze uitkomsten zijn bedoeld als een bijdrage voor de beslissing van diegenen die hier te beslissen of te adviseren hebben'. Als criterium om de juistheid van de voorspelling van de uitkomsten van de toets mee te vergelijken gebruikt hij het advies dat door de toeleidende lagere school gegeven is: 'men kan de percentielscores natuurlijk in verband brengen met het feit, dat in Amsterdam ca. 20% van de zesde-klassers naar het v.h.m.o. gaat' (Samenwerkende Instituten, 1967: 69). De Groot vergelijkt de scores met de conclusie van de

lagere school, vertaald in het al dan niet aangemeld zijn voor het v.h.m.o. De correlaties zijn in de optiek van De Groot 'zeer hoog' wat hem brengt tot de conclusie dat 'de tests [bedoeld worden de 'subtests'] goed in staat zijn om te differentiëren tussen goede en slechte leerlingen – 'goed' en 'slecht' dan volgens de opgave voor het v.h.m.o.' (Samenwerkende Instituten, 1967: 108, tabel 5). De Groot acht het gegeven 'de aanmelding (tegenover niet-aanmelding) een bruikbaar "substituut-criterium" (Samenwerkende Instituten, 1967: 133), waarmee hij de validiteit van de Amsterdamse Schooltoetsen onderbouwt: 'De correlaties met dit criterium worden al 'validiteiten' genoemd in dit rapport.' (Samenwerkende Instituten, 1967: 133). En hij stelt dat 'deze validiteitsbevindingen een krachtige steun verlenen aan de veronderstelling, dat (ook) de schooltoets in staat is de v.h.m.o.-geschikten eruit te halen' (Samenwerkende Instituten, 1967: 133).

De Samenwerkende Instituten hebben de scores op de toetsonderdelen ook vergeleken met de achtergrondvariabele van milieu van herkomst, een milieu-indicatie die zij op basis van een serie vragen aan leerlingen, hebben kunnen herleiden. De Amsterdamse Schooltoetsen laten een verband zien tussen sociaal milieu en toetsprestaties: alle vier de hoofdtests blijken 'ongeveer even sterk te correleren met de beoordeling van het sociale milieu' (Samenwerkende Instituten, 1967: 112). Daarmee zijn de uitkomsten van de Amsterdamse Schooltoetsen in lijn met bijvoorbeeld de uitkomsten van IQ-tests zoals in het onderzoek van Van Heek gebruikt: een hogere score op de toe blijkt verband te houden met een hogere inschaling wat betreft sociaal milieu.

Hoewel De Groot (Samenwerkende Instituten, 1967: 137, tabel 16) laat zien dat de gebruikte toetsen gepaard gaan met de nodige onnauwkeurigheid bij de meting en dat het risico van een verkeerde score er wel is, blijkt hij niet ontevreden. Wat betreft de nauwkeurigheid van de juistheid van het individuele advies 'zou men ze graag nauwkeuriger hebben' zo stelt hij. 'Maar voor statistisch gebruik is de nauwkeurigheid van de schooltoets 'goed te noemen'. En, voegt hij eraan toe: 'Ongetwijfeld beter dan van de meeste, zo niet alle andere instrumenten die er voor schoolkeuze- en selectiedoeleinden in gebruik zijn.' (Samenwerkende Instituten, 1967: 137).

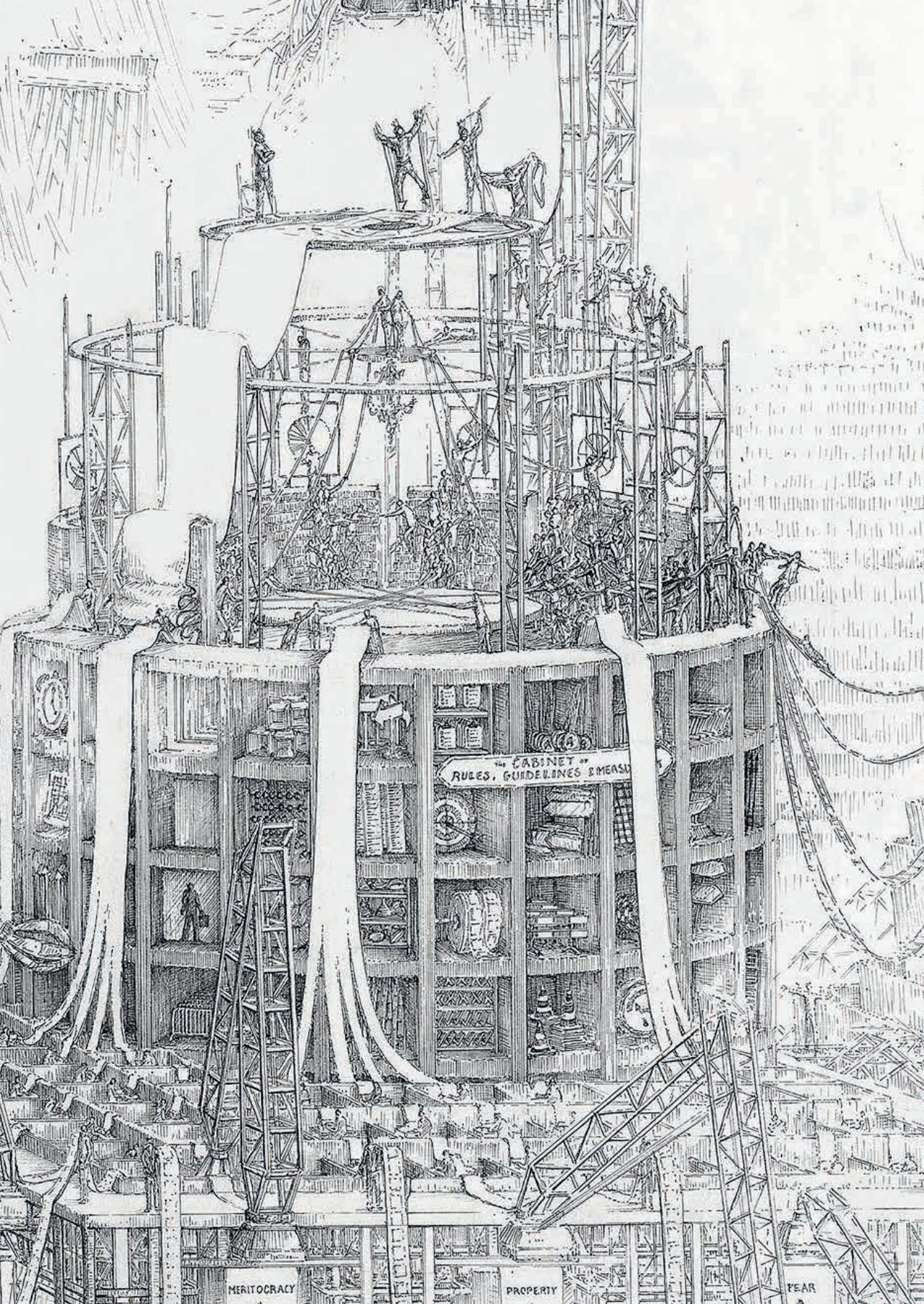
In 1968 krijgt Nederland zijn 'eigen' toetsinstituut, het Centraal Instituut voor Toetsontwikkeling, Cito. Hoewel De Groot al vanaf 1958 heeft geijverd voor een dergelijk instituut, naar voorbeeld van het ETS in de Verenigde Staten, is hij er nooit in dienst geweest. Maar bij Cito is hij de boeken ingegaan als 'Cito's geestelijk vader' (Cito, 2019: 24). De Amsterdamse Schooltoetsen worden direct bij de oprichting van het Cito ondergebracht bij dit nieuwe toetsinstituut. De benaming verandert in Cito-toets maar inhoud, vorm en opzet blijven in hoge mate gelijk, tot op de dag van vandaag. De toets is een zelfstandig product van Cito en weet, geholpen door het KB, uit te groeien tot een toets waaraan in 1969 zo'n 40.000 leerlingen uit heel Nederland deelnemen en die begin van deze eeuw door ongeveer 85% van de scholen (170.000 leerlingen) wordt gebruikt bij de overgang van basis- naar voortgezet onderwijs (De Regt, 2004).

## 4.9 Samengevat

De veranderingen in de samenleving en de mogelijkheden die het volgen van onderwijs lijken te bieden voor het stijgen op de maatschappelijke ladder leiden tot druk op instroom in middelbaar onderwijs. Naast zorg om de uitval is er nu vanuit financieel en economisch oogpunt een extra reden voor selectie. Daarmee is selectie ook niet meer uitsluitend een vraagstuk geworden voor discussie tussen lager en middelbaar onderwijs. Vanuit nieuwe perspectieven krijgt het vraagstuk belangstelling. Het belang van het individu en zijn vorming zoals dat in de jaren zestig opkomt,

wordt ingepast in het meer economisch ingegeven streven naar onderwijs waar 'iedere leerling op de juiste plek' terecht komt zodat economische doelen, nationaal en internationaal, gediend kunnen worden. De komst van de Amsterdamse Schooltoetsen biedt nieuwe perspectieven op het vraagstuk van selectie. In het licht van bijna een eeuw van zoeken naar een adequate vorm van selectie, is het voorstel van de Amsterdamse Schooltoetsen verfrissend, onder andere door de toepassing van nieuwe technologische ontwikkelingen op het gebied van digitalisering die samenhangen met het gebruik van meerkeuzevragen. Het nieuw toegevoegde element van meten met het doel zicht te krijgen op opbrengsten van (onderwijs)investeringen past goed bij de meer onderwijs-economische manier van kijken die nationaal en internationaal opkomt. Terugkijkend kan worden gezegd dat er met de komst van de eerste vorm van eindtoetsing minder nieuws onder de zon was dan het leek. De aanpak bleef gericht op het verbeteren van selectie en niet op een andere, nieuwe aanpak van het middelbaar onderwijs zelf. Wat betreft inhoud zijn de Amsterdamse Schooltoetsen niet veel anders dan de toelatingsexamens zoals die al jaren circuleerden waar taal en rekenen een prominente plek innamen. En het type opgaven heeft veel overeenkomsten met gangbare opgaven uit intelligentie-metingen. De psychometrische uitgangspunten die bij de Amsterdamse Schooltoetsen werden toegepast, zijn overgenomen van de manier van werken zoals die dan al een halve eeuw gangbaar is bij het testen van intelligentie: differentiëren tussen leerlingen is het primaire doel waarbij leerlingen met elkaar worden vergeleken en worden geordend in percentielgroepen uitgaand van een normaalverdeling. De uitkomsten worden met behulp van correlatieve technieken vergeleken met de schooladviezen waarmee de bestaande toewijzing aan het middelbaar onderwijs het ijkpunt vormt voor de validering van de nieuwe toetsen. Daarmee rijst van deze Amsterdamse Schooltoetsen eerder een beeld van 'padafhankelijkheid': gebeurtenissen en processen in het verleden hebben eerder een dynamiek in gang gezet waarvan nauwelijks afgeweken wordt bij de ontwikkeling van deze 'nieuwe' toetsen. Dat vormt tevens een verklaring voor de vrij gemakkelijke acceptatie van de Amsterdamse Schooltoetsen. De toetsen, beperkt in hun vernieuwingen wat betreft inhoud en selectie-uitkomsten, hebben er juist daardoor voor kunnen zorgen dat de acceptatie van de toetsen door het bestaande onderwijs redelijk eenvoudig is verlopen. En dat heeft weer bijgedragen aan de geloofwaardigheid en waarde die aan deze toetsen is toegekend.

Het onderwijs wordt gezien als de vijver waaruit de besten gevestigd moeten worden zodat Nederland mee kan in de vaart der volkeren. Planmatig moet worden gezorgd dat Nederland voldoende (maar niet te veel) hoger opgeleiden heeft. Onderwijsplanning wordt onderdeel van beleid en de concrete taak van selectie wordt belegd bij het onderwijs zelf. De ontwikkelingen in de samenlevingen leiden wel tot 'anders vissen' dan in de tijd dat afkomst bepalend was voor deelname aan vormen van onderwijs: verdiensten in termen van talent en inzet worden meer bepalend. Omdat het bepalen van talent geen waarde vrije invulling kent, wordt vooral gezocht naar prestaties, gekoppeld aan resultaten van eerder genoten onderwijs, als voorspellers van potentie voor de toekomst. Die prestaties van individuen worden behalve als beoordelingscriteria voor selectie van passend vervolgonderwijs ook bruikbaar geacht als indicator van geleverde kwaliteit van onderwijs. Dit staat centraal in hoofdstuk 5.



THE CABINET OF  
RULES, GUIDELINES & MEASURES

MERITOCRACY

PROPERTY

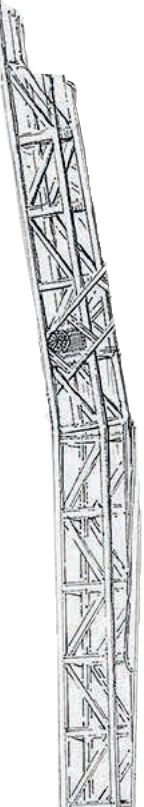
FEAR





# HOOFDSTUK 5

Onderwijskwaliteit en toetsing



## Hoofdstuk 5 Onderwijskwaliteit en toetsing

*'Elke normering, elk protocol, elk model houdt een reductie in van de veelkleurige werkelijkheid waarin burgers leven en uitvoerders hun functie uitoefenen. Alleen door ruimte voor de uitvoerders kan werkelijk 'maatwerk' worden geleverd. Alleen zo kan samenwerking tussen verschillende uitvoerders bij de oplossing van complexe problemen succesvol zijn.'*

*Uit: Herman Tjeenk Willink, 'Groter denken, kleiner doen' (2019)*

Ter ere van het 25-jarig bestaan van de Stichting voor Onderzoek van het Onderwijs (SVO) in 1990 wordt een internationale conferentie georganiseerd over het meten van de kwaliteit van onderwijs. In de introductie van de bundel naar aanleiding van de conferentie zegt Vedder dat het 'Leitmotive' van de bundel is: 'measuring the quality of education is possible, but we have not yet managed to do so in a satisfactory manner' (Vedder, 1992: 19). Het willen en kunnen meten van kwaliteit van onderwijs op nationaal en internationaal niveau is eind van de vorige eeuw een belangrijk onderwerp van onderzoek en debat in de westerse wereld. Niet omdat het nieuw is, maar omdat de belangen en de problemen die ermee gepaard gaan groot zijn. Idenburg schrijft in 1960 al over onderwijs in relatie tot opbrengsten. Daarbij begint hij met een waarschuwing: 'Nu wij voor het eerst het begrip "rendement" hebben geïntroduceerd, moet hier evenwel een waarschuwing worden ingelast. Het begrip komt uit het bedrijfsleven. Het duidt daar op de opbrengst, welke met bepaalde investeringen wordt verkregen. Het diploma-resultaat is uiteraard slechts één van de aspecten, waaronder de rentabiliteit van het onderwijs kan worden gezien. Er zijn andere, veel belangrijkere vruchten van het onderwijs, welke zich niet of niet zo eenvoudig laten meten, maar die daarom niet minder wezenlijk zijn. [...] Laten wij ons niet schuldig maken aan het misverstand, dat datgene wat niet exact kan worden vastgesteld en zich daardoor aan de statistiek onttrekt, niet zou bestaan!' (Idenburg, 1960: 485). Hij concludeert dat voor onderwijs geldt: 'De vruchten, die het afwerpt, kunnen niet worden gewogen.' (Idenburg, 1960: 517). Vanuit de onderwijswereld zelf is er vooral de intentie problemen in het onderwijs te kunnen definiëren en te verbeteren met behulp van instrumenten, veelal met het belang van het individuele kind op het netvlies. Die zoektocht lijkt aan de oppervlakte goed samen te gaan met de het denken over opbrengsten en verantwoording, 'accountability', zoals dat ook in de publieke sector onder invloed van het New Public Management gestalte krijgt. In de context van onderwijs leidt het accountability-denken tot de notie dat scholen en hun docenten verantwoordelijk gehouden kunnen en moeten worden voor succes en falen van onderwijs. Dat maakt dat vanuit het onderwijs zelf, zeker in het begin, aarzelend wordt gekeken naar instrumenten die opbrengsten van onderwijs in kaart zouden moeten brengen. Onderzoek laat zien dat de invloed van scholen op het succes van hun leerlingen beperkt is. Bovendien ontbreekt consensus over wat kwaliteit van onderwijs is; kwaliteit van onderwijs is een 'creative process of ongoing redefinition' (Vedder, 1992: 4). Voor het bepalen van geschiktheid van instrumenten voor het meten van onderwijskwaliteit betekent dit dat het op voorhand lastig te bepalen is of met bepaalde instrumenten daadwerkelijk de kwaliteit van onderwijs kan worden gedefinieerd. Veelal alleen achteraf, als blijkt hoe een meting uitwerkt, bij de discussie over de uitkomsten, wordt duidelijk hoe een instrument zich tot onderwijskwaliteit verhoudt (Vedder, 1992: 3). Dit zoeken en twijfelen van onderwijskundigen over nut en noodzaak van het vraagstuk van rendement in relatie tot onderwijs en onderwijskwaliteit blijkt in de praktijk ruimte te bieden voor het dominante discours over accountability,

toegepast op onderwijs. Hoe denken over kwaliteit en onderwijs zich heeft ontwikkeld en welke rol de overheid daar in is gaan spelen, met behulp van toetsen, staat centraal in dit hoofdstuk. Eerst worden de belangrijkste economische en maatschappelijke ontwikkelingen geschetst waarin het accountability-denken ook in Nederland gestalte heeft gekregen (5.1). Vervolgens wordt onderzocht hoe in Nederland tegen kwaliteit van onderwijs wordt aangekeken (5.2), hoe toezicht in relatie tot kwaliteit vorm en inhoud heeft gekregen (5.3) en welke rol toetsing, en vooral de eindtoets, daarbij is gaan vervullen (5.4).

## 5.1 Veranderende rol van de overheid

De oliecrisis en de economische recessie van de jaren zeventig versterken bij met name westerse regeringen het gevoel dat de economische politiek, met name gebaseerd op het gedachtegoed van de Britse econoom Keynes, geen bescherming heeft kunnen bieden tegen deze crises: hoge inflatie, financieringstekort, werkloosheid rond de 10 procent en verdamping van de winsten in het bedrijfsleven wat leidt tot een gebrek aan investeringen (Van Dam, 2009: 17). Omdat de Keynesiaanse politiek in de decennia na de Tweede Wereldoorlog nauw verweven is geraakt met de waarden van solidariteit en volledige werkgelegenheid waarop de Europese verzorgingsstaat is gebouwd, maakt dat kritiek op de economische politiek van Keynes ook al snel betrekking heeft op de verzorgingsstaat (Achterhuis, 2010: 100). De hoop en het vertrouwen van de samenleving van de jaren zestig en zeventig slaan om in onzekerheid en gebrek aan vertrouwen (Van Dam, 2009: 18). In dat vacuüm ontstaat ruimte voor de politieke wind van het neoliberalisme. Belangrijk uitgangspunt van het neoliberalisme is dat groeiend overheidsingrijpen in de economie in de vorm van sociale zekerheid en directe economische interventies het marktmechanisme (te veel) aan banden legt. De overheid krijgt met haar ingrijpen dan te veel politieke en economische macht, wat leidt tot het verdwijnen van individuele vrijheid en verantwoordelijkheid. De markt zelf kan deze samenballing van macht tegengaan door marktwerking centraal te stellen en de overheid daarop te laten toezien. De overheid moet alleen ingrijpen daar waar het vrije spel van de markt gehinderd wordt. Het neoliberalisme heeft ook gevolgen voor het denken over sociale zekerheid, de gemeenschap en het gezin. Voor deze domeinen wordt ruimte bepleit voor het individu en zijn eigen verantwoordelijkheid, tegen de 'betutteling' van de staat. Neoliberalen stellen het functioneren van de overheid ter discussie of zelfs in een kwaad daglicht, als eerste in de Verenigde Staten en het Verenigd Koninkrijk. Overheidsinstanties zouden te bureaucratisch zijn en te weinig de belangen van de samenleving en burgers dienen. Er is sprake van 'bureaucracy bashing' (Korsten, 2011: 74). Er komt een roep om verandering en er wordt gezocht naar wegen voor daadwerkelijke vernieuwing van de overheid. Het beleid van president Reagan in de Verenigde Staten – Reaganomics – en van premier Thatcher in het Verenigd Koninkrijk – Thatcherism – staan model voor veranderingen waar ook andere westerse landen naar op zoek zijn. Dat als liberaal te karakteriseren beleid heeft de kenmerken van een overheid die terugtreedt en bezuinigt en een markt die meer ruimte moet krijgen (Noordegraaf, 2004: 85).

### 5.1.1 Overheid op afstand

Het neoliberale denken geeft een boost aan nieuwe ideeën over het functioneren van overheidsinstanties. In de jaren tachtig en negentig vertaalt zich dat in een meer bedrijfsmatige aansturing van publieke organisaties, wat het label New Public Management (NPM) krijgt (Hood, 1991; Pollitt, 1993): 'Het management van en in het openbaar bestuur moest bedrijfsmatiger, zakelijker en prestatiegerichter (oriëntatie), teneinde de efficiency, de effectiviteit en de kwaliteit van

publieke dienstverlening te vergroten (waarden).’ (Noordegraaf, 2004: 78). NPM positioneert zichzelf niet als een wetenschappelijke stroming maar als een beweging die opteert voor de hervorming van overheidsinstanties naar meer marktgerichte organisaties. Het NPM, gestart in de VS en het VK krijgt vervolgens navolging in Australië en Nieuw Zeeland. De successen daar maken dat NPM ook bij andere OECD-landen onder de loep genomen wordt en ingebed wordt in tal van OECD-landen waarmee het gedachtegoed gezien kan worden als een wereldwijde stroming. Bij de meeste auteurs die zich met NPM bezighouden, komen omschrijvingen terug waarin NPM gezien wordt als een verzameling opvattingen en instrumenten over hoe publieke dienstverlening verbeterd kan worden, ontleend aan het bedrijfsleven (Hood, 1986; Osborne & Gaebler, 1993; Pollitt & Bouckaert, 2004; Korsten, 2011; Noordegraaf, 2004). NPM is niet te zien als een duidelijk gedefinieerde en afgebakende benadering van het managen van publieke organisaties, maar als een verzameling van inzichten, modellen, en vooral praktijken die tot doel hebben publieke organisaties resultaatgerichter, gecoördineerder en efficiënter te laten werken (Noordegraaf, 2004: 86). Hood heeft op basis van onderzoek naar de uitingen van NPM in verschillende landen, zeven, elkaar deels overlappende, doctrines afgeleid die in de meeste discussies over NPM terugkomen:

1. ‘hands-on professional management;
2. explicit standards and measures of performance;
3. greater emphasis on output controls;
4. shift to disaggregation of units;
5. shift to greater competition;
6. stress on private sector styles of management practices;
7. stress on greater discipline and parsimony in resource use’ (Hood, 1991: 4/5).

Een belangrijke verklaring voor het succes van NPM ziet Hood in het feit dat het gepresenteerd is als een algemeen toepasbaar model; ‘a public management for all seasons’. De universele waarde ervan komt enerzijds door het feit dat binnen het algemene raamwerk tal van variaties mogelijk zijn en anderzijds door de claim dat het een a-politiek raamwerk zou zijn (Hood, 1991: 8). Lorenz geeft een meer taalkundige verklaring voor de populariteit van de stroming. NPM wordt in de jaren ’80 volgens hem neergezet als een aanpak om de publieke sector efficiënt en transparant en daarmee ‘accountable’ te maken. In het dagelijks taalgebruik betekent ‘accountability’ volgens Lorenz: ‘Being responsible, with the connotation of giving an account of one’s actions.’ Het woord ‘accountable’ heeft in het dagelijks spraakgebruik een positieve waarde terwijl het tegenovergestelde van ‘accountable’, ‘inaccountable’, min of meer synoniem is met ‘irresponsible’. Daarmee is, volgens Lorenz, ‘accountability assumed to be an intrinsically desirable goal, and nobody ever claims that one can have ‘too much’ accountability—the pressure is always for more’ (Lorenz, 2012: 618).

Belangrijke beïnvloeders van de NPM-beweging, Osborne en Gaebler, zien oplossingen voor het verbeteren van het functioneren van overheidsinstanties in meer prestatiegerichtheid, innovatie, creativiteit, meer marktgerichtheid en vraagsturing van publieke organisaties. Zij komen met tien aanbevelingen voor een nieuwe overheid die in hun optiek meer gericht moet zijn op een missie (in plaats van op regels), op wat klanten willen, op het bereiken van resultaat, op bevordering van concurrentie in plaats van het vasthouden aan een overheidsmonopolie en op benutting van

expertise en energie van ambtenaren die niet alles zelf hoeven uit te voeren (Korsten, 2011: 75). Zij maken hun denkbeelden ook meer specifiek voor het onderwijs dat in hun ogen op dat moment meer lijkt op een communistische economie dan op een vrije markteconomie (Osborne & Gaebler, 1993: 319). Voor wat betreft de onderwijs-'markt' wordt duidelijk dat zij de ouders zien als de klanten die op de vrije markt op zoek gaan naar het best passende onderwijs voor hun kinderen. Scholen moeten zorgen dat ze zoveel mogelijk leerlingen aan zich weten te binden door te komen met voor ouders aantrekkelijk onderwijs met goede resultaten die aan het publiek bekend gemaakt worden. Dat moeten de scholen zelf doen, niet de districten, de staten of de federale overheid. Scholen moeten daartoe eigen budgetten hebben en zelf over hun personeelsbeleid kunnen gaan (Osborne & Gaebler, 1993: 316).

Het willen veranderen van de overheid en aan de overheid gelieerde instanties is door NPM veelal gelinkt aan een gebrek aan vertrouwen van de burger in de overheid vanwege achterblijvende prestaties van de overheid. Op basis van het idee van 'slecht functioneren en te weinig vertrouwen' wordt verandering bepleit van de 'falende' overheid. Want, zo is de redenering, als overheden gaan leveren wat burgers willen, kan het vertrouwen hersteld worden: congruentie tussen vraag en aanbod is daarbij van belang. De hervormingen van de publieke sector hebben echter (nog) niet geleid tot een toename van vertrouwen van de burger in de overheid (Pollitt en Brouckaert, 2004). In 2001 wijst de OECD erop dat door de vereisten ten aanzien van transparantie en de verbetering van communicatie zoals door NPM nodig geacht, de gebreken van de publieke sector juist aan het licht komen en daarmee eerder wantrouwen kunnen aanmoedigen dan wegnemen (Van de Walle, 2011: 11).

NPM heeft invloed op het denken over het functioneren van de overheid. Quené stelt dat de ondernemende overheid wordt gezien als een reactie op de bureaucratische overheid die te duur en niet efficiënt is en niet flexibel genoeg in de nieuwe tijd (Quené, 2018: 143). Modelmanagers van het NPM hebben opvattingen over hoe de overheid moet functioneren: de overheid moet niet 'roeien' maar 'sturen' waarbij 'sturen' wordt gezien als 'zorgen dat er geroeid wordt'. Daarnaast moeten overheidsorganisaties zich richten op hun missie, moeten ze hun middelen efficiënt besteden en moeten ze ondernemend zijn (Quené, 2018: 143). Scheiding tussen beleid en uitvoering maakt het mogelijk om flexibel op veranderingen te reageren volgens de opvattingen van NPM (Quené, 2018: 144). Quené stelt dat de scheiding tussen beleid en uitvoering sinds de jaren negentig ook in Nederland een mantra is van overheidsbeleid (Quené, 2018: 144). Belangrijk in het licht van deze ontwikkeling is de ambtelijke nota van Geelhoed uit 1983 met als titel 'De Intervenierende Staat: Aanzet voor een instrumentenleer'. De nota bevat kritiek op het gangbare overheidsbeleid en de recente pogingen tot organisatorische verbetering daarvan. Het handelen van de overheid moet zich volgens Geelhoed gaan ontwikkelen van een overheid die roeit naar een overheid die stuurt. In een reactie op dit rapport stelt de politiek wetenschapper Van Gunsteren, die de analyse onderschrijft, dat 'controle herwinnen door afstand te nemen, zoals het stuk voorstelt, een zinnige strategie kan zijn maar even goed kan leiden tot niets doen, staatsonthouding, het afsterven van de staat' (Van Gunsteren, 1984: 159). De nota van Geelhoed heeft ook voor onderwijs geleid tot inspanningen om deregulering en autonomievergroting voor de context van onderwijs nader uit te werken. Van Wieringen stelt dat 'deregulering en autonomievergroting ingrijpende beleidsprogramma's zijn geweest die meer dan 10 jaren lang het onderwijs beheersen' (Van Wieringen, 1996: 155). Van Wieringen citeert Bekke die, terug-

kijkend op het onderwijsbeleid in die periode, de Nederlandse overheid in de jaren tachtig als 'heroverweger' typeert en het beleid in die periode als 'afstandelijk onderwijsbeleid' (Van Wieringen, 2000: 32). Dereguleringsoperaties gaan vanaf 1985 van start en hebben hun effecten vanaf het begin van de jaren negentig. Deregulering in onderwijs is volgens Bekke gericht op het verminderen van regels van de landelijke overheid, wat ervoor moet zorgen dat scholen en schoolbesturen in openbaar en bijzonder onderwijs autonomer kunnen optreden (Meijnen & Van Wieringen, 2000: 34). De ingezette deregulering verandert echter binnen een periode van tien jaar in een actieve regulering waarbij meetbare prestaties centraal worden gesteld en in hoge mate bepalend worden in het licht van de bewaking van de kwaliteit van onderwijs.

### 5.1.2 Prestaties meten

Het werken met prestatiemetingen gebaseerd op prestatie-indicatoren is inherent verbonden geraakt met de veranderende taakopvattingen van en voor de overheid. Het gebruik van prestatie-indicatoren is toegenomen en ze worden vooral ingezet voor externe verantwoording (Pollitt & Brouckaert, 2004: 90-93). Meten is daarmee een belangrijk onderdeel geworden van management: metingen bepalen wat er gebeurt (Quené, 2018: 151).

Van de Walle wijst op het feit dat vormen van verantwoording op basis van wederzijds vertrouwen worden ontmanteld en vervangen door vertrouwen gebaseerd op cijfers en kennisinzage, vooral gebaseerd op prestatie-standaarden (Van de Walle, 2011: 11). Hij noemt het ironisch dat NPM beoogt het publieke vertrouwen in de overheid terug te brengen door op wantrouwen gebaseerde controlemechanismen en klachtenprocedures in te voeren (Van de Walle, 2011: 2). NPM kan zo feitelijk leiden tot nieuw wantrouwen als er geen vertrouwen is in de mechanismen die gehanteerd worden bij het vormgeven van meer supervisie en regulering van de publieke sector zoals audits en inspecties. Voor het gebruik hiervan is juist weer vertrouwen nodig, vertrouwen in de metingen, de meetinstrumenten en de informatiebronnen die gebruikt worden om het goed functioneren van overheidsinstanties te kunnen aantonen (Lorenz, 2012: 608). De Bruijn stelt dat een gematigd gebruik van prestatiemetingen aan te bevelen is omdat er dan 'minder prikkels voor pervertering bestaan en dat de positieve effecten van prestatiemetingen meer kansen krijgen' (De Bruijn, 2004).

Korsten et al. stellen dat wie prestaties van organisaties wil meten, ze wil vergelijken en rangordenen, en daarna wil verklaren welke organisatie, opleiding of andere eenheid van handelen waarom op welk vlak de beste is en wil aangeven wat nodig is om de beste of in elk geval een betere te worden, doet aan 'benchmarking' (Korsten et al., 2019: 21). Zij wijzen erop dat achter benchmarking hardnekkige mythes schuilgaan zoals dat benchmarking daadwerkelijk kan en zal leiden tot prestatieverbeteringen bij overheidsorganisaties (Korsten et al., 2019: 34). Volgens hen vertoont benchmarkonderzoek nogal eens tekorten die samenhangen met valkuilen die gelden voor dit soort onderzoek. Zij onderscheiden zestien valkuilen (Korsten et al., 2019: 219) en vullen die aan met nog eens tien methodische valkuilen (Korsten et al., 2019: 230). Benchmarking blijkt volgens hen vaak niet meer dan 'het opmaken van een klassement (lees: vergelijking en rangordening)' (Korsten et al., 2019: 240). Veelal ontbreekt een analyse van het waarom van de verschillen en ook het overtuigend aanwijzen van een goede praktijk en hoe daar te komen.

De vertaling die NPM-discussies en doelen, waarden en instrumenten krijgen is volgens Lin-

demann te zien als een performatief proces, gericht op het realiseren van verandering waarbij de interactie tussen de bij de verandering betrokkenen, de actanten, uiteindelijk in hoge mate de uitkomst van het veranderproces bepaalt (Lindemann, 2014: 26). Op het praktische operationele niveau is NPM volgens Lindemann te zien als een 'boodschappenmand' met diverse tools ontleend aan het bedrijfsleven, waar hervormers uit kunnen kiezen waarbij ze de lokale omstandigheden en belangen mee kunnen wegen in hun keuzes (Lindemann, 2014: 23). Hoewel implementaties van NPM een eigen 'vertaling' zijn van wat onder NPM wordt verstaan, bewegen de discussies erover zich volgens Lindeman rondom drie thema's: competitie, controle en competentie (Lindemann, 2014: 24). De drie thema's zijn nauw met elkaar verbonden en versterken elkaar bij het optimaliseren van prestaties van publieke organisaties waarmee een bedrijfsgerichte bureaucratie wordt vormgegeven.

Prestatiemetingen hebben schaduwzijden die zich slecht verhouden tot de ambities die met prestatiemetingen beoogd worden. Dit maakt dat meten zelden de karakteristieken heeft die het op voorhand worden toebedacht, integendeel. Tonkens stelt: 'Meten is niet zomaar meten: meten is ingrijpen.' (Tonkens, 2008: 91). Dat ingrijpen is niet neutraal in haar optiek, maar heeft effect op de waarden van een samenleving. In de praktijk blijkt het zoeken naar goede 'proxies' voor prestaties, uit te drukken in meetbare eenheden, lastig uitvoerbaar te zijn. Er kleven veel technische en operationele problemen aan het meten wat maakt dat er volgens Tonkens twee kanten op gegaan kan worden: ofwel er wordt gekozen voor betere indicatoren zodat metingen verbeterd kunnen worden, ofwel er wordt gezorgd voor beter gebruik van en meer realistische verwachtingen ten aanzien van de gegeneerde informatie (Tonkens, 2008: 85). Naar de eerste kant, het verbeteren van prestatiemetingen, is het nodige onderzoek gedaan. Tonkens rapporteert over de variatie die onderzoeken laten zien van onbedoelde gevolgen van (de techniek van) het meten in organisaties. Daarbij wijst zij op het belang van de meer cultuurfilosofische vraag naar de betekenis van de meetcultuur voor de samenleving omdat meten organisatieprocessen, normen, identiteiten en samenlevingsvormen beïnvloedt, ook als processen op het oog gewoon doorgang blijven vinden, zoals in het geval van onderwijs (Tonkens, 2008: 91). Vosselman wijst op de mogelijkheden die een beter gebruik en meer realistische verwachtingen ten aanzien van informatie bieden om te komen tot prestatieverbetering. Want, zo stelt hij, getalsmatige (en ook andere) informatie zal wat zich op de werkvloer en bij het management afspeelt hooguit gebrek-kig kunnen representeren, want cijfers zijn op zichzelf het resultaat van een bewerkingsproces en zijn zeker geen opdoemende feiten. Vosselman spreekt van een 'crisis van representatie' waarmee hij bedoelt dat er blijft worden geloofd in de mogelijkheid om de werkelijkheid, de waarheid, in cijfers te doen reflecteren, wat volgens Vosselman nooit zal lukken (Vosselman, 2015: 6).

## 5.2 Basisonderwijs en kwaliteit

In de lange traditie van eerst het lager onderwijs en vervolgens het basisonderwijs is in Nederland het doel van dat onderwijs gericht geweest op het zorgen voor een 'ononderbroken ontwikkelingsproces'. Ook in de huidige Wet op het Primair Onderwijs, artikel 8, is dit terug te vinden. In artikel 8 wordt het uitgangspunt en de doelstelling van het primair onderwijs uitgewerkt als:

- 1) 'het onderwijs wordt zodanig ingericht dat de leerlingen een ononderbroken ontwikkelingsproces kunnen doorlopen. Het wordt afgestemd op de voortgang in de ontwikkeling van de leerlingen;



- 2) het onderwijs richt zich in elk geval op de emotionele en de verstandelijke ontwikkeling, en op het ontwikkelen van creativiteit, op het verwerven van noodzakelijke kennis en van sociale, culturele en lichamelijke vaardigheden;
- 3) het onderwijs gaat er mede van uit dat leerlingen opgroeien in een pluriforme samenleving, is er mede op gericht actief burgerschap en sociale integratie te bevorderen en is er mede op gericht dat leerlingen kennis hebben van en kennismaken met verschillende achtergronden en culturen van leeftijdgenoten; en
- 4) ten aanzien van leerlingen die extra ondersteuning behoeven, is het onderwijs gericht op individuele begeleiding die is afgestemd op de behoeften van de leerling' (WPO, artikel 8).

Het zorgen voor kwaliteit van onderwijs is historisch gezien een taak van de scholen. Onder zorg dragen voor de kwaliteit van het onderwijs wordt volgens de WPO in elk geval verstaan: 'de zorg voor het personeelsbeleid, voor zover het betreft de duurzame borging van de kwaliteit van het onderwijspersoneel, en het uitvoeren van het in het schoolplan, bedoeld in artikel 12, beschreven beleid op een zodanige wijze dat de wettelijke opdrachten voor het onderwijs en de door het bevoegd gezag in het schoolplan opgenomen eigen opdrachten voor het onderwijs worden gerealiseerd' (WPO, artikel 12). In artikel 10a wordt vervolgens uitgewerkt wanneer scholen niet aan de kwaliteitseisen voldoen. Als de leerresultaten voor taal en rekenen op de school aan het eind van het achtste schooljaar op groepsniveau ernstig of langdurig tekortschieten, gemeten over een periode van 3 schooljaren, voldoen scholen niet aan de kwaliteitseisen. Het tekortschieten definieert de wet als dat de leerresultaten 'liggen onder de minimum normering die daarvoor geldt in vergelijking tot de leerresultaten over diezelfde schooljaren van scholen met een vergelijkbaar leerlingenbestand' (WPO, artikel 10a). Prestaties van andere scholen zijn bepalend voor de te hanteren beoordelingsnorm. Het beoordelen van kwaliteit van onderwijs kent daarmee een relatieve norm; de prestatie van een school is te definiëren als een prestatie ontleend aan de scores van de individuele leerlingen die aan het eind van groep 8 een eindtoets maken én van de prestaties van leerlingen van andere scholen. Kwaliteit van primair onderwijs kent in Nederland geen 'normatieve' omschrijving maar wordt ontleend aan relatieve metingen op basis van relatieve prestaties op een beperkt deel van de beoogde doelen van primair onderwijs, namelijk dat deel dat onderdeel uitmaakt van de eindtoets.

Biesta stelt dat de vraag naar wat goed onderwijs is in Nederland niet alleen nauwelijks gesteld wordt, maar dat die vraag is vervangen door andere gesprekken en discussies die niet gaan over de *normatieve* vraag naar wat goed onderwijs is. Die vraag is volgens hem vervangen door technische en organisatorische vragen over de efficiëntie en effectiviteit van processen (Biesta, 2012: 15). Kwaliteitscontrole van onderwijs in Nederland gaat volgens Biesta niet over datgene dat die processen moeten bewerkstelligen (Biesta, 2012: 58). Daarmee is volgens Biesta reenschap over kwaliteit van onderwijs veranderd van een professionele en democratische notie in een fundamenteel bureaucratische notie (Biesta, 2012: 19). Onderwijs is in de optiek van Biesta geherpositioneerd als een publieke dienst, waarin door de overheid wordt voorzien en die wordt betaald met belastinggeld: 'In deze constructie is er niet langer een directe verantwoordingsrelatie tussen ouders (of leerlingen) en de scholen. De verantwoordingsrelatie is indirect. Directe verantwoording is er tussen de school en de overheid. Burgers worden erin gepositioneerd als consumenten die kunnen 'stemmen' over de kwaliteit van de diensten die de overheid levert, maar ze hebben geen democratische zeggenschap over de richting of inhoud van hetgeen gele-

verd wordt.' (Biesta, 2012: 59). Scholen verantwoorden zich aan een overheid die meer en meer sturend is voor de onderwijspraktijk. Die relatie is inherent aan de gekozen verantwoordingsrelatie maar daarmee is de overheid voor de scholen steeds meer een inhoudelijke bepaler van beleid geworden en raken leerlingen en hun ouders meer uit beeld.

Vanuit wat gezien wordt als belangrijk voor het realiseren van kwaliteit stuurt de overheid op vorm en inhoud van het onderwijs. Uit de aanbevelingen van het parlementair onderzoek Onderwijsvernieuwingen (2008) blijkt dat de Commissie Dijsselbloem pleit voor een stimulerende rol van de overheid bij het tot stand brengen van 'evidence based' of 'practice-based' vernieuwingen in de scholen (Dijsselbloem, 2008: 154). De nadruk die onder andere naar aanleiding van deze aanbeveling is ontstaan op het gebruik van feitelijke informatie voor het nemen van beslissingen over de richting van onderwijsbeleid en de vorm van de onderwijspraktijk, is een manier van redeneren waarmee volgens Biesta twee fundamentele problemen samenhangen: het eerste probleem is dat uit wat de situatie is nooit logisch afgeleid kan worden wat er gedaan zou moeten worden. En het tweede probleem waar Biesta op wijst, betreft de validiteit van metingen waarmee hij niet de technische validiteit bedoelt maar de normatieve validiteit. De normatieve validiteit heeft volgens Biesta te maken met de beantwoording van de vraag: 'Meten we daadwerkelijk wat we waardevol vinden of meten we wat we makkelijk kunnen meten waardoor we waarderen wat we kunnen meten?' (Biesta, 2012: 25). Er is volgens Biesta sprake van een performativiteitscultuur die opkomt in het onderwijs waarmee hij bedoelt dat kwaliteitsindicatoren worden verward met kwaliteit als zodanig en het middel een doel op zichzelf wordt (Biesta, 2012: 25). Het 'afrekenen' (accountability) op leerresultaten ziet Biesta als een heel beperkte vorm van rekenschap afleggen: 'Verantwoording dragen voor kind en wereld houdt meer in' (Biesta, 2012: 10). Biesta stelt vast dat de huidige cultuur van verantwoording zeer problematisch is. De manier waarop nu in het Nederlandse onderwijs het afleggen van verantwoording wordt vormgegeven, is volgens Biesta 'een apolitek en anti-democratische strategie die alle bestaande relaties herdefinieert in economische termen en ze daarom beschouwt als formele, in plaats van substantiële verhoudingen (...) waardoor verantwoording een doel op zich wordt, in plaats van een middel om andere doelen te bereiken.' (Biesta, 2012: 62). Hij concludeert dat geen oordeel geveld kan worden over de kwaliteit van onderwijs zolang niet helder is wat precies van onderwijs verwacht wordt (Biesta, 2012: 76).

Het denken over kwaliteit van onderwijs is in Nederland traditioneel overgelaten aan de scholen om inhoud en invulling aan te geven en daarover verantwoording af te leggen, vooral naar ouders en leerlingen. Bij het terugtreden van de overheid, ingezet vanaf de jaren tachtig, is de overheid gaan zoeken naar indicatoren voor kwaliteit op basis waarvan aansturing en verantwoording kan plaatsvinden. Dit is ook voor onderwijs gebeurd. Frissen et al. (2015) hebben aan de hand van een media-analyse de thema's in het debat over onderwijskwaliteit van 2002 tot 2014 nader onderzocht. De betekenis van onderwijskwaliteit is een terugkerend discussiepunt in de media. In de berichtgeving zijn volgens de analyse drie fasen te herkennen waarin de toonzetting verandert. In de eerste fase (2002-2004) is er volgens Frissen et al. vooral aandacht voor de bijdrage die het onderwijs moet leveren aan burgerschapsvorming en het bevorderen van integratie. Onderwijskwaliteit wordt hier sterk als brede maatschappelijke vorming besproken. In de tweede fase (2005-2010) wordt de nadruk juist gelegd op 'basisvaardigheden' (taal en rekenen) die veel minder maatschappelijk en juist heel cognitief zijn. Vanaf 2011 verschiet volgens de analyse het

debat opnieuw van kleur, waarbij de essentie van de discussie vooral is dat onderwijskwaliteit méér behelst dan taal en rekenen. Wat dat 'meer' is kristalliseert zich niet helder uit. Het belang van de basisvaardigheden wordt daarbij niet ontkend, maar onderwijskwaliteit moet hier niet toe worden beperkt. De nadruk in het maatschappelijke debat over onderwijskwaliteit ligt volgens Frissen et al. op de waarden en regels waarin de overheid een initiërende rol speelt en het overheidsbeleid het belangrijkste onderwerp van gesprek is (Frissen, 2015: 32).

### 5.2.1 Onderwijsbeleid en kwaliteit

De beleidsprogramma's deregulering en autonomievergroting krijgen in 1985 in de vorm van de nota HOAK, 'Hoger onderwijs: Autonomie en Kwaliteit, een andere besturingswijze', voor het eerst concreet gestalte. Volgens Van Wieringen heeft deze nota 'als weinig andere nota's het denken en handelen in het onderwijsbeleid bepaald' (Van Wieringen, 1996: 155). Met deze nota en 'haar derivaten' is er een 'ontkoppeling gekomen tussen het pedagogische en het administratieve' wat volgens hem 'niet altijd even gelukkig is geweest' (Van Wieringen, 1996: 156). Verbonden aan autonomievergroting en deregulering wordt gezocht naar nieuwe vormen van verantwoording om grip op kwaliteit van de kant van de overheid te kunnen behouden.

In 1991 brengt het Instituut voor Toegepaste Sociale Wetenschappen (ITS) in opdracht van de WRR een adviesrapport uit waarin de resultaten van een haalbaarheidsstudie naar de invoering van vergelijkende kwaliteitsbeoordeling van scholen in het basisonderwijs worden gepresenteerd. Het rapport presenteert drie scenario's. Bij twee ligt het initiatief voor kwaliteitsverantwoording bij de school zelf en bij het derde scenario wordt geopteerd voor een door de overheid 'gemandateerde schoolkwaliteitsstoets' (Pouwels & Jungbluth, 1991: 133). Het doel van een dergelijke toets is volgens het rapport te 'controleren of wordt voldaan aan van tevoren door de wetgever en/of de overheid vast gestelde kerndoelen of minimumstandaarden' (Pouwels&Jungbluth, 1991: 133). Pouwels en Jungbluth achten het 'terugvallen op toetsing van leerlingprestaties [...] van strategische betekenis voor de bescherming van de onderwijsconsumenten (i.c. vooral de ouders)' (Pouwels&Jungbluth, 1991: 133). Een systeem voor schoolkwaliteitstoetsing 'kan en moet bestaan uit prestatiemetingen bij de leerlingen' (Pouwels&Jungbluth, 1991: 138). Zij wijzen op ervaringen die met dergelijke toetsen op dat moment worden opgedaan in de Verenigde Staten waar onderlinge vergelijkingen tussen regio's en scholen gemaakt kunnen worden 'om de prestaties te verhogen' (Pouwels&Jungbluth, 1991: 134). Voor de ontwikkeling van een test zoals zij die voor zich zien in hun derde scenario stellen zij dat het noodzakelijk is dat er landelijke standaarden worden ontwikkeld op basis waarvan vergelijkingen mogelijk worden (Pouwels&Jungbluth, 1991: 135).

Het eerste paarse kabinet (1994) introduceert in de nota 'De school als lerende organisatie' van onderwijsstaatssecretaris Netelenbos (1995) een nieuwe benadering van verantwoording over kwaliteit. Het idee van een lerende organisatie is ontwikkeld in het bedrijfsleven maar trekt als concept ook de belangstelling van de overheid in relatie tot onderwijs. Tjepkema stelt dat dit begrijpelijk is omdat ook scholen merken dat er een steeds groter beroep op hun leervermogen wordt gedaan; zij moeten steeds creatiever en sneller inspelen op veranderingen (Tjepkema, 2002: 2). In de nota is nader uitgewerkt hoe verantwoording over kwaliteit er volgens de overheid meer concreet uit moet komen te zien. De overheid, zo stelt de nota, tracht op diverse wijzen de kwaliteit van het onderwijs te waarborgen, te bewaken en te stimuleren. Het doel is

hierbij tweeledig: enerzijds worden scholen aangespoord de informatie te gebruiken voor de verbetering van het onderwijs aan individuele leerlingen maar aan de andere kant zegt de nota de informatie te kunnen en te willen benutten voor de vergelijking van scholen met elkaar (OCW, 1995: 13). Dit kan volgens de nota niet langer alleen met de eigen invulling van evaluaties door de scholen zelf, maar hiervoor zijn gestandaardiseerde meetresultaten nodig. De nota acht daarom ook de categorie van sturingsinstrumenten voor externe kwaliteitsbewaking van belang die worden gevormd door de landelijke peilingsinstrumenten en internationale vergelijkingen. Naast dit soort externe toetsgegevens en door de school zelf gegenereerde informatie ziet de nota als belangrijke externe toets op de kwaliteit een rol voor de inspectie weggelegd. Volgens Mertens, voormalig inspecteur generaal van het onderwijs, 'nam de nota, als breuk met het verleden, de afzonderlijke school als uitgangspunt en introduceerde een verantwoordingsbenadering (accountability) in samenhang met grotere bestuurlijke zelfstandigheid' (Mertens, 2014: 18). De school krijgt meer bevoegdheden om zelf integraal beleid te voeren en moet over dat beleid jaarlijks verantwoording afleggen in de vorm van de schoolgids. De inspectie zou toezicht houden aan de hand van die verantwoording. Een benadering, volgens Mertens, waarbij de toezichthouder zich 'proportioneel tot de school verhoudt' (Mertens, 2014: 18); het toezicht is afhankelijk van de opbrengsten en de aangetroffen kwaliteit van de kwaliteitszorg.

In 2004 verschijnt de nota 'Koers Primair Onderwijs, Ruimte voor de school'. Deze nota, gebaseerd op tal van gespreksrondes die het ministerie heeft gehouden met betrokkenen uit het onderwijsveld, wil streven naar een nieuw evenwicht tussen ruimte en rekenschap. Dit komt, volgens de nota, nadat scholen dan al jaren 'last' hebben van een overheid die zich bijna tot in detail bemoeit met wat goed is voor de leerling, de school en het stelsel. Die bemoeienis gaat tot op het didactisch niveau, gegeven de discussie over de invoering van het 'probleemgestuurd onderwijs' of het 'natuurlijk leren'. De nota stelt: 'De visie op de toekomst van de sector sluit aan bij de inzet van het kabinet om verantwoordelijkheden te beleggen waar zij thuishoren en de regel-druk te verminderen. Net als in andere sectoren van de samenleving, zoals de gezondheidszorg en de politie, lijkt de regellast in het onderwijs de afgelopen jaren alleen maar te zijn toegenomen – met als gevolg dat de professionals in het veld minder ruimte hebben voor het uitoefenen van hun eigenlijke taken' (OCW, 2004: 7). Met de nota zegt het ministerie enerzijds ruimte te willen bieden aan de vrijheid van onderwijs. Want, zo stelt de nota, er is behoefte aan een omkering van verhoudingen: de leraar of het schoolteam hoort de architect van het onderwijsprogramma te zijn. Als zij de ruimte krijgen, zijn ze in staat om onderwijs op maat te bieden. Scholen worden anderzijds geacht rekenschap af te leggen door duidelijk te communiceren over hun zelf gestelde doelen en daarover ook verantwoording af te leggen. De rekenschapscomponent krijgt in de nota concreet karakter door de verplichting op te leggen aan scholen om een jaarverslag te maken, kengetallen aan te leveren in het kader van mogelijke benchmarks en bij te dragen aan stelselindicatoren wat betreft toegankelijkheid, kwaliteit, rechtmatigheid en doelmatigheid. De autonomie van scholen brengt in eerste instantie met zich mee dat scholen zelf verantwoordelijk zijn voor de uitvoering van het onderwijs waar het al dan niet toetsen gezien wordt als een integraal onderdeel van de lespraktijk en daarmee van de vrijheid van onderwijs.

De criticasters van de ruimte en autonomievergroting krijgen de wind in de zeilen mede naar aanleiding van signalen dat het niet goed zou gaan met het onderwijs. Die signalen zijn voor de Tweede Kamer aanleiding om uiteindelijk te besluiten tot het instellen van een parlementair

onderzoek naar de vernieuwingen in het onderwijs. Het 'Parlementair onderzoek Onderwijs-vernieuwingen' uitgevoerd door de Commissie Dijsselbloem start in 2007 en komt in 2008 met bevindingen. De Commissie tikt de overheid nadrukkelijk op de vingers daar waar de overheid naar het oordeel van de Commissie te veel is getreden in waar het onderwijs over moet gaan. De Commissie stelt dat de Inspectie van het Onderwijs 'zich ontegenzeggelijk in de beleidsvrijheid van de scholen begaf' (Dijsselbloem, 2008: 137). De Commissie verklaart dat door te stellen dat 'dit mede een gevolg was van het feit dat de politiek naliet helder de onderwijsdoelen te definiëren. Wanneer het onmogelijk wordt op concrete uitkomsten toe te zien, gaat men het toezicht richten op het proces, op datgene wat zich in de klassen afspeelt' (Dijsselbloem, 2008: 137). De Commissie komt tot de conclusie dat het maatschappelijk vertrouwen in het onderwijs onder druk is komen te staan en komt met de nodige aanbevelingen voor de rijksoverheid, die verweten wordt onvoldoende zicht te hebben gehouden op de resultaten van het onderwijs. Leerstandaarden ontbreken naar het oordeel van de Commissie evenals adequate en effectieve instrumenten. De Commissie stelt vast dat er op dat moment geen effectief instrument is om de kwaliteit van het onderwijs te bewaken waarmee een objectief beeld over het niveau van leerlingen aan het einde van de basisschool kan worden verkregen in relatie tot bestaande leerdoelen. In het primair onderwijs wordt tot dan toe gewerkt met zogenaamde 'streefdoelen' die zijn wat de naam zegt, namelijk doelen om naar te streven, maar zijn niet te zien als 'minimum standaarden' of 'minimale opbrengsten'. De eindtoets wordt weliswaar door het overgrote deel van de scholen gebruikt, maar is door het relatieve karakter van de meting niet geschikt om een oordeel te geven over het al dan niet behalen van leerstandaarden. Dat maakt dat de Commissie pleit voor de invoering van leerstandaarden zodat voor iedereen helder is wat de leerlingen aan het einde van hun schoolperiode moeten kennen en kunnen. Toetsen, gekoppeld aan die leerstandaarden zijn daarbij, naar het oordeel van de Commissie, onmisbaar. Ondanks het op dat moment ontbreken van duidelijke leerstandaarden en het ontbreken van effectieve instrumenten stelt de Commissie wél dat er wat betreft de basisvaardigheden lezen en rekenen/wiskunde 'reële zorgen' bestaan, omdat een 'zorgwekkende dalende trend zichtbaar is' (Dijsselbloem, 2008: 189). De Commissie pleit heel concreet voor een eindtoets, aan het eind van groep 8, verplicht voor alle leerlingen en gekoppeld aan leerstandaarden. De Commissie wil niet alleen een toets om leerstandaarden te meten maar pleit tevens voor de afname van een begin- en een eindtoets zodat de toegevoegde waarde van basisscholen kan worden beoordeeld. Er zijn, samengevat, naar het oordeel van de Commissie, metingen nodig op het niveau van de leerling, de school en het stelsel. De Commissie pleit voor een goede verdeling van verantwoordelijkheden tussen overheid enerzijds en scholen anderzijds en vat dat handelen samen met de woorden 'de overheid gaat over het 'wat' en de scholen over het 'hoe'. Als motto heeft de uitspraak veel gevolgen, maar minder in de toepassing. Ten eerste is daar de vraag hoe het uitgangspunt dat de overheid gaat over het 'wat' van onderwijs zich verhoudt tot artikel 23 van de Grondwet ten aanzien van onderwijs en ten tweede blijkt het onderscheid tussen het wat en het hoe in de praktijk tot veel onduidelijkheid te leiden en minder goed te onderscheiden te zijn dan het op het eerste gezicht lijkt.

Ten aanzien van toetsing in het primair onderwijs, kan het rapport worden gezien als een formalisering van een kentering die dan al langer gaande was. Met een belangrijke fundamentele wijziging. Tot dan toe is toetsing aan het eind van het basisonderwijs vooral gepositioneerd als een instrument in handen van de onderwijsprofessional die tests en uitkomsten naar eigen inzicht gebruikt bij het opstellen van een advies voor (vervolg)onderwijs. De invoering van een ver-

plichte eindtoets, een verplicht gebruik van een Leerlingvolgsysteem (LOVS) en de discussie over de ontwikkeling van een test aan het begin van het onderwijs, de 'kleutertoets', waarmee dan leerwinst zou moeten worden vastgesteld (toegevoegde waarde) leggen tests en uitkomsten in de handen van de overheid. Daarmee krijgt de overheid ook een belangrijke sleutel ten aanzien van het 'hoe' van het onderwijs in handen omdat vanuit terugslageffecten van toetsen op het onderwijs bekend is dat toetsing sturend is voor het onderwijs: hoe belangrijker de uitkomsten van een test zijn, hoe meer scholen zich zullen richten naar het willen behalen van goede resultaten.

De Commissie realiseert zich dat onderwijsvernieuwingen zich in de toekomst weer zullen voordoen. Dat is de reden dat aan het rapport een zogenaamd 'Toetsingskader' is toegevoegd dat beleidsvoorwaarden bevat die volgens de Commissie een zorgvuldig beleidsproces borgen. De Commissie stelt dat 'afwijken van dit toetsingskader alleen kan wanneer dit nadrukkelijk en adequaat wordt gemotiveerd richting parlement' (Dijsselbloem, 2008: 152).

#### **Voorwaarden voor een zorgvuldig beleidsproces:**

- De probleemanalyse is helder, wetenschappelijk onderbouwd en wordt breed gedragen door betrokkenen.
- Er is overtuigend aangetoond dat overheidsinterventie noodzakelijk is. – Er is een evaluatie beschikbaar van voorafgaand beleid. – Er is verantwoord welke beleidsalternatieven zijn overwogen en gekozen. – Neveneffecten en samenhang met overig beleid zijn in beeld gebracht en betrokken bij de verdere uitwerking van het beleid.
- De gekozen beleidsopatie is wetenschappelijk gevalideerd. Zo niet, dan wordt de beleidsvernieuwing onder wetenschappelijke begeleiding eerst kleinschalig in pilots (met controlegroep) uitgetoetst.
- De resultaten van deze pilots zijn adequaat geëvalueerd en zichtbaar verwerkt in het beleid.
- Aan de voorwaarden voor een goede implementatie, waaronder geld, expertise en tijd is voldaan.
- De uitvoeringsorganisatie is helder gepositioneerd. – Diegenen die geacht worden de vernieuwing in de praktijk uit te voeren zijn actief betrokken geweest bij de totstandkoming van de vernieuwing en hebben zich een helder beeld kunnen vormen over de consequenties voor hun eigen werk.
- Er is voldoende draagvlak onder alle betrokkenen maar in ieder geval onder de professionals die de vernieuwing in de praktijk moeten brengen.
- Er is voorzien in (tussentijdse) evaluaties. Er zal niet worden overgegaan tot overhaaste bijstellingen dan nadat eerst nut en noodzaak zijn onderzocht.

*Uit: Parlementair Onderzoek Onderwijsvernieuwingen, 2008*

Bronneman-Helmers stelt dat de commissie vooral een politieke functie had. De commissie wilde, volgens Bronneman-Helmers, na een periode van heftige discussie over de rol die de politiek bij de vernieuwingen had gespeeld, schoon schip maken, om daarna een nieuwe start te kunnen maken. Van een goede analyse van het eigen onderwijsbeleid was volgens haar geen sprake (Bronneman-Helmers, 2011: 121/122). Het zorgvuldige beleidsproces zoals uitgewerkt in het 'Toetsingskader' is bij de invoering van de verplichte afname van eindtoets en leerlingvolgsysteem niet herleidbaar gevolgd bij de verantwoording van de wetswijziging.

De verschuiving van toezicht op en verantwoording over kwaliteit van proces naar resultaat vergroot de nadruk op het belang van meetbare, objectieve opbrengsten. In ruil voor het gevraag-

de loslaten van de kant van de overheid ontstaat meer druk op verantwoording en aantonen dat geleverde diensten de financiering waard zijn gebleken. Voor onderwijs worden individuele prestaties (prestaties op het microniveau van het onderwijs) als uitgangspunt genomen voor het beoordelen van prestaties op het microniveau, namelijk op het niveau van de docent, maar ook op meso- en macroniveau. Zowel de scholen, de schoolbesturen als het stelsel als geheel worden wat betreft prestaties afhankelijk van de prestaties van hun leerlingen. De Vlaamse pedagoog Standaert wijst erop dat dit gevolgen heeft voor de pedagogische relatie die er is tussen de leerlingen en hun leerkracht(en) en school: leerlingen moeten zorgen dat de scholen voldoen aan de eisen, terwijl het andersom zou moeten gelden. Scholen moeten leerlingen als het centrum van hun handelen zien (Standaert, 2014: 37). In de tekst bij het toezichtkader voor het primair onderwijs van de Inspectie van het Onderwijs uit 2005 wordt de verschuiving zichtbaar: 'Scholen worden in toenemende mate aangesproken op de behaalde onderwijsresultaten in plaats van op de verrichte inspanningen.' Dit vraagt van de inspectie een andere manier van kijken om het vergelijken en het waarderen van verschillen mogelijk te maken: 'Het vigerende waarderingskader is te veel gelaagd en discrimineert te weinig tussen goede en zwakke scholen.' In 2005 is er vooral sprake van stimulering tot het behalen van resultaten: 'De inspectie beoordeelt of de school onderwijs van voldoende kwaliteit verzorgt (waarborg) en de best mogelijke resultaten wil stimuleren te behalen met hun leerlingen (stimulering)' (Inspectie van het Onderwijs, 2005). Vijf jaar later, in 2010, zijn toetsen naar het oordeel van de inspectie een onmisbaar onderdeel geworden in het kader van de bewaking van kwaliteit: 'Toetsen moeten niet alleen meten of leerlingen een bepaald niveau halen, maar ook hoever ze boven of onder dat niveau zitten. Zulke gegevens zijn onmisbaar voor verbeteringen op leerling-, klas-, school- en stelselniveau.' (Inspectie van het Onderwijs, 2010).

In het najaar van 2010 treedt het eerste kabinet Rutte aan. Het regeerakkoord van VVD en CDA krijgt de naam 'Vrijheid en Verantwoordelijkheid'<sup>21</sup>. Het kabinet stelt daarin dat 'een krachtige, kleine en dienstverlenende overheid zal worden gevormd met minder belastinggeld, minder ambtenaren, minder regels en minder bestuurders'. Voor wat betreft het onderwijs wil het kabinet de kwaliteit van het onderwijs versterken en hogere prestaties bevorderen. Meer concreet betekent dat het kabinet een predicaat 'excellent' wil invoeren dat scholen kunnen verdienen in een wedstrijd met elkaar. De kwaliteit van het onderwijs versterken vertaalt het kabinet in een kerntaak voor taal en rekenen en in 'de toegevoegde waarde (leerwinst) zwaarder laten wegen' bij het beoordelen van scholen. Uit de gekozen aanpak blijkt welke ruimte de overheid voor ogen heeft als het gaat om sturen op kwaliteit: dit kan alleen indirect. Bronneman-Helmers vat de kern van het beleid zo samen dat er 'minder werd gestuurd op inputkenmerken van scholen en meer op basis van geleverde prestaties' (Bronneman-Helmers, 2011: 100). Peeters, Hofman en Frissen spreken van een 'U-bocht constructie': omdat de overheid niet rechtstreeks mag sturen op de relatie tussen leraar en leerling organiseert de overheid gedragsprikkel richting scholen, ouders, leraren en schoolbesturen. De gedragsprikkel hebben twee vormen volgens Peeters, Hofman en Frissen, namelijk 'normering aan de achterkant en 'capacity-building' aan de voorkant'. De normering aan de achterkant leidt tot het sturen op 'eenduidigheid in onderwijsnormen en transparantie in onderwijsopbrengsten' (Peeters et al., 2013: 13). Zo kan een situatie ontstaan waarin de 'overheid zich niet rechtstreeks bemoeit met het onderwijs in de klas, maar een samenhan-

21] <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2010/09/30/regeerakkoord-vvd-cda/regeerakkoord-vvd-cda.pdf>

gend stelsel van normering, toezicht en sanctionering organiseert om de onderwijsopbrengsten te meten, en een pakket aan ondersteuningsmaatregelen om scholen en leerkrachten te helpen aan deze normen te voldoen' (Peeters et al., 2013: 16).

In november 2010 neemt de Kamer een motie aan van PVV-kamerlid Beertema en VVD-kamerlid Elias waarin wordt geformuleerd dat een 'landelijk stelsel van eindtoetsen in het primair onderwijs wordt ingevoerd, vastgesteld en gecontroleerd door onafhankelijke instituties'. Dit wordt onderbouwd door te stellen dat de kwaliteit van diploma's boven elke twijfel verheven moet zijn. Daarover zijn op dat moment twijfels omdat, zo stelt de motie, 'onderwijsinstellingen nu regelmatig eigen kwaliteitsnormen vaststellen, die willekeurig aanpassen en in eigen beheer toetsen'. In lijn met het regeerakkoord, de motie Beertema/Elias en de brief die de minister eerder dat jaar schrijft aan de Tweede kamer (31 293, nr. 89) komt minister van Bijsterveldt in 2011 met haar actieplan 'Basis voor Presteren, naar een ambitieuze leercultuur voor alle leerlingen'. 'Zonder ambities geen betere prestaties' vormt de belangrijkste inhoudelijke onderbouwing van het actieplan waarin een grondige analyse van de stand van het onderwijs en aantoonbare effectiviteit van maatregelen ontbreken. Om de prestaties te verhogen is volgens de minister een opbrengstgerichte manier van werken nodig. Meer scholen moeten hiermee aan de slag want, zo luidt de redenering, dan gaan de prestaties op het gebied van taal en rekenen omhoog. De nadruk op taal en rekenen wordt in het actieplan onderbouwd door te stellen dat 'een goede beheersing van de basisvaardigheden voor een belangrijk deel het succes in de schoolloopbaan en daarmee het toekomstig opleidingsniveau bepaalt' (OCW, 2011: 2).

### 5.2.2 Resultaten volgen

Met het wetsvoorstel van 2011 komt er een eind aan de vrijblijvendheid van het stimuleren van onderwijsopbrengsten. Met de wetwijziging stelt de minister een eindtoets verplicht alsmede een verplichting voor het gebruik van een extern goedgekeurd en gevalideerd Leerlingvolgsysteem (LVS). Het gebruik van een LVS wordt gezien als een belangrijk middel om inzicht te krijgen in de leervorderingen en leerresultaten van leerlingen vanaf groep 3 tot aan het eind van groep 8. De toelichting bij het wetsvoorstel stelt dat het verplichte gebruik van een LVS past in het beleid gericht op meer 'opbrengstgericht werken'. Het past ook bij de groeiende aandacht die wordt gevraagd voor het onderkennen en bieden van kansen aan excellentie. De 'opgelegde plicht' tot differentiëren waarop de Inspectie van het Onderwijs in het kader van opbrengstgericht werken is gaan toezien heeft ertoe bijgedragen dat op jonge leeftijd behaalde scores en daaraan gekoppelde oordelen nu een belangrijke rol spelen in het onderwijs. Ze bepalen in hoge mate het onderwijs dat een leerling krijgt. De gevolgen van de verplichting tot het diagnosticeren van achterstanden, om die met een gedifferentieerd onderwijsaanbod te kunnen aanpakken, worden nu in de praktijk zichtbaar. In zijn oratie (2017) gaat Denessen in op onderzoek dat hij gedaan heeft naar differentiatie in het onderwijs in relatie tot kansengelijkheid. Hij spreekt over een 'differentiatiedilemma' waar een leerkracht in de praktijk van het onderwijs voor komt te staan. Want de twee beleidsinitiatieven van de overheid, namelijk de aanpak van toptalenten en het bevorderen van gelijke kansen zijn volgens Denessen strijdig met elkaar. Want hoe verdeelt een leerkracht zijn aandacht op een rechtvaardige manier? Is dat door alle kinderen evenveel aandacht, belangstelling, affectie en waardering te geven (divergente differentiatie) of is het rechtvaardiger leerlingen onevenredig veel tijd en aandacht te geven die bij het leren meer ondersteuning nodig hebben (convergente differentiatie)? Denessen wijst erop dat het huidige beleid de leerkracht



zelf laat bepalen welk rechtvaardigheidsprincipe hij aanhangt en verkiest voor zijn handelen. Een bijkomend probleem is dat als leerkrachten lage prestaties van leerlingen toeschrijven aan een gebrek aan aanleg of inzet zij minder geneigd zijn om in deze leerling te investeren dan leerkrachten die lage prestaties toeschrijven aan een gebrek aan ondersteuning van de ouders of aan onrechtvaardigheid van het onderwijssysteem (Denessen, 2017: 7). Ook hier is het aan de leerkracht zelf om een keuze te maken afhankelijk van zijn mensbeeld. Differentiëren is daarmee zeker geen neutrale ingreep en niet zonder meer na te streven zoals de Inspectie van het Onderwijs beoogt en betoogt in het kader van het verbeteren van de kwaliteit van onderwijs. Denessen stelt dat als het streven is burgerschapsdoelen, sociale cohesie en integratie te realiseren het wenselijk is om juist *niet* te differentiëren (Denessen, 2017: 10).

Onderzoek heeft sinds lange tijd het belang van het verwachtingseffect (of het pygmalioneffect) voor het gedrag van kinderen aangetoond: zij internaliseren de definities die volwassenen hun geven en willen die definities waarmaken (Rosenthal&Jacobson, 1968). Daar komt bij dat de manier waarop je het leren tegemoet treedt volgens Dweck in hoge mate bepalend is voor het succes in leren. Een statische mindset maakt die kans op succes kleiner en een negatief zelfbeeld bevordert een dergelijke mindset (Dweck, 2011). De Franse pedagoog Meirieu wijst op het gevaar te streven naar het systematisch identificeren van groepen kinderen die door hun afwijkende gedrag of zelfs delinquentie een potentieel gevaar zouden vormen. Het zou gevaarlijk zijn hen van tevoren als zodanig te beschouwen en te onderwerpen aan een door de institutie voorschreven sociale behandeling. Het gaat dan om opsluiting in een gegeven. Meirieu stelt dat het daarom van groot belang is altijd vraagtekens te zetten bij de legitimiteit van classificaties die in onderwijs worden gebruikt (Meirieu, 2016: 104). Het werken met tests waarmee leerlingen worden geordend ziet hij als een vorm van classificeren; een leerling wordt op een schaal gesitueerd waarmee een hiërarchie wordt opgesteld. Dit betekent, in de optiek van Meirieu, dat 'we een zware verantwoordelijkheid dragen voor de mogelijk zeer forse gevolgen, zowel voor de leerlingen onder aan de ladder als voor de leerlingen aan de top' (Meirieu, 2016: 105). Het ordenen kan leiden tot ofwel het besluit dat een leerling erbij hoort ofwel het besluit dat een leerling er niet (meer) bij hoort. Daarbij raakt differentiëren aan de problemen die samenhangen met het vraagstuk van 'inclusie of exclusie'. Inclusie kan volgens Biesta 'met recht een van dé kernwaarden – zo niet de kernwaarde – van de democratie worden genoemd' (Biesta, 2012: 106). En juist die kernwaarde wordt met onderwijs niet nageleefd. De Vlaamse psycholoog Verhaeghe (2019) stelt dat wat er in het onderwijs gebeurt verder gaat dan 'uitsluiting als maatregel voor disciplineren' in de betekenis die Foucault daaraan toekent. In de context waarin 'excelleren' het nieuwe normaal lijkt te zijn geworden, schemert volgens Verhaeghe op de achtergrond het idee van maakbaarheid door, wat volgens hem 'een stuk verder gaat dan disciplineren'. Met maakbaar geluk wordt mislukking een persoonlijk falen, volgens Verhaeghe: 'Je bent al goed, maar niet goed genoeg; jezelf nog beter maken is een morele plicht. Excelleren is mogelijk, mits er voldoende inspanning wordt geleverd.' (Verhaeghe, 2019: 74).

De socioloog Schuyt wijst in deze context op het belang van de verhouding tussen differentiatie en classificatie: wanneer slaat een op goede gronden te maken differentiatie – bijvoorbeeld naar voorkeur voor een bepaald beroep – om in een op vele andere manieren terugkerende classificatie, een ordeningsinstrument, waarmee 'rechten' en 'verdere kansen' worden uitgedeeld of definitief worden ontzegd? (Schuyt, 1991: 335). Formele classificatie leidt volgens Schuyt alleen

onder bepaalde omstandigheden tot grotere gelijkheid. Generaliserend stelt Schuyt: 'Slechts onder omstandigheden waarin de toegang tot de formele classificatie min of meer voor alle betrokkenen gelijkelijk aanwezig is, valt het effect te verwachten dat formele classificaties rechtvaardig zullen functioneren.' (Schuyt, 1991: 336).

Tests die nadrukkelijk als doel hebben leerlingen van elkaar te onderscheiden, te ordenen en op basis daarvan classificaties te maken, lopen het gevaar bij te dragen aan het vergroten van ongelijkheid. Met de verplichting tot het inzetten van differentiërende instrumenten zoals LVS en eindtoets heeft de overheid een streven naar kansgelijkheid en inclusie in gevaar gebracht. Tegelijkertijd wordt van de kant van de overheid vooral benadrukt dat testen, en dan vooral het monitoren van uitkomsten, juist bedoeld is om kwaliteit van onderwijs te stimuleren, zodat alle leerlingen optimaal van onderwijs kunnen profiteren.

### 5.2.3 Rol van de Onderwijsraad

Het denken over kwaliteit, verantwoordelijkheid en verantwoording heeft de afgelopen decennia de nodige ontwikkeling laten zien. De Onderwijsraad, een van de belangrijkste adviseurs van de minister van onderwijs heeft diverse rapporten uitgebracht over beleidsvoorstellen ten aanzien hiervan.

In 2004 komt de Onderwijsraad met het rapport 'Werken aan gezamenlijke doelen' waarin begrenzing van de autonomie van de scholen wordt bepleit. Ruimte voor autonomie wil de Raad beperken door te komen tot invoering van leerstandaarden en het ontwikkelen van doorlopende leerlijnen van po naar vo om daarmee beoogde opbrengsten voor het onderwijs vast te leggen. De Onderwijsraad wil een resultaatverplichting gekoppeld aan het behalen van een minimum leerstandaard. Ook stelt de Raad, in lijn met de Lissabon-doelstellingen, streefcijfers voor ten aanzien van het aantal gediplomeerden voor havo/vwo.

De Onderwijsraad komt in 2006 met het rapport dat de titel 'Doortastend onderwijstoezicht' meekrijgt. De Onderwijsraad wil grenzen aan een toezicht gebaseerd op vertrouwen. Terwijl het ministerie nog spreekt over ruimte voor vertrouwen, stelt de Raad dat er sprake is van 'onzekere tijden die vragen om bakens'. Die bakens ziet de Raad in de vorm van heldere resultaatnormen die als minimumstandaarden moeten gelden voor taal en rekenen en in de vorm van een intensivering van het toezicht op waarborgen zoals examens. De Raad wijst erop dat doortastend toezicht nodig is dat grenzen bewaakt van het algemeen belang, gebaseerd op normen die de democratische wetgever vaststelt.

In 2006 verschijnt nog een rapport van de Onderwijsraad: 'Versteving van kennis in het onderwijs'. Dit rapport gaat in op signalen die er in de optiek van de Raad zijn over niveauverlies en achteruitgang van verworven kennis. Die signalen lijken voor wat betreft de achteruitgang van het niveau van Nederlands en wiskunde ondersteund te worden door peilingen, nationale en internationale. Het brengt de Onderwijsraad ertoe te wijzen op het gevaar van het hebben van een onvoldoende systematiek in de manier waarop onderwijsinhouden worden vastgesteld en onderhouden. De Raad pleit voor het verhogen van de onderwijsnorm want, zo stelt de Raad 'er is geen reden tot tevredenheid als met extra inspanning zo veel meer kan worden bereikt'. De Raad pleit voor landelijke, structurele peilingen van het onderwijsniveau voor Nederlands en

wiskunde en vat de adviezen van de voorgaande periode kort en krachtig samen door te stellen: 'Bij vrijheid in het 'hoe' hoort duidelijkheid over het 'wat'. (Onderwijsraad, 2006: 10). Hiermee loopt de Raad al vooruit op de terminologie die naar aanleiding van de bevindingen van de parlementaire enquête, de Commissie Dijsselbloem, gemeengoed zal worden.

De Onderwijsraad komt in 2011 met een beleidsreactie op de plannen van de minister om over te gaan tot invoering van een centrale eindtoets in combinatie met een verplicht gebruik van een Leerlingvolgsysteem (LVS) met de titel 'Toetsing in het primair onderwijs'. Daaruit blijkt dat de Onderwijsraad opbrengstgericht werken ziet als middel voor het verbeteren van onderwijsprestaties (Onderwijsraad, 2011: 13). Visscher en Ehren definiëren opbrengstgericht werken als 'het op basis van een goed beeld van beginsituatie en beoogde situatie selecteren en hanteren van de meest geschikt geachte instructiestrategie (in de brede betekenis, dus zowel leerstof, didactiek als klasmanagement)' (Visscher&Ehren, 2011: 1). Hoewel deze definitie ruimte biedt voor een ruim repertoire aan pedagogisch handelen, blijkt in de praktijk dat met name van de kant van beleid opbrengstgericht werken toch vooral gezien wordt als de inzet van (gestandaardiseerde) vormen van begin- en eindmetingen waarbij 'opbrengst' gezien wordt als iets dat met 'harde' gegevens kan worden onderbouwd. In de optiek van de Onderwijsraad betekent opbrengstgericht werken dat steeds opnieuw de stappen van de evaluatieve cyclus worden doorlopen: het stellen van doelen, het meten van resultaten, het analyseren van deze resultaten en het aanpassen van het onderwijsaanbod. Volgens de Raad werkt in 2011 een derde van de scholen opbrengstgericht, waarmee wordt bedoeld dat ze handelen in lijn met de eisen die de inspectie op dit gebied stelt. De Onderwijsraad acht een leerlingvolgsysteem 'onmisbaar voor opbrengstgericht werken' (Onderwijsraad, 2011: 13). In het advies van 2011 schaarde de Onderwijsraad zich ook achter het voorstel een eindtoets verplicht te stellen maar adviseert daarbij scholen de keus te laten zelf een eindtoets te kiezen. De Raad adviseert de overheid wel met een 'eigen' toets, ontwikkeld onder regie van het College voor Toetsen en Examens, te komen. De Raad stelt ook dat de eindtoets, mits 'geijkt' op de Referentieniveaus tevens kan dienen als een landelijke peiling onder leerlingen in groep acht op stelselniveau in het kader van een 'benchmark' en het vaststellen van 'leerwinst'. De Raad stelt dat de huidige toetsgegevens (nog) niet gebruikt kunnen worden voor het vergelijken van scholen onderling maar vooral het lerend vermogen van scholen zelf moeten stimuleren (Onderwijsraad, 2011: 7). De Raad wijst op de risico's die samen kunnen hangen met een te grote sturing op resultaat die volgens de Raad kunnen leiden tot 'neveneffecten in de vorm van ongewenst strategisch gedrag' (Onderwijsraad, 2011: 18).

In 2013, nauwelijks twee jaar later, komt de Onderwijsraad met een kritisch rapport over de ingezette ontwikkelingen. In het advies over de kwaliteit van onderwijs met als titel 'Een smalle kijk op onderwijskwaliteit' pleit de Raad voor meer aandacht voor de volle breedte van de onderwijsambities. Dit ter voorkoming van de risico's die de Raad in 2011 al signaleerde. De Raad stelt dat de aandacht te eenzijdig gericht is geraakt op meetbare doelen, in het bijzonder op het verhogen van taal- en rekenprestaties. Dat heeft ertoe geleid dat er veel minder beleidsaandacht is voor het bredere vakkenaanbod, algemene vorming en beroepspraktijkvorming in de optiek van de Raad. De voor de marktwerking noodzakelijke openheid van onderwijsresultaten is de laatste decennia vergemakkelijkt door de mogelijkheden en het gebruik van de informatietechnologie. Het bewaren en opslaan van gegevens van steeds meer digitaal afgenomen toetsen is de laatste jaren enorm vereenvoudigd. De groei van de digitale mogelijkheden heeft ook een keerzijde, volgens

de Raad. Digitale toetsing, al dan niet in combinatie met leermiddelen die via digitale platforms meer en meer hun intrede doen in de basisscholen, betekent veelal een primaire focus op vooral cognitieve toetsen die bestaan uit meerkeuzevragen en op vaardigheden die zich lenen voor digitale toetsing. Tal van 'lastig te meten variabelen' dreigen daarmee onder de radar te verdwijnen zonder dat dit het gevolg is van een debat over het belang van deze variabelen. De Raad stelt dat juist door prestatie-verhogende maatregelen scholen onvoldoende ruimte hebben om accenten te leggen in hun onderwijsaanbod of om te vernieuwen. Tot slot wijst de Raad op de gevolgen die zich aftekenen voor de eigenwaarde van leerlingen die niet goed presteren op basisvaardigheden. Dit is onder druk komen te staan volgens de Raad: 'Het vroege keuzemoment voor vmbo, havo of vwo accentueerde al cognitieve verschillen tussen jongeren van verschillende sociale achtergronden, doordat afkomst nog steeds een rol speelt bij schoolkeuze. Door de toename van categorale klassen dreigen nu ook de sociale verschillen groter te worden.' (Onderwijsraad, 2013: 9). Een van de aanbevelingen van de Onderwijsraad is om scholen meer de ruimte te geven 'in ruil voor' meer aantoonbare professionaliteit (Onderwijsraad, 2013: 9).

In 2016 verschijnt het rapport 'De volle breedte van onderwijskwaliteit; van smal beoordelen naar breed verantwoorden'. Daarin stelt de Raad vast dat het beperkt inzichtelijk maken van kwaliteit ook de erkenning hiervan belemmert. Er lijkt volgens de Raad sprake te zijn van een vicieuze cirkel: Kwaliteit in brede zin wordt in toezicht en verantwoording nog nauwelijks gewaardeerd, daardoor brengen scholen kwaliteit beperkt in beeld, wat de erkenning van de bijdrage van scholen op het brede terrein van onderwijs in de weg staat. De raad stelt in dit advies een drievoudige verbreding van onderwijskwaliteit voor, namelijk (1) van de opvatting van kwaliteit, (2) van de manieren waarop deze kwaliteit inzichtelijk gemaakt wordt en (3) van de verantwoording over onderwijskwaliteit (Onderwijsraad, 2016: 7-8).

In het rapport over de stand van educatief Nederland met als titel 'Doorgeschoten differentiatie in het onderwijsstelsel' uit 2019 komt de Onderwijsraad tot de conclusie dat differentiatie te ver is doorgeschoten. De Raad signaleert de volgende problemen met het stelsel: '(1) Jongeren uit verschillende sociale groepen komen elkaar niet meer vanzelfsprekend tegen in een sterk gedifferentieerd onderwijsstelsel; (2) overgangsbepalingen vroeg in de schoolloopbaan worden steeds bepalender door afnemende mobiliteit in het stelsel; en (3) permanente educatie is geen formeel onderdeel van het stelsel.' (Onderwijsraad, 2019: 7-9). Dit heeft volgens de Raad twee belangrijke gevolgen. Het 'versterkt de tendens van sociale segmentering en verkleint de mogelijkheden van het onderwijs om bij te dragen aan sociale samenhang in de maatschappij (1), en het 'zet de toegankelijkheid van en doorstroom binnen het onderwijs onder druk, waardoor sommige groepen leerlingen en studenten minder kansen krijgen in het onderwijs (2)'. De Raad komt tot de conclusie dat het huidige onderwijsstelsel op onderdelen herziening behoeft. Ook bestaat het risico dat onderdelen van het stelsel zich los van elkaar ontwikkelen (Onderwijsraad, 2019: 7-9). Volgens de Raad wordt het daardoor steeds moeilijker voor de overheid om haar stelselverantwoordelijkheid waar te maken. De Raad wijst op de versnippering in aandacht vanuit onderwijsbeleid dat vooral gericht is op afzonderlijke onderwijssectoren, waardoor volgens de Raad het risico bestaat 'dat opeengestapelde kleine maatregelen het stelsel wel veranderen, maar niet altijd in de gewenste richting en zonder dat onderliggende knelpunten voldoende worden aangepakt.' (Onderwijsraad, 2019: 25).

De diverse rapporten in de afgelopen twee decennia laten een wisselend beeld zien van de advisering ten aanzien de bewaking van kwaliteit en de verantwoording daarover, al dan niet met behulp van tests.

### 5.3 Kwaliteit en testen

Evalueren is een kernbegrip in onderwijs. Het gaat bij evalueren over het bekijken en waarderen van het resultaat aan de hand van de doelen.<sup>22</sup> Evaluatie kan betrekking hebben op verschillende actoren, zoals de leerling, de leerkracht, de school of het onderwijs. En evaluatie kan betrekking hebben op resultaten of op proces. Onderwijskundige Wiliam stelt dat evalueren in de context van onderwijs twee functies kan hebben. In het ene geval is het doel van evalueren de balans op te maken van het gegeven onderwijs om te komen tot een oordeel of een leerling wel of niet voldoende begrip heeft van de leerstof in termen van voldoende of onvoldoende. Dit evalueren wordt summatief genoemd. De andere functie van evalueren is formatief. Bij formatief evalueren gaat het om het verbeteren van het functioneren van de leraar. De leraar kan op basis van de uitkomsten het begrip of het leren of zijn uitleg verbeteren. De aandacht ligt bij het proces van leren. Evaluatieresultaten worden in dat geval gebruikt om lesinstructies aan te passen en beter te voldoen aan behoeften van leerlingen. Formatief evalueren blijkt te behoren tot de meest effectieve onderwijspraktijken (Wiliam, 2013: 6). Goede feedback is volgens Wiliam motiverend om te leren. Het gros van de feedback, waaronder het geven van cijfers, is volgens Wiliam echter contraproductief. Bij formatief evalueren gaat het om het zoeken naar het genereren van die vormen van feedback die de leerling motiveren om te leren. Een test wordt gezien als een instrument in het kader van evaluatie van onderwijs; de uitkomsten van een test kunnen worden meegenomen in een evaluatie. Tests in de context van formatief evalueren zijn tools in de handen van de onderwijsprofessional die hij gebruikt om te kunnen evalueren. Bij summatief evalueren ligt dat anders. Dit geldt zeker daar waar gestandaardiseerde tests ingezet worden om te evalueren die door andere instanties dan de docent of school zelf worden opgelegd aan het onderwijs. De vraag is welke feedback daarmee gegenereerd wordt en welke functie de feedback heeft voor wie. En hoe die feedback op basis van tests wordt meegenomen in de evaluatie. Popham stelt dat voor de Verenigde Staten geldt dat tests in onderwijs zich ontwikkeld hebben van 'teachers' tools into teachers' terrors' (Popham, 2001: 2). Het soort tests waar Popham hier aan refereert maken geen deel uit van de cyclus van leren maar hebben de functie om aan de hand van uitkomsten op tests een oordeel te vellen over de kwaliteit van het gegeven onderwijs, een summatieve vorm van evalueren. Wiliam stelt dat evaluatie niet zowel formatief als summatief gebruikt kan worden, dus niet tegelijkertijd gericht kan zijn op het verbeteren van leren en op het beoordelen van leren. Dit is wel wat in de praktijk van onderwijs door elkaar loopt. Het (extern) beoordelen van het leren wordt gezien als prikkel om het leren te verbeteren. Van een summatieve test wordt dan gezegd dat die formatieve waarde heeft waardoor verwarring ontstaat en bij veel tests het zicht op de daadwerkelijke functie van een test lastig te bepalen wordt. Dit geldt ook voor de eindtoets. De eindtoets is een extern ontwikkeld en gevalideerd instrument dat ingezet wordt om, onafhankelijk van het gegeven onderwijs en de onderwijsgegenden, te evalueren. Dit vaststellen van de kwaliteit is niet vrijblijvend maar gebeurt vanuit de gedachte dat dit aanzet tot verbetering van onderwijs.

De legitimiteit van het gebruik van toetsen in het onderwijs wordt in het maatschappelijk debat

22| <https://wij-leren.nl/evalueren.php>

aanvankelijk vooral onderbouwd vanuit het streven naar gelijke kansen ofwel het wegwerken van achterstanden. De waarde van toetsen wordt vooral benadrukt door te wijzen op de diagnostische werking die van toetsing kan uitgaan. Ledoux et al. spreken van 'data-driven teaching' en definiëren dat als een vorm van onderwijs waarbij leraren, schoolmanagers én leerlingen zich in hun taakuitoefening laten leiden door uitkomsten van metingen van (leerling)vorderingen (Ledoux et al. (2009). Daarbij wordt voorbij gegaan aan het feit dat de belangrijkste functie van bestaande gestandaardiseerde toetsen de aanlevering van voor leerkrachten en leerlingen relevante, diagnostische informatie in de weg staat. Want het vergelijken van leerlingen met elkaar staat centraal en niet de vergelijking met leerdoelen en de stand van ontwikkeling ten opzichte van de leerdoelen. Daar komt bij dat de gevolgen die de uitkomsten op deze toetsen hebben voor leerlingen deze toetsen tot 'high stakes' maken. Internationale ervaringen met highstakesvormen van testen onder regie van de overheid laten niet zomaar een rooskleurig beeld zien van het effect van dergelijke tests op de kwaliteit van onderwijs. De ervaringen wijzen meer in de richting van een veelheid aan ongewenste, negatieve gevolgen voor alle betrokkenen, met name voor kinderen die het meest gebaat zijn bij kwalitatief hoogwaardig onderwijs.

### 5.3.1 Kwaliteit van onderwijs vanuit internationaal perspectief

Tests zijn in de afgelopen decennia een steeds belangrijker rol gaan spelen in het bepalen van kwaliteit van onderwijs. Onderwijs is daarbij steeds meer beschouwd als een productieproces waar met een bepaalde input bij een gevolgd proces een bepaalde output verwacht kan worden. Daar waar overheden meer zijn gaan sturen op onderwijs als productieproces, is gezocht naar vormen om opbrengsten van onderwijs niet alleen inzichtelijk te maken maar ook op te eisen in termen van minimaal te behalen prestaties. Op het eerste gezicht lijken tests die gericht zijn op het al dan niet behalen van minimumdoelen vooral in het belang te zijn van de leerlingen gericht op het verbeteren van onderwijs zodat alle leerlingen de minimumdoelen zouden kunnen gaan behalen. Belangrijk in dit verband zijn ervaringen met testen in de Verenigde Staten. Gestandaardiseerd testen is in de Verenigde Staten ingevoerd door politici die hun twijfel etaleren over de kwaliteit van publiek onderwijs (Popham, 2001: 5). In 1983 verschijnt in de VS een rapport met de alarmerende titel 'A Nation at Risk'. In dit rapport wordt op basis van tal van nationale en internationale onderzoeken de noodklok geluid over de opbrengsten van het Amerikaanse onderwijs. Een van de zorgpunten wordt gevormd door de scores op de SAT<sup>23</sup> die in de periode 1963-1980 gemiddeld gesproken dramatisch zouden zijn gedaald. De act 'No Child Left Behind' (NCLB) is de actie die wordt ingezet om wat te doen aan de alarmerende uitkomsten en die beoogt met behulp van uniforme, landelijk gestandaardiseerde tests scholen te stimuleren meer opbrengsten te genereren. Een belangrijk doel van NCLB is het verkleinen van de 'Achievement Gap'. Of zoals Koretz het formuleert: 'To reduce inequality by imposing high standards on low-performing schools.' (Koretz, 2017: 129). De tests gebaseerd op minimale doelen voor rekenen en lezen zijn vanaf 2002 jaarlijks ingezet om vast te stellen of scholen en leerkrachten voldoende kinderen naar een minimaal niveau weten te brengen, met grote gevolgen voor leerkrachten en scholen bij het niet halen van de minimale niveaus. De ambities zijn zodanig geformuleerd dat verzet hier tegen niet of nauwelijks mogelijk is. Met brede steun is het programma NCLB dan ook gestart.

---

23| In eerste instantie stond SAT voor 'The Scholastic Aptitude Test'. Tegenwoordig is het een betekenisloze afkorting. De SAT is een vorm van een intelligentiemeting en is te beschouwen als een variant van de Army Test (Lemann, 1999: 31). De score op de SAT is onder andere bepalend voor toelating tot de prestigieuze universiteiten van de Verenigde Staten (Sandel, 2020: 164).

Ravitch, onderwijshistoricus en voormalig lid van de National Assessment Governing Board van de Verenigde Staten, beschrijft hoe het testen het Amerikaanse onderwijs in zijn greep heeft gekregen en wat daarvan de gevolgen zijn. In eerste instantie is zij voorstander van de NCLB-act maar door alle gebleken effecten is zij nu tegenstander van het willen meten van onderwijsopbrengsten met 'high stakes tests'.

De NCLB Act is vanwege alle effecten en commotie afgezwakt in de vorm van de Every Student Succeeds Act (ESSA) die sinds 2015 de NCLB vervangt. Koretz stelt dat van 'testbased accountability' zeker niet gezegd kan worden dat het geen verbeteringen heeft gebracht, maar dat de negatieve effecten vooral tegen de magere, positieve resultaten moeten worden afgezet en dat het van belang is naar de 'netto score' te kijken (Koretz, 2017: 45). En dan komt Koretz niet tot een positieve slotsom, al gelooft hij wel dat velen gedacht hebben dat deze aanpak een effectieve manier zou zijn om ongelijkheid tegen te gaan (Koretz, 2017: 129): 'On balance, then, the reforms have been a failure.' (Koretz, 2017: 192). En de negatieve effecten zijn in zijn optiek groot. Hij wijst op de volgende vier grote problemen die samengaan met testbased accountability zoals in het geval van NCLB:

1. 'Er gaat veel onderwijstijd verloren met allerlei ongewenste testvoorbereiding wat ten koste gaat van zinvolle instructietijd, met name voor de kinderen aan de onderkant die meer gebaat zijn bij goede instructie;
2. het leidt tot corruptie van onderwijsidealen zoals manipulatie van de toetspopulatie, manipulatie of bedrog op basis van resultaten en de creatie van stress bij alle betrokkenen (docenten, ouders en leerlingen);
3. beide voorgaande punten leiden tot score-inflatie en geen verbetering van onderwijsprestaties waarmee het eigenlijke doel van testbased accountability ondermijnd wordt;
4. vooral kinderen met achterstanden hebben hier last van zodat de verschillen in resultaten tussen achterstandskinderen en niet-achterstandskinderen alleen maar groter zijn geworden terwijl 'testbased accountability' juist ingevoerd wordt met het argument om achterstanden tegen te gaan' (Koretz, 2017: 6-7).

Gestandaardiseerde tests met een dergelijk high stakes karakter zijn volgens Koretz te beschouwen als sociale indicatoren in de zin zoals Campbell bedoelt. Campbell komt in 1975 tot de slotsom dat iedere kwantitatieve sociale indicator die gebruikt wordt voor het nemen van vergaande beslissingen het risico loopt ten prooi te vallen aan corruptie: 'The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.' (Ravitch, 2010: 170). Van standaardtests gaan effecten uit die met deze naar Campbell genoemde wetmatigheid samenhangen. Tests doen dingen die je niet wilt dat ze doen en wat ze meten is onderhevig aan inflatie zodat ze een te rooskleurig beeld geven van wat ze beoogd worden vast te stellen en te verbeteren (Koretz, 2017: 38). Ook Popham wijst op de kwalijke gevolgen die vormen van gestandaardiseerde high stakes tests hebben voor onderwijs. Hij spreekt in dit verband over 'curricular exodus' waarmee hij bedoelt dat onderwijs zich door deze vormen van testen steeds meer is gaan beperken tot de stof die in de tests aan bod komt en niet meer richt op het beoogde curriculum (Popham, 2001: 19). Hij is daarbij kritisch over de rol van het onderwijs zelf (inclusief hemzelf als docent) en stelt dat de leraren te weinig hebben gedaan om tegenwicht te

bieden aan de inzet van gestandaardiseerde tests: 'We allowed students' test scores to become the indicator of our effectiveness. We failed to halt the profound misuse of standardized achievement tests to judge our educational quality. We let it happen to ourselves. And more grievously, we let it happen to the children we are supposed to be educating.' (Popham, 2001: 13).

Ravitch benadrukt het belang dat moet worden gehecht aan professionele oordelen waar testcores eventueel een aanvulling op zijn (Ravitch, 2010: 173). Testresultaten kunnen in haar optiek geen vervanging zijn van de input van de professional, hooguit een aanvulling. Relativering van de waarde van metingen voor het handelen van de professional vraagt om sterke professionals die die relativering kunnen aanbrengen. Omdat burgers geloven dat hoge scores een teken zijn van goed onderwijs hebben de jaarlijkse vergelijkingen van scholen in de vorm van openbaar gemaakte rankings een enorme druk gelegd op docenten om de scores van leerlingen te verbeteren (Popham, 2001: 11). Met de introductie van dergelijke tests is tegelijkertijd gelegitimeerd dat kwaliteit van onderwijs ontleend kon worden aan leerlingsscores op gestandaardiseerde tests (Popham, 2001: 8).

Rothstein, Jacobsen en Wilder stellen dat het curriculum van public schools zich onder druk van accountability tests in het kader van de act 'No Child Left Behind' (NCLB) in hun dagelijkse praktijk meer en meer blijkt te zijn gaan beperken tot dat wat in de tests aan bod komt (met name rekenen en lezen). De aandacht voor andere vakken, zoals bijvoorbeeld geschiedenis, is gedaald, gebleken uit een daling van het aantal staten dat hierop test, van 27 naar 19 (Rothstein et al., 2008: 47). In een accountability-systeem geldt volgens Rothstein dat wat niet wordt gemeten, niet aan bod komt (Rothstein et al., 2008: 49). De kans dat scholen toch aandacht besteden aan inhoud die niet in de tests is opgenomen, is volgens hem het grootst op elitescholen omdat zij relatief de minste tijd nodig hebben om de 'accountability norm' te halen voor taal en rekenen (Rothstein et al., 2008: 49). Het werken met vaardigheidsniveaus die door scholen moeten worden behaald, hoe redelijk die ook lijken gesteld, brengt daarnaast met zich mee dat de onderwijsaandacht vooral uitgaat naar de leerlingen die net onder de 'voldoende' zitten, leerlingen die in de VS worden aangeduid als 'bubble kids', een term uit poker en basketbal waarmee spelers en teams worden aangeduid die dreigen te degraderen. Voor hen in het bijzonder wordt veel gedaan in het drillen en oefenen met tests in plaats van het geven van goed onderwijs (Rothstein et al., 2008: 67). Eén standaardscore die moet worden behaald door alle kinderen onder het motto dat 'alle kinderen kunnen leren' doet volgens Rothstein geen recht aan de werkelijkheid dat niet alle kinderen de gevraagde doelen op hetzelfde moment beheersen, simpelweg omdat ze niet op hetzelfde punt beginnen en niet allemaal even snel kunnen. Professionals voor de klas weten dit, maar het accountability-systeem maakt dat velen van hen niet langer werkzaam willen zijn op achterstandsscholen of op public schools in het algemeen (Rothstein, 2008: 71). Koretz stelt dat het werken met uniforme standaarden in gestandaardiseerde toetsen twee kanten op kan gaan: of we stellen de standaarden zo hoog, onrealistisch hoog, dat manipulatie van scores nodig is om die scores voor alle kinderen te behalen ofwel we verlagen de standaarden zodanig dat alle kinderen ze kunnen halen (Koretz, 2017: 131). Nichols en Berliner stellen dat high-stakes tests vanwege dit effect de praktijk van het onderwijs zelf beïnvloeden. Het trainen voor tests brengt met zich mee dat de indicator waar de test oorspronkelijk informatie over zou geven, geen nuttige informatie meer kan verschaffen omdat training de resultaten heeft beïnvloed (Nichols, 2007: 123). Zij concluderen dat zodra het belang van een test toeneemt (high stakesness), er



veel inspanning en middelen zullen worden besteed aan het goed scoren op die test, wat niet direct leidt tot grondiger leren of breder leren: 'Effort is not necessarily directed to learning at deeper levels nor at learning unrelated but interesting and useful content.' (Nichols, 2007: 141). Popham stelt dat gestandaardiseerde tests niet geschikt zijn om de kwaliteit van onderwijs vast te stellen vanwege het probleem van causatie. Er zijn volgens hem drie factoren die bijdragen aan de scores op een test van een individuele leerling, namelijk dat wat op school geleerd is, het aangeboren intellectuele vermogen en wat een leerling leert buiten de school om. Daarmee zijn gestandaardiseerde tests ongeschikt om effectiviteit van onderwijs te kunnen vaststellen (Popham, 1999). Tests zoals gebruikt in dit soort high stakes contexten zijn geen middel meer om het leerproces te bevorderen maar brengen met zich mee dat het leren wordt gebruikt om de testprestaties omhoog te duwen. Ook Koretz stelt dat testcores altijd gezien moeten worden als incomplete maten van leerlingprestaties waar altijd een aanvulling op basis van docentobservaties en maatwerktests op nodig is (Koretz, 2017: 17). Standaardtests zijn beperkt in de informatie die ze kunnen geven. Niet alles leent zich even goed voor gestandaardiseerd meten wat leidt tot het maken van keuzes. De meeste tests kunnen daarbij de last die op hun schouders is gelegd niet dragen, omdat het fenomeen van onnauwkeurigheid of 'error' een rol speelt in de scores van individuele kinderen en daarmee ook in alles wat vervolgens met die scores gebeurt. Omdat tests geen goede afspiegeling vormen van het domein dat ze willen meten, vernauwen ze het onderwijs dat zich focust op wat er in de test aan bod komt. Daarmee zijn standaardtests ernstige bedreigers van onderwijsdoelen, aldus Koretz (Koretz, 2017: 17).

Broadfoot beschrijft de ervaringen die in de jaren zeventig in Engeland en Wales zijn opgedaan met nationale tests voor het monitoren van de ontwikkeling van belangrijke curriculumonderdelen als natuurkunde, wiskunde en taal. Zij stelt dat het in de praktijk extreem moeilijk is gebleken, zo niet onmogelijk, om hiervoor indicatoren op te stellen. En nog minder is het gelukt om daarmee de lespraktijk en dus de kwaliteit van onderwijs te verbeteren (Vedder, 1992: 114). Ook Broadfoot wijst op het gevaar van versmalling van het curriculum tot dat wat in de test aan bod komt. Dit heeft er in Engeland toe geleid dat de nationale test zo breed dekkend is gemaakt voor het curriculum dat de afname een onacceptabele hoeveelheid tijd van leerkrachten en leerlingen in beslag nam (Vedder, 1992: 116).

Niet alle westerse landen kiezen voor tests gericht op het kunnen vergelijken van opbrengsten. Finland heeft bijvoorbeeld vanaf de jaren zeventig met een andere visie op onderwijs gewerkt aan testen. Kangasniemi stelt dat in de jaren zeventig en tachtig tests vooral hebben gediend om scholen en meer specifiek docenten inzicht te geven in prestaties van leerlingen door scholen informatie te geven over scores van leerlingen die mee hadden gedaan aan nationale tests zodat die vergeleken konden worden met de eigen scores van eigen leerlingen. Dit met als doel het eigen referentiekader van de docent te scherpen. Volgens Kangasniemi is het belangrijkste doel van testen in de jaren negentig de invoering van het nationale curriculum te monitoren. Dit betekent volgens hem dat de uitkomsten vooral bedoeld zijn voor curriculumontwikkelaars en voor docenten en leerlingen. De uitkomsten zijn niet bedoeld voor ouders (Vedder, 1992: 107).

Het gebruik van gestandaardiseerde, vergelijkende manieren van meten is internationaal gestimuleerd door de OECD. Met behulp van periodieke, internationaal vergelijkende onderzoeken, zoals PISA, PIRLS en TIMMS, hebben dit soort vergelijkende onderzoeken die in de jaren negen-

tig zijn begonnen aan belang gewonnen. De internationale aandacht voor het meten van leeropbrengsten is onder leiding van de OECD in veel landen een belangrijke vorm van input geworden voor het bepalen van een oordeel over kwaliteit van onderwijs. En die waarde die aan de uitkomsten wordt toegekend leidt vervolgens tot een verdere legitimering van dergelijke onderzoeken.

Finland vormt in Europa nog altijd een belangrijke uitzondering. Sahlberg laat in zijn boek 'Finnish Lessons' zien dat in Finland op een geheel andere manier wordt gekeken naar onderwijskwaliteit. Hoewel het streven naar transparantie en verantwoording tot een verbeterde informatiestroom richting ouders en politici leidt, zorgt het volgens Sahlberg ook voor wantrouwen, een lage motivatie en professioneel cynisme (Sahlberg, 2016: 180). Sahlberg baseert zich daarbij op de Canadese socioloog Michael Fullan die stelt dat leiders (...) de neiging hebben om verkeerde drijvende krachten te kiezen. Dat zijn volgens Fullan: verantwoording (versus professionaliteit), individuele kwaliteiten van de leraar (versus collegialiteit), technologie (versus pedagogiek) en een gefragmenteerde aanpak (versus systeemdenken) (Sahlberg, 2016: 145). Fullan komt met vier criteria waarop een onderwijsaanpak moet worden bevestigd wil de aanpak effectief zijn in het kader van de verbetering van onderwijs: 'Specifically, do the drivers, sooner than later, 1) foster intrinsic motivation of teachers and students; 2) engage educators and students in continuous improvement of instruction and learning; 3) inspire collective or team work; and 4) affect all teachers and students 100 per cent?' (Fullan, 2011: 3). Als deze vragen niet positief kunnen worden beantwoord, is er volgens Fullan sprake van een foute prikkel, 'a wrong driver', die weinig kans heeft beoogde ambities te realiseren terwijl een goede prikkel, 'a right driver', die wel aan de vier eisen voldoet, in staat is om betere resultaten voor leerlingen dichterbij te brengen: 'A wrong driver is a deliberate policy force that has little chance of achieving the desired result, while a right driver is one that ends up achieving better measurable results for students.' (Fullan, 2011: 3).

De OECD laat elke drie jaar een test afnemen in een groot aantal landen waarmee de landen die meedoen worden vergeleken op de leesvaardigheid, wiskunde en natuurwetenschappen van hun 15-jarigen: Programme for International Student Assessment (PISA). Ook voor deze meting geldt dat de vergelijking het uitgangspunt is. PISA beoogt 'te voorzien in internationaal vergelijkbare evidentie over kwaliteit, de gelijkheid en de doelmatigheid van schoolsystemen' (Standaert, 2017: 78). Daarmee zijn de metingen van PISA volgens Standaert niet zozeer hulpmiddelen voor het lokale onderwijsbeleid, maar gaat het om een arbitrage van de bestaande onderwijsystemen in de wereld. PISA is niet alleen diagnosticus maar ook beoordelaar en adviseur voor lokaal onderwijsbeleid (Standaert, 2017: 78). Standaert wijst op tal van manieren waarop ruis bij het ontwikkelen van tests en bij het verzamelen en interpreteren van data kan ontstaan en stelt dat PISA als onderneming kan worden gekarakteriseerd als een 'bijna ondoordringbaar woud' (Standaert, 2017: 82). Hij pleit voor veel meer bescheidenheid ten aanzien van de uitkomsten van de kant van de OECD. De fascinatie voor de rankings van landen vindt Standaert niet terecht (Standaert, 2017: 83). Toch hebben de uitkomsten impact op de meeste deelnemende landen. Nederland neemt elke drie jaar een relatieve positie in ten opzichte van andere landen. Over de conclusies wordt veelal gesproken alsof het hier gaat om 'harde' feiten. Zo meldt de onderwijsinspectie in de Staat van het Onderwijs dat 'Nederlandse leerprestaties dalen' en dat 'Nederland 1 van de 7 landen is waarbij sinds 2006 de scores op de 3 onderdelen op de lange termijn teruglopen' (Inspectie van het Onderwijs, 2020). We kunnen echter tot de beste twintig procent van de wereld behoren op wiskunde en wel degelijk achteruit zijn gegaan en we kunnen behoren tot

de zwakste twintig procent en wel degelijk vooruit zijn gegaan. Bij een relatieve meting wordt de uitkomst mede bepaald door de andere deelnemers. Standaert stelt dat 'de ambities van de PISA-toetsen overgeaccentueerd worden en dat de resultaten een betekenis krijgen die de validiteit van de data en de dataverzameling ver te boven gaat' (Standaert, 2017: 83).

### 5.3.2 Ontwikkelingen in Nederland

In Nederland ligt de focus in onderwijsbeleid en in onderwijsonderzoek in de laatste decennia op de vraag of en in hoeverre scholen van elkaar verschillen in effectiviteit en welke factoren daar een rol in spelen. Sociologisch onderzoek in de Verenigde Staten zoals het grote onderzoek van Coleman (zie ook hoofdstuk 6) is aanvankelijk vooral gericht op het zoeken naar effecten van macro-variabelen zoals locatie, faciliteiten of budget. Dit levert weinig resultaat op. Dit leidt ertoe dat onderzoek vooral gericht wordt op schoolinterne factoren. Vanaf het eind van de jaren zeventig ontstaat in verschillende westerse landen 'schooleffectiviteitsonderzoek' gericht op 'het achterhalen van factoren die een school effectief kunnen maken, in de manier waarop een school te werk moet gaan om goede resultaten bij de leerlingen te bewerkstelligen' (Dekkers, 1999: 16). De vraag of de ene school effectiever is dan de andere gaat volgens Wilbrink voorbij aan de vraag welke effecten twee scholen gemeenschappelijk hebben. Onderzoek naar het verbeteren van macro-effecten lijkt volgens hem aan de aandacht te ontsnappen terwijl daar grote verbeteringen van het stelsel en in het verlengde daarvan voor de kwaliteit van de samenleving wellicht mee mogelijk zouden zijn (Wilbrink, 1995). De focus op het vergelijken op basis van verschillen, heeft ook het toezicht op het onderwijs in Nederland bepaald. In hun kritische beschouwing over de validiteit van een vijftal meer of minder gangbare indicatoren voor schooleffectiviteit stellen Blok et al. dat alle gebruikte indicatoren gevoelig zijn voor invaliderende effecten. Zij stellen dat de begripsvaliditeit van schooleffectiviteitsindicatoren niet te beoordelen is en dat de indicatoren vooral het resultaat zijn van een 'black box benadering' (Blok et al., 2016).

Zowel bruto leeropbrengsten als toegevoegde-waarde-indicatoren (Blok et al., 2016) worden gebruikt en van waarde geacht voor het bepalen van de kwaliteit van onderwijs zoals de Inspectie van het Onderwijs die vaststelt. Het voorwoord bij het jaarlijkse rapport dat de Inspectie van het Onderwijs aanlevert aan de minister van onderwijs stelt: 'In deze Staat van het Onderwijs 2020 rapporteren wij over het functioneren van het onderwijsstelsel in het afgelopen jaar (...). Daarnaast gaat de Staat van het Onderwijs 2020 dieper in op verschillen tussen de prestaties die scholen met vergelijkbare leerlingpopulaties realiseren, op het onderwijs aan leerlingen met extra ondersteuningsbehoeften en op vragen rond de inrichting van en sturing op het onderwijsstelsel.' (Inspectie van het Onderwijs, 2020b).

Uit gegevensanalyses blijkt volgens deze Staat dat scholen met 'eenzelfde leerlinggewicht sterk verschillen in hun resultaten op de Centrale Eindtoets (Inspectie van het Onderwijs, 2017). Dit is ook zichtbaar als we scholen met eenzelfde CBS-schoolweging (tussen 28 en 31, dat is rond het gemiddelde) vergelijken op het aandeel leerlingen dat het streefniveau 2F/1S bereikt.' (Inspectie van het Onderwijs, 2020).

Tot aan het schooljaar 2020-2021 is de gemiddelde standaardscore van de leerlingen van een school het uitgangspunt geweest voor toezicht. Het vergelijkend karakter van de eindtoets impliceert dat het verhogen van de resultaten van de ene school 'ten koste' gaat van de resultaten van

een andere school. Met een relatieve meting is het eenvoudigweg onmogelijk om alle scholen te laten eindigen bij de beste twintig procent. Dit maakt dat gesprekken van de Inspectie van het Onderwijs met scholen over de uitkomsten van onderwijs wat bevreemdend overkomen. In de Staat van 2020 rapporteert de Inspectie van het Onderwijs: 'Op 41 basisscholen zochten we in gesprekken met leerlingen, leraren en managementteams naar verklaringen voor de leerresultaten die zij realiseren, vanuit de vraag: 'wat werkt op deze school om tot hoge(re) leerresultaten te komen?' (Inspectie van het onderwijs, 2020).

Vanaf het schooljaar 2020-2021 gaat de Inspectie van het Onderwijs met een aangepast waarderingskader werken. Het toevoegen van referentieniveaus als 'absolute norm' lijkt een 'oplossing' voor de beperkte informatie die een relatieve meting gebaseerd op gemiddelde standaardcores heeft. Het probleem van de absolute norm is wel dat deze is toegevoegd aan een bestaande toets met een selectie-oogmerk. Dat laatste maakt dat de toets juist het verschil moet kunnen maken tussen leerlingen. De eindtoetsen gebruiken inhoud en opgaven waarmee verschillen te maken zijn. Want dat is inherent aan een instrument gericht op selectie. Een toets primair bedoeld voor selectie heeft voldoende items nodig die zodanig moeilijk zijn dat daarmee met voldoende zekerheid de beste twintig procent van de leerlingen kan worden geselecteerd. Dit verhoudt zich slecht tot een toets die meer criteriumgerelateerd moet vaststellen hoeveel procent (liefst honderd procent) van de leerlingen bijvoorbeeld 2F voor lezen hebben behaald. De doelstelling van het willen verhogen van het aandeel scholen dat 2F behaalt (lezen en taalverzorging) of 1S (rekenen) staat op gespannen voet met de ambitieuze doelstelling om met voldoende zekerheid passende schooladviezen te geven: de ene meetambitie kan niet anders dan ten koste gaan van de andere.

Omdat de eindtoets primair gericht is op selectie is de eindtoets volgens Cito niet geschikt voor het doen van 'harde uitspraken over een daling of stijging van de gemiddelde vaardigheid die door de Eindtoets wordt gemeten' (Van Boxtel et al., 2011: 40). Dat is de reden dat de overheid in de loop van de jaren tachtig opdracht geeft voor het ontwikkelen en uitvoeren van zogenaamde 'cohortonderzoeken'.

### **Cohortonderzoeken**

In de jaren tachtig wordt onderzoek gedaan naar het rendement van inspanningen die plaatsvinden om achterstanden van specifieke doelgroepen tegen te gaan. Dit leidt tot divers onderzoek naar het rendement van het onderwijskansenbeleid. Van dit beleid dat de rijksoverheid in de jaren tachtig heeft gedelegeerd aan de gemeenten (Van Wieringen, 2000: 37) is er de wens de effectiviteit ervan te onderzoeken. Om vast te stellen of er verbetering optreedt in de onderwijskansen van leerlingen die behoren tot de doelgroepen van het achterstandenbeleid, wordt vanaf het schooljaar 1988/89 in het basisonderwijs en vanaf schooljaar 1989/90 in het voortgezet onderwijs cohortonderzoek uitgevoerd om (verbetering van) prestaties vast te stellen. Tot 2005 ging het om het LEO- het PRIMA-cohortonderzoek in het basisonderwijs en PRIMA-VO en VOCL in het voortgezet onderwijs. Daarna zijn deze cohorten geïntegreerd voortgezet onder de naam COOL5-18. Gaat het in eerste instantie nog om onderzoek naar het al dan niet verbeteren van de kansen van achterstandsléerlingen, afgemeten aan hun prestaties, in de loop van de tijd verschuift de focus naar het meten van onderwijsopbrengsten van onderwijs in het algemeen. Dit is in lijn met de discussies en de ontwikkeling van de rol van toetsing in de Verenigde Sta-

ten. In 1983 maakt de Nederlandse directeur-generaal en tevens inspecteur-generaal van het onderwijs, Veldhuis, een studiereis naar de Verenigde Staten op uitnodiging van de Amerikaanse regering. Hij schrijft daarover een verslag met als titel 'Een opzienbarend Amerikaans rapport over de kwaliteit van het onderwijs in de Verenigde Staten van Noord-Amerika' (Veldhuis, 1983). Hij doelt op het rapport 'A Nation at Risk' dat dan net verschenen is en veel stof doet opwaaien, ook buiten de grenzen van de VS zelf. In zijn verslag vat hij de discussie samen die in de VS woedt over de kwaliteit van het onderwijs in het algemeen en het rapport in het bijzonder en vergelijkt dat debat met de discussie over onderwijskwaliteit in Nederland. Hij vindt drie punten opvallend. Het eerste verschil is volgens Veldhuis dat de Amerikanen niet blijven hangen in een beschouwende discussie over wat nu kwaliteit van onderwijs is maar hier pragmatischer in zijn en de discussie meer laten gaan over instrumenten waarvan in brede kring wordt aangenomen dat ze de kwaliteit kunnen beïnvloeden. Het tweede verschil ziet hij in de cijfermatige onderbouwing op basis van testresultaten van de toestand van het onderwijs. Ten slotte stelt hij dat cijfermatige onderbouwingen op basis van toetsresultaten om de kwaliteit van het onderwijs vast te stellen in combinatie met een zo groot mogelijke openbaarheid noodzakelijk zijn om verantwoording af te kunnen leggen. Het gebruik van meer kwantitatieve gegevens ziet hij als een belangrijk ontwikkelpunt voor Nederland (Veldhuis, 1983: 25). Bij Cito is al eerder belangstelling ontstaan voor het uitvoeren van grootschalige onderwijs- en evaluatieprojecten naar aanleiding van een studiereis in 1979. In 1984-1985 heeft het toenmalige SCO onderzocht of zo'n project ook voor het Nederlandse basisonderwijs zinvol zou zijn (Wesdorp et al., 1986). De conclusie is bevestigend. In 1986 krijgt Cito van het Ministerie van Onderwijs en Wetenschappen de opdracht peilingonderzoeken te verrichten: Periodieke peiling van het Onderwijsniveau (PPON). In zijn Diesrede 'Het zwalkende niveau van ons onderwijs' uit 1987 wijst Van der Linden op het belang van periodieke peilingen als aanvulling op wat er dan aan toetsen en examens in het funderend onderwijs bestaat. Want, zo stelt hij, 'we bezitten dus een onderwijsstelsel waarin van begin tot eind de vorderingen van leerlingen relatief vastgesteld worden. Vaste kwaliteitsdrempels ontbreken. En het is onmogelijk de prestaties over de jaren heen te vergelijken' (Van der Linden, 1987: 11). In een context waarin 'onderwijs en maatschappij op drift zijn geraakt' is het niet te bepalen wat hiervan de effecten zijn op de prestaties van leerlingen: 'De beste hypothese lijkt dat het niveau maar wat meezwalkt met al deze ontwikkelingen en dat niemand weet waar we in feite staan.' (Van der Linden, 1987: 11). Hij pleit voor periodieke peilingen en is verheugd dat de overheid tot het inzicht is gekomen dat periodieke peilingen belangrijk zijn 'nadat andere landen ons reeds zijn voorgegaan'. Het peilingonderzoek PPON wordt door Cito gepositioneerd als een 'typische vorm van systeemevaluatie' waarbij het object van evaluatie de inhoudelijke kwaliteit van de leeropbrengsten is. In 1987 krijgt Cito de opdracht van de overheid om via een meting te 'bepalen of we nationaal gezien onze onderwijskwaliteit weten te handhaven in het basisonderwijs' (Cito, 2019: 126). Er worden daartoe aparte toetsen ontwikkeld. Van 1987 tot 2014 voert Cito de Periodieke Peiling van het Onderwijsniveau (PPON) uit en rapporteert de uitkomsten daarvan in publicaties die de naam 'balans' krijgen. Van der Schoot blikt in de publicatie 'Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON' terug op PPON. Hij vat de functie ervan als volgt samen: 'Positieve trends vragen om ondersteuning en negatieve om bijsturende maatregelen. De kwaliteit van ons onderwijs staat centraal. En telkens weer moeten we ons bezinnen. Wat is het rendement van onze onderwijsinspanning? En sluit dat rendement aan bij wat we ervan verwachten? Resultaten van peilingonderzoek leveren stof tot nadenken en input voor afwegingen over de kwaliteit en het niveau van het onderwijs.' (Van der Schoot, 2008: 14). De ingevoerde

stelselmetingen van PPOON betreffen in principe alle leer- en vormingsgebieden aan het einde van het basisonderwijs én – voor de gebieden Nederlandse taal en rekenen/ wiskunde – ook halverwege het basisonderwijs, aan het begin van jaargroep 5. De PPOON-onderzoeken vinden nog altijd plaats maar zijn nu onder regie geplaatst van de Inspectie van het Onderwijs zelf. Zij gebruiken de PPOON-onderzoeken om meer in de breedte van het onderwijs zicht te krijgen op de stand van zaken op stelselniveau. Het betreft altijd een steekproef van scholen waarop de PPOON-onderzoeken betrekking hebben. Daarmee zijn de PPOON-onderzoeken niet geschikt om individuele scholen te beoordelen. Dit in tegenstelling tot LVS en eindtoets. Die toetsen moeten door alle scholen bij alle leerlingen worden afgenomen en vormen daarmee voor de Inspectie van het Onderwijs een belangrijke bron van input voor de beoordeling van kwaliteit van onderwijs.

### 5.3.3 Inspectie van het Onderwijs en toezicht op kwaliteit

Het toezien op kwaliteit zoals dat in Nederland in de Wet op het Onderwijstoezicht (WOT) in 2002 is vastgelegd, is niet een vanzelfsprekende invulling van toezicht. Uit vergelijkend onderzoek van De Volder blijkt dat externe schoolevaluaties in Europa op macroniveau in twee groepen te onderscheiden zijn: inspecties die de klemtoon leggen op verantwoording en inspecties die het ontwikkelingsperspectief centraal stellen. Nederland valt volgens De Volder in de eerste groep (De Volder, 2012: 251). Binnen de tweedeling op macroniveau vond De Volder bij onderzoek op mesoniveau een verdere verfijning van onderscheid binnen de groepen. Binnen de inspecties waar het verantwoordingsperspectief overweegt, maakt De Volder onderscheid tussen inspecties met een risicogerichte benadering (waartoe zij Nederland rekent), inspecties waar de zelfevaluatie centraal staat en inspecties die een globale holistische benadering hanteren (waar zij Vlaanderen toe rekent). Bij de ontwikkelingsgerichte inspectiesystemen maakt zij onderscheid in inspecties waarbij zelfevaluatie hét uitgangspunt is en inspecties die een procesgerichte benadering hanteren. Vooral de analyse op mesoniveau, waarmee De Volder de uitgevoerde schoolevaluaties bedoelt, geven volgens haar goed inzicht in de typering van schoolevaluaties die zij heeft uitgevoerd aan de hand van een 'veertigpuntenbalk' (De Volder, 2012: 258). In zijn boek 'Inspectorates of Education in Europe' (2001) werkt de Vlaamse pedagoog Standaert een raamwerk uit voor het typeren van vormen van onderwijstoezicht. Centraal daarbij staat de visie op de invulling van het begrip 'kwaliteit van onderwijs'. De kwaliteit van onderwijs hangt volgens Standaert samen met drie belangrijke aspecten: wat wordt gezien als waardevolle doelstellingen, wat als efficiënte strategieën en wat als voldoende resultaten die worden gerealiseerd (de effectiviteit of de opbrengst)? (Standaert, 2001: 13). Standaert maakt onderscheid tussen twee manieren van denken over kwaliteit: de economisch-technische benadering die Standaert 'technocratisch' noemt en de pedagogisch-didactische benadering anderzijds die Standaert 'leerlinggericht' noemt (zie ook Hoofdstuk 3). De invalshoek bepaalt in hoge mate de invulling die aan de verschillende aspecten van kwaliteit worden gegeven.

De ontwikkeling van de Nederlandse inspectie naar een verantwoordingsinspectie met een risicogerichte benadering is een ontwikkeling die zich, tegen de achtergrond van tal van maatschappelijke ontwikkelingen, internationaal en nationaal, heeft voltrokken. De Inspectie van het Onderwijs in Nederland heeft zich tot aan de invoering van de Wet op het Onderwijstoezicht (2002) gericht op het rapporteren over de 'deugdelijkheid' van het stelsel, conform de Grondwet art. 23. De informatie van de Inspectie heeft tot aan het begin van deze eeuw geen betrekking op het geven van informatie over afzonderlijke scholen of over de kwaliteit van het gerealiseerde

onderwijs op schoolniveau (Bronneman-Helmers, 2011: 95). De kwaliteitseisen en maatstaven die de inspectie is gaan gebruiken bij het toezicht op schoolniveau krijgen pas een wettelijke basis na de invoering van de WOT (2002) (Bronneman-Helmers, 2011: 97). Met de WOT verschuift het toezicht op 'deugdelijkheid' zoals de Grondwet in artikel 23 voorschrijft naar een nieuwe interpretatie, namelijk toezicht op 'kwaliteit'. In de WOT wordt het de inspectie mogelijk gemaakt onderwijskwaliteit inhoudelijk te beoordelen op aspecten als het onderwijsleerproces, het schoolklimaat, de leeropbrengsten en de kwaliteit van de leerlingenzorg. Het toezicht moet niet alleen controleren zonder onnodig te sturen, maar ook stimuleren. Tussen deze twee uitgangspunten kan een spanningsveld ontstaan waarop de Onderwijsraad wijst (Onderwijsraad, 2006: 27). Na de invoering van de WOT komt de Inspectie van het Onderwijs in 2005 met een nieuw 'Toezichtkader PO'. Dit toezichtkader sluit dan vooral aan op de lijn die door het ministerie met de 'Koers Primair Onderwijs, Ruimte voor de school' is geformuleerd ten aanzien van de ruimte die er voor maatwerk in het onderwijs zou moeten komen. Het tot aan de WOT gehanteerde kader zou naar het oordeel van de inspectie te veel gelaagd zijn, zou te weinig discrimineren tussen goede en zwakke scholen, zou te massief zijn en zou te weinig mogelijkheden tot proportionaliteit en variëteit in het toezicht bieden. De Inspectie stelt in het nieuwe toezichtkader voor te werken met een basis-set indicatoren die voor alle scholen het waarderingskader vormt en tegelijkertijd scholen de ruimte biedt zelf indicatoren te selecteren als een vorm van maatwerk, passend bij een specifieke school. Zelfevaluatie krijgt nog een prominente plek in het voorstel voor toezicht. Het voorstel heeft daarmee twee gezichten: enerzijds is er ruimte voor de eigenheid en eigen keuzes van een school en anderzijds is het waarderingskader te zien als een streven en opstap naar standaardisering en meer uniformering in het toezicht. Het nieuw gezochte evenwicht tussen ruimte en rekenschap krijgt geen tijd om tot ontwikkeling te komen. Er wordt gezocht naar nieuwe vormen van verantwoording op basis van uniforme, niet door de school deels zelf te selecteren criteria om zo vergelijkingen op stelselniveau en vergelijkingen tussen scholen mogelijk te maken. Met de wettelijke basis van de WOT voor het toezicht van de inspectie op onderwijskwaliteit krijgt de inspectie de verantwoordelijkheid om het kwaliteitsbegrip nader uit te werken

#### **Missie: Effectief toezicht voor beter onderwijs.**

*Elk kind heeft recht op goed onderwijs. Leerlingen, studenten en ouders moeten erop kunnen vertrouwen dat het onderwijs op een school goed is. Het schoolbestuur is verantwoordelijk voor de kwaliteit van het onderwijs en moet zich verantwoorden over de resultaten. Hierbij gaat het om resultaten in brede zin; krijgen alle leerlingen/studenten onderwijs van voldoende kwaliteit, voldoen scholen en instellingen aan wet- en regelgeving en hebben ze hun financiën op orde? De Inspectie van het Onderwijs houdt hierop toezicht. Daarnaast rapporteert ze gevraagd en ongevraagd over de ontwikkelingen binnen het onderwijs, met als doel het onderwijs als geheel te verbeteren. Goed onderwijs is de ambitie; met ons toezicht willen we bijdragen aan continue kwaliteitsverbetering op alle scholen. In ons werk stellen we de leerling/student en het leerproces centraal. We willen alle scholen in Nederland stimuleren zichzelf te verbeteren; op het niveau van besturen en hun scholen en op het niveau van het stelsel. We gaan er daarbij vanuit dat verbeteringen de klas moeten bereiken. We werken vanuit verdiend vertrouwen, namens de overheid en samenleving. Beter onderwijs is het publieke belang dat wij centraal stellen in ons toezicht. Effectief zijn we, als we eraan bijdragen dat het onderwijs in Nederland beter wordt.*

*Uit: Meerjarenbeleidsplan Inspectie van het Onderwijs, 2015-2020, in: ONDERZOEKSKADER 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs*

in een algemeen toezichtkader. Het mogen beoordelen van kwaliteit en kwaliteit als centrale notie in onderwijsbeleid en -wetgeving geeft een ander gezicht aan het toezicht en het bijbehorende kader. Het impliceert volgens Peeters, Hofman en Frissen een meer inhoudelijke opvatting over wat goed onderwijs is (Peeters et al., 2013: 11). En een inhoudelijke opvatting over kwaliteit geeft ruimte voor normstelling op basis waarvan kan worden beoordeeld of er wel of niet sprake is van (voldoende) kwaliteit zoals ook Mertens aangeeft: 'De toename van heftige opvattingen, meningen over de kwaliteit van scholen riep om meer precisie en meer feiten. Hoe kunnen wij weten of scholen wel 'voldoende' goed zijn? Niet alleen uit wat we kunnen vernemen van de scholen zelf, maar externe toetsing en validering zou daarbij toch vanzelfsprekend moeten zijn?' (Mertens, 2014: 21). De media-analyse in het rapport van Frissen et al. over onderwijskwaliteit, uitgevoerd over de periode van 2002 tot 2014, laat zien dat er vanaf 2002 sprake is van een veranderend klimaat ten aanzien van de vraag hoe en waarover scholen verantwoording horen af te leggen, waarin er toenemende aandacht komt voor de manier waarop de overheid een rol heeft in het bewaken van onderwijskwaliteit. De terughoudendheid ten aanzien van overheidsbemoediging met het onderwijs wordt voor een belangrijk deel losgelaten en de schroom om scholen af te rekenen op hun resultaten wordt afgeworpen (Frissen et al., 2015: 29). Uit de media-analyse blijkt ook dat afstand is genomen van de eerder gedeelde opvatting dat scholen niet op hun resultaten mogen worden afgerekend. Het beeld dat zich vanaf 2002 ontwikkelt is er een van verantwoording afleggen en prestatiemeting.

Bulterman en De Muynck (2014) wijzen op een omkering van zaken die samenhangt met de wens resultaten van onderwijs te willen vatten in getallen. Volgens hen kan bij managers of beleidsmakers de indruk ontstaan dat men daarmee grip heeft op de kwaliteit van onderwijs. Maar het leidt tot een omkering van de relatie tussen doel en middelen: 'Toetsen zijn dan geen middelen meer om het leerproces te bevorderen. Het leerproces wordt een middel om de toetsresultaten te bevorderen (...) Zo dreigt toetsing losgemaakt te worden van vorming en telt alleen nog maar wat gemeten wordt.' (Bulterman&De Muynck, 2014: 133). Dit heeft gevolgen voor de inhoud van het onderwijs maar ook voor de rol van de leerkracht. De onderwijsprofessional die regie voert over wat er in het onderwijs gebeurt en zelf de regie heeft over de bewaking van de kwaliteit, maakt plaats voor een overheid die zich de rol toe-eigent van bewaker van wat er in het onderwijs moet gebeuren en beoordelaar van de kwaliteit van onderwijs. Daarmee worden leraren en onderwijsinstellingen vooral gezien als uitvoerders van het door de overheid vormgegeven beleid. De scholen mogen zich buigen over hoe de doelen van de overheid het beste kunnen worden gerealiseerd en zullen op hun resultaten door diezelfde overheid op basis van door de overheid te hanteren criteria en instrumenten worden beoordeeld. De inspectie kan vanuit haar twee hoofdtaken enerzijds 'controleren' op naleving van wet- en regelgeving en anderzijds 'evalueren' waaronder wordt verstaan het systematisch beoordelen van (bepaalde aspecten van) de kwaliteit. Timmer wijst er in zijn onderzoek op dat het in Nederland onduidelijk is geworden wie bepaalt wat kwaliteit is van onderwijs. En wat de formele status is van het toezichtkader van de Inspectie van het Onderwijs. Timmer stelt dat toezichtkaders niet bedoeld zijn als regels voor de scholen en hun onderwijsgeevenden maar als instructies voor toezichthoudende inspectie-teams. Daarmee zijn toezichtkaders niet normstellend volgens Timmer ook al lijken bepalingen in het toezichtkader gericht te zijn op de onderwijsgeevenden en de school (Timmer, 2011: 179) (zie ook Hoofdstuk 3). In de praktijk hebben de toezichtkaders wel degelijk een normstellend effect gehad.



### 5.3.4 Effect van toezicht gebaseerd op testen

Hoewel van diverse kanten steeds is geprobeerd de breedte van onderwijsdoelstellingen onder de aandacht te brengen en wordt gepoogd de ontwikkeling van kinderen in een breder perspectief te zien, zorgt de eindtoets, en de belangen die eraan gekoppeld zijn voor wat betreft de plaatsing in het voortgezet onderwijs, voor een stevige verankering van het belang van taal en rekenen in het basisonderwijs. Zolang plaatsing in het voortgezet onderwijs wordt gebaseerd op toetsing van een combinatie van (een beperkte set van) taal en rekenen, zullen deze onderwerpen veel aandacht krijgen in het onderwijs zelf, in het toezicht op het onderwijs en in de waardering van het onderwijs. Hoe de Inspectie van het Onderwijs naar scholen kijkt, heeft gevolgen voor het handelen van scholen. Emmelot, Karsten, Ledoux en Vermeulen stellen op basis van hun onderzoek naar het inspectietoezicht dat scholen de neiging hebben zich 'inspectiegericht' te gedragen (Emmelot et al., 2004: 18). Dit gebeurt omdat het oordeel van de inspectie als belangrijk wordt ervaren en men graag een positief oordeel wil krijgen. Ook Borghans, Schils, Van Langen en Leest hebben meer recent onderzoek gedaan naar de invloed die opbrengstindicatoren uitoefenen op het functioneren van scholen. Hoewel zij stellen dat het lastig te achterhalen is of aanpassing van het gedrag een gevolg is van de indicatoren, stellen zij wel dat de mate waarin scholen hun gedrag lijken aan te passen aan het toezichtkader, met name de gemiddelde score op de eindtoets, opvallend groot is. Uit de in het kader van hun onderzoek gebruikte enquête komt naar voren dat scholen 'veel doen om hun door de indicatoren gemeten resultaten een klein beetje te verbeteren'. Borghans et al. noemen hiervoor twee oorzaken, namelijk dat bij de beoordeling van de scholen erg veel afhangt van de gemiddelde score op de eindtoets en dat de invloed van toeval bij deze score vrij groot is. Dat maakt volgens Borghans et al. dat scholen het gevoel hebben dat zij overgeleverd zijn aan een score die geen recht doet aan hun feitelijke prestaties (Borghans et al., 2015: 63). Scholen gaan zich richten naar de criteria waarop ze beoordeeld worden. In lijn met 'Campbell's Law' wordt een indicator, als er veel waarde aan wordt toegekend, leidend voor het handelen van degenen die op basis van de indicator worden beoordeeld. En dat richten heeft gevolgen voor de waarde van de indicatoren. Ehren wijst op het gevaar van 'isomorphisme' dat gepaard kan gaan met een (te) grote nadruk op verantwoordingsstandaarden. Daarmee bedoelt ze dat de variëteit tussen scholen verloren gaat omdat scholen allemaal hun best doen om zo goed mogelijk aan de verantwoordingsstandaarden te willen voldoen. En isomorphisme draagt niet bij tot het verbeteren van onderwijs (Ehren, 2019: 11). Om bij te dragen aan verbetering van de kwaliteit van onderwijs is het volgens Ehren zaak te zoeken naar een combinatie van vertrouwen en 'intelligente' vormen van verantwoording (Ehren, 2019: 13).

### 5.4 Eindtoets en kwaliteit van onderwijs

De gedachte dat onderwijsresultaten meetbaar zijn, heeft in combinatie met de beschikbaarheid van tests ertoe geleid dat de overheid voor de nieuwe rol van sturen op afstand snel en eenvoudig de beschikking kan krijgen over tests die ingezet kunnen worden als indicatoren van opbrengsten van onderwijs en daarmee over de kwaliteit van onderwijs. Bestaande tests zijn daarbij als uitgangspunt genomen zonder eerst grondig onderzoek te doen naar de kwaliteit van de in te zetten instrumenten in de specifieke context en de waarde van beschikbare testresultaten in relatie tot de realisatie van beleidsambities en de mogelijke negatieve effecten ervan op alle bij het onderwijs betrokkenen.

### 5.4.1 Relatief meten en kwaliteit

In de verantwoording van 1984 geeft Cito aan dat de ambitie van De Groot om met een eindtoets een oordeel te kunnen geven over 'de stand van het onderwijs' (de evaluatieve functie), bij Cito 'nooit veel aandacht heeft gekregen' omdat de eindtoets volgens Cito, 'te beperkt van opzet is om te functioneren in een onderzoek naar de stand van het onderwijs met name omdat slechts een deel van het totale curriculum wordt bestreken en dan nog met alleen meerkeuzevragen' aldus Cito (Wijnstra, 1984: 10). Vanaf 1971 wordt wel meewerkt aan rapportages ten behoeve van de evaluatieve functie van de eindtoets, op verzoek van schoolbegeleidingsdiensten en het Bureau van Statistiek van de Gemeente Amsterdam (Wijnstra, 1984: 10). Uitgangspunt daarbij is het vergelijken van de resultaten van een school met de resultaten van scholen uit de regio en met het landelijk gemiddelde. Op grond van die laatste vergelijking worden scholen geïnclassificeerd van hoog naar laag: 'De laagste en hoogste 14% kregen als oordeel 'laag', respectievelijk 'hoog', de overige gemiddelden werden als 'gewoon' aangemerkt.' (Wijnstra, 1984:17). In de loop van de jaren blijkt de evaluatieve functie van de eindtoets door Cito te worden uitgebreid en verder te worden geprofessionaliseerd. De evaluatieve functie krijgt concreet gestalte in de evaluatierapporten die scholen standaard ontvangen. Scholen worden vergeleken op de verschillende onderdelen van de eindtoets. Ook hier is het landelijk gemiddelde het uitgangspunt en wordt visueel inzichtelijk gemaakt of en in hoeverre scholen op de verschillende toetsonderdelen 'afwijken' van die gemiddelde scores: 'Op het evaluatierapport wordt per doelstellingenrubriek, per toetsonderdeel en van de toets als geheel de positie van de school aangegeven in de verdeling van schoolgemiddelden.' (Uiterwijk&Theunissen, 1996: 19). Hiermee kan per school een 'profiel getekend worden van de relatief sterke en zwakke onderdelen' van een school (Uiterwijk&Theunissen, 1996: 19). Cito is er ook toe overgegaan gemiddelde standaardscores te rapporteren per regio en die te vergelijken met het landelijk gemiddelde (Uiterwijk&Theunissen, 1996: 20). Tegelijkertijd is Cito aarzelend over de interpretatie van toetsuitkomsten in relatie tot bepaling van onderwijskwaliteit. Voor het 25-jarig jubileum in 1993 organiseert Cito een conferentie rond het thema 'Telt alleen het resultaat?' (Cito, 1993b). Hier wijzen diverse sprekers op de beperkte informatie die leerresultaten bieden voor het kunnen vaststellen van kwaliteit van onderwijs (Cito, 1993b: 40, 57). In de verantwoording van 2010 spreekt Cito over het 'schoolrapport' als rapportage die gebruikt kan worden ter evaluatie van het onderwijs, omdat het inzicht geeft in 'wat de plaats is van de school in de landelijke verdeling van schoolgemiddelden en in de verdeling van de schoolgroep waartoe de school behoort' (Boxtel et al., 2010: 9). De rapportages op schoolniveau worden vanaf het begin gebaseerd op de gecombineerde scores van de individuele leerlingen. Voor het vergemakkelijken van de (jaarlijkse) vergelijkbaarheid is het gebruik van standaardscores ingevoerd door Cito (zie hoofdstuk 3). De manier van rapporteren, ondanks de zorgvuldig geformuleerde toelichting, heeft bijgedragen aan het idee dat een oordeel over een school en de kwaliteit van een school kunnen worden ontleend aan de gemiddelde scores van deelnemende leerlingen.

In het Actieplan 'Basis voor Presteren; naar een ambitieuze leercultuur voor alle leerlingen' van onderwijsminister Bijsterveldt-Vliegthart (Ministerie Onderwijs en Wetenschap, 2011a) stelt de minister dat de prestaties voor taal en rekenen in het primair onderwijs moeten verbeteren. In de memorie van toelichting bij de wijziging van de wet PO stelt de minister dat 'toetsing helpt om onderwijsverbetering mee aan te zwengelen en gaande te houden omdat toetsing duidelijk maakt waar nog verbetering te behalen is. Opbrengstgericht werken leidt aantoonbaar tot betere

prestaties'. Met de verplichting van de eindtoets worden 'geobjectieerde en genormeerde minimumleerresultaten in een wettelijke basis vastgelegd: de schoolscore wordt dan de basis voor de geobjectieerde opbrengstennorm van alle scholen.' Daarmee is de eindtoets een 'instrument voor risicogericht toezicht van de inspectie' ten behoeve van de 'externe verantwoording en beoordeling van de leerresultaten van scholen' (Ministerie OCW, 2012). Op basis van de scores van de eindtoets formuleert zij ambities voor het jaar 2015. Daarbij stelt zij onder andere dat de gemiddelde score op de Cito-toets per school omhoog moet. Deze ambities geven aan dat de minister niet voldoende op de hoogte is van het feit dat het gemiddelde een gecreëerd en gefixeerd gemiddelde is en dat de percentielen daarvan afgeleid zijn en dus niet door prestaties van scholen te beïnvloeden zijn, anders dan door verschuivingen in de rij waarbij de vooruitgang van de ene school ten koste gaat van de positie van de andere school in de rij.

Voor de vergelijkbaarheid van scores door de jaren heen 'vereist' het model de aanname dat de populatie die aan de eindtoets deelneemt jaarlijks niet substantieel verandert in het niveau van presteren, zo stelt Cito in de verantwoording van 2011: 'We gaan uit van de veronderstelling dat er in de loop van een gering aantal jaren geen dramatische veranderingen plaatsvinden in het landelijke algemene niveau van de leervorderingen.' (Van Boxtel et al., 2011: 18). De aanname dat het gemiddeld niveau in een groep constant is gebleven maakt dat Cito stelt dat de scores van twee eindtoetsen lineair met elkaar kunnen worden geëquivalereerd. Dit uitgangspunt maakt de eindtoets niet geschikt voor het vaststellen van voor- of achteruitgang in behaalde resultaten in termen van standaardscores, niet van individuen en niet van scholen. Scholen kunnen niet door de jaren heen worden gevolgd wat betreft hun gemiddelde standaardscores. Scholen kunnen zich wel binnen een cohort met elkaar vergelijken op behaalde standaardscores. De ambities van de overheid om te kunnen sturen op de opbrengsten van onderwijs in combinatie met de ongeschiktheid van de eindtoets van Cito voor uitspraken in die richting, hebben ervoor gezorgd dat de overheid bij de overname van de eindtoets heeft besloten de toets te verankeren in meer vaste criteria in de vorm van referentieniveaus voor taal en rekenen. Dit omdat de overheid wil kunnen beschikken over 'een maat om uiteindelijk de toegevoegde waarde/leerwinst van scholen te kunnen vaststellen.' (Ministerie OCW, 2012). Vanaf 2014 zijn leerstofonafhankelijke normen, in de vorm van referentieniveaus, aan de eindtoetsen toegevoegd.

#### 5.4.2 Referentieniveaus en kwaliteit

De meer recente opdracht om referentieniveaus in te voegen in de eindtoets heeft niet geleid tot fundamentele aanpassingen in de psychometrische uitgangspunten van de eindtoets. Het kunnen vaststellen van het al dan niet behalen van beheersingsstandaarden in de vorm van referentieniveaus is ingevoegd in het al bestaande psychometrisch model, gericht op het selectiedoeleinde. De primaire functie van de eindtoets is daarmee nog altijd het leveren van een 'objectieve gegevens bij overdracht po naar vo' (Ministerie OCW, 2012). In dat licht is het feit dat de scores voor taal en rekenen bij elkaar zijn opgeteld om zo de betrouwbaarheid van de toets als geheel te vergroten, begrijpelijk. Maar een gecombineerde score verliest aan waarde als het gaat om inzicht in de scores op de deoltoetsen. Dit inzicht is wel beoogd met de evaluatieve functie van de eindtoets als hulpmiddel voor zelfevaluatie en als externe verantwoording en beoordeling van leerresultaten van scholen. Een gecombineerde score levert weinig informatie ten behoeve van evaluatie en verbetering.

De invoeging van de referentieniveaus in de bestaande eindtoets heeft in de huidige manier van werken niet geleid tot een inhoudelijke norm waarmee resultaten kunnen worden vergeleken. In de praktijk is er namelijk een 'eis van groei' gekoppeld aan de te behalen scores. In de verantwoording van de eindtoets van 2019, 'Terugblik Centrale eindtoets 2019', wordt gesteld dat als een leerling referentieniveau 1F beheerst, de leerling dan taal en rekenen in voldoende mate beheerst (CvTE, 2019: 15). Dit is in lijn met de normen voor beheersing van de referentieniveaus zoals oorspronkelijk door de Expertgroep Doorlopende leerlijnen Taal en Rekenen bedacht. Deze Expertgroep stelde destijds voor Taal: 'Samenvattend: een school voor primair onderwijs functioneert op dit moment goed als tenminste 75% van de leerlingen op niveau 1F functioneert (liever nog meer leerlingen dan 75%); een deel van die leerlingen is al verder, en functioneert op het volgende niveau.' (Expertgroep, 2008: 26). Voor Rekenen komt de Expertgroep met de aanbeveling dat 'het percentage leerlingen dat minimaal het referentieniveau 1F behaalt, moet toenemen van 75% naar 85% en dat het percentage leerlingen dat minimaal het referentieniveau 1S behaalt moet toenemen van 50% naar 65%' (Expertgroep, 2008: 50). Uit de eerste rapportage van de Inspectie van het Onderwijs over de resultaten op de referentieniveaus, zoals gemeten in de centrale eindtoets van 2014/2015 en 2015/2016 blijken de resultaten voor het behalen van het drempelniveau 1F positief, zo stelt de Inspectie: 'Bij het opstellen van de referentieniveaus is de ambitie geformuleerd dat minimaal 85 procent van de leerlingen aan het einde van het basisonderwijs niveau 1F behaalt en minimaal 65 procent niveau 1S/2F. In 2015/2016 beheersen vrijwel alle leerlingen niveau 1F voor lezen en taalverzorging (respectievelijk 99% en 96%). Voor deze inhoudsgebieden wordt het ambitieniveau van 85 procent dus (ruimschoots) bereikt. Ook voor rekenen wordt het ambitieniveau van 85 procent bereikt, al beheersen daar minder leerlingen 1F (87%) dan bij lezen en taalverzorging. Kijken we hoeveel leerlingen het streefniveau (1S/2F) beheersen, dan zien we dat alleen bij lezen de ambitie van minimaal 65 procent wordt gehaald: 77 procent van de leerlingen beheerst het streefniveau. Voor taalverzorging en rekenen beheersen veel minder leerlingen dit niveau, respectievelijk 57 en slechts 44 procent.' (Inspectie van het Onderwijs, 2017: 9). Daarmee zijn de eerste resultaten op de leerstofonafhankelijke norm positiever dan aanvankelijk door de opstellers van de referentieniveaus gedacht.

Het toezichtkader biedt geen mogelijkheid voor tevredenheid over de behaalde resultaten. Daar wordt gesproken over ruimte voor verbetering. De uitdaging zit niet in het behalen van 1F voor het gros van de kinderen zoals met de invoering van de referentieniveaus de bedoeling was. Het doel blijkt nu zoveel mogelijk leerlingen naar de streefniveaus 2F/1S te krijgen. Alleen groei leidt tot een positieve beoordeling door de inspectie. De rekensom die de inspectie gaat hanteren voor het toezicht is geen absolute norm maar een formule die uitgaat van een continue verbetering van resultaten. Dit is niet in lijn met de curriculaire gedachte die aan het Referentiekader ten grondslag ligt: het ligt niet in de lijn der verwachtingen dat het gros van de kinderen 2F (Lezen, Taalverzorging) of 1S (rekenen) zal beheersen.

Kinderen verschillen in hun ontwikkeling, in tempo en in te behalen doelen. Het nieuwe toezichtskader gaat uit van groei en ruimte voor verbetering: 'Landelijk behaalt nog lang geen 65 procent van de leerlingen het referentieniveau 1S/2F' (Inspectie van het Onderwijs, 2020: 5). Hiermee heeft de inspectie de ambities naar boven bijgesteld. Daarmee werken minister en inspectie naar eigen zeggen aan een 'competitieve opbrengstcultuur'.

Ook de PO-raad komt op de site<sup>24</sup> met een handreiking hoe scholen 'ambitieuze en realistische schoolnormen' ten aanzien van de te behalen niveaus voor de komende jaren kunnen opstellen. Het vergelijken met andere scholen met een vergelijkbare schoolpopulatie wordt gestimuleerd in de handreiking om scholen zo uit te dagen het maximale uit hun leerlingen te halen. Het opstellen van eigen schoolnormen is een nadere uitwerking van het streven naar meer regie over onderwijskwaliteit waartoe de PO-raad schoolbesturen en teams uitdaagt. En zo is de eindtoets niet langer te bezien als een instrument om een 'voldoende' niveau van opbrengsten ten aanzien van de vaardigheden Lezen, Taalverzorging en Rekenen te realiseren, maar is de eindtoets een instrument om scholen aan te zetten tot competitie en een voortdurende verbetering van resultaten. Door uit te gaan van permanente ruimte voor verbetering, ligt permanente ontevredenheid over de resultaten op de loer. Binnen de kaders van de huidige eindtoets die het kunnen maken van onderscheid tussen leerlingen als primaire functie heeft, zal een toename van resultaten weer leiden tot het toevoegen van weer moeilijkere opgaven om het selectiedoeleinde te kunnen dienen. Door bij de invoering van referentieniveaus opnieuw een relatieve norm toe te voegen, is van een criteriumgerelateerde meting nauwelijks sprake meer. Opnieuw is er nu sprake van een relatieve norm voor de beoordeling van scholen ten opzichte van elkaar, weliswaar anders dan voorheen, maar wel opnieuw een norm op basis waarvan competitie wordt gestimuleerd. Daarmee blijft de eindtoets een test met een voor scholen en hun leerkrachten high stakes karakter. Voor de leerlingen zijn de belangen vooral gekoppeld aan de selectiefunctie en voor de scholen zijn de belangen vooral gekoppeld aan het steeds weer verbeteren van prestaties op de referentieniveaus. Met als tragiek dat als alle scholen hun prestaties op de referentieniveaus steeds maar verbeteren, dat geen gevolgen heeft voor de doorstroom naar vervolgonderwijs.

#### Inspectie van het onderwijs

Het oordeel over de basisvaardigheden (de standaard OR1 Resultaten) baseren we op de behaalde referentieniveaus op de eindtoets in de laatste drie schooljaren. Daarbij onderscheiden we twee indicatoren, waarin we de behaalde referentieniveaus voor lezen, taalverzorging en rekenen samen nemen:

1. Het percentage leerlingen dat aan het einde van het basisonderwijs het fundamentele niveau 1F haalt voor taal en rekenen. In principe zou elke leerling dit niveau aan het einde van de basisschool moeten beheersen.
2. Het percentage leerlingen dat aan het einde van het basisonderwijs het hogere niveau (streefniveau) 1S voor rekenen en 2F voortaalverzorging en lezen haalt. Het streven moet zijn dat zoveel mogelijk leerlingen dit niveau beheersen.

Voor alle leerlingen die in de laatste drie schooljaren een eindtoets hebben gemaakt, beschikt de school over de behaalde referentieniveaus (1F of 1S/2F) voor lezen, taalverzorging en rekenen. Wij kijken vervolgens welk percentage van de behaalde referentieniveaus in de afgelopen drie schooljaren op of boven 1F ligt en welk percentage op of boven 1S/2F ligt. We berekenen dus twee percentages, die zowel de referentieniveaus voor de drie vakgebieden als de drie schooljaren omvatten.

Uit: ONDERWIJSRESULTATENMODEL PO *Inspectie van het Onderwijs, 2020b: 4-5*

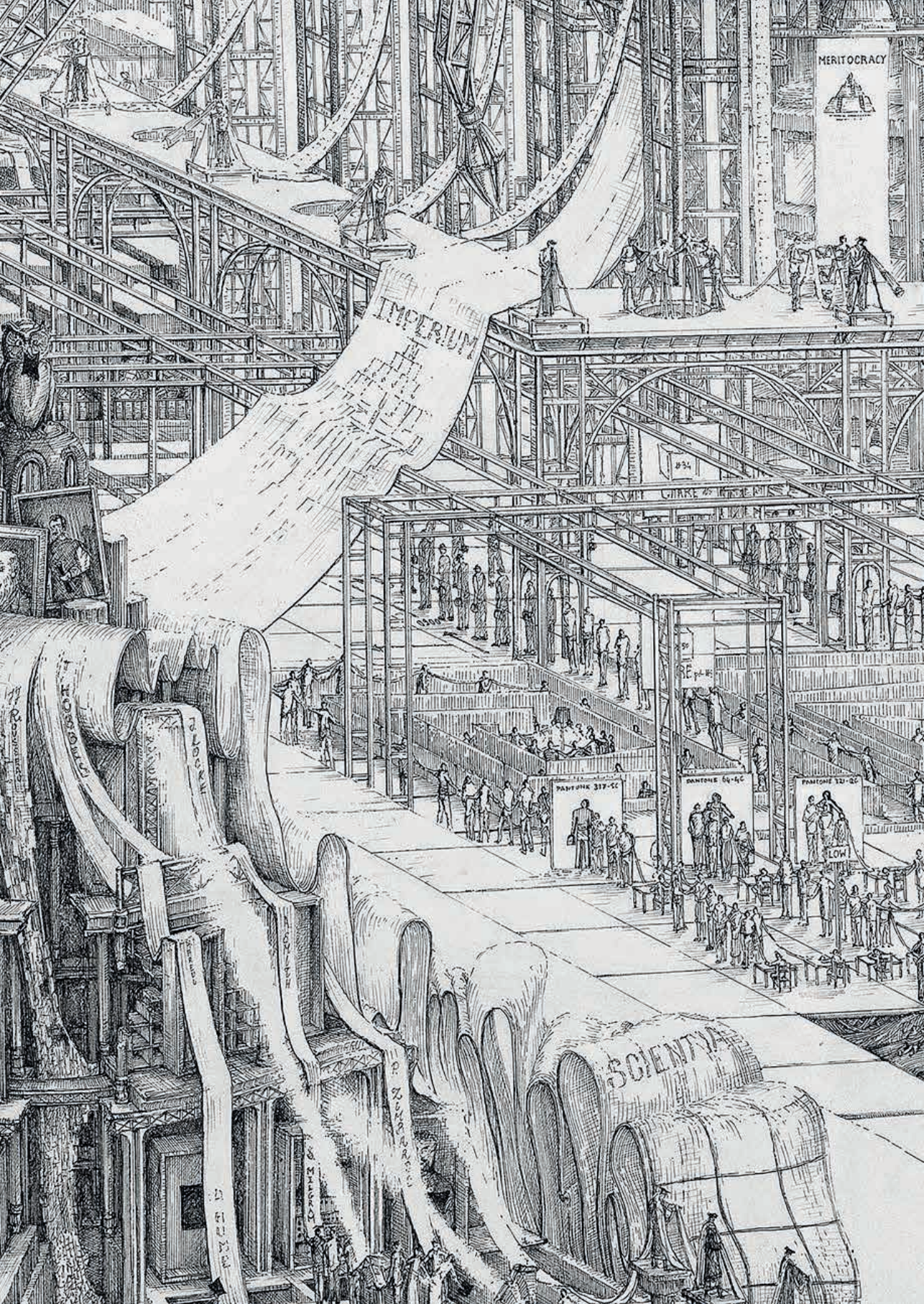
## 5.5 Samengevat

Met de verplichting tot het gebruik van een eindtoets in combinatie met een LVS is definitief een einde gekomen aan een tijdperk waarin de school ruimte en autonomie heeft om zich op basis van eigen onderwijskundige inzichten te verantwoorden over de kwaliteit van het gegeven onderwijs. Eigen keuzes en accenten zijn verdrongen door scores op gestandaardiseerde toetsen,

24] <https://www.poraad.nl/nieuws-en-achtergronden/zo-gaat-de-inspectie-straks-beoordelen-of-je-school-het-goed-doet>

opgelegd door beleid en toezicht, die vergelijkingen tussen scholen mogelijk maken. De overheid is met het willen hanteren van objectief vaststelbare indicatoren voor opbrengsten en kwaliteit met behulp van gestandaardiseerde toetsresultaten een nieuwe weg ingeslagen. Het toepassen van een relatieve norm op de prestaties van de scholen als geheel door scholen met elkaar te vergelijken op basis van hun eigen gemiddelde prestaties, en dat als uitgangspunt nemen voor het vaststellen van kwaliteit van onderwijs, heeft maar een zeer beperkte invulling aan het begrip 'onderwijskwaliteit' gegeven. In combinatie met het toezicht door de Inspectie van het Onderwijs heeft het er wel toe geleid dat scholen aangezet worden tot competitie waarbij de vraag of dit ook zal leiden tot verbetering van de kwaliteit van onderwijs zelf weliswaar reëel maar moeilijk te beantwoorden is geworden. De huidige manier van vergelijken, inhoudelijk beperkt tot cognitieve aspecten, en die weer beperkt tot een selectie van taal en rekenen, die met meerkeuzevragen kunnen worden bevraagd, biedt summier zicht op 'de staat van het onderwijs' in termen van de realisatie van de onderwijsinhoudelijke doelen zoals voorzien voor de drie domeinen kwalificatie, socialisatie en subjectificatie (Biesta, 2015: 20). De belangrijkste informatie die de eindtoets nu oplevert is dat de ene school hoger scoort dan de andere school. Daarmee is de evaluatieve waarde van de eindtoets beperkt. Die beperking geldt voor de overheid maar ook voor de scholen, hun leerlingen en hun ouders. Dit zelfde geldt voor de uitkomsten van de toetsen van het LVS; ook die bieden een zeer beperkt zicht op kwaliteit van onderwijs en geven nauwelijks handvatten voor de verbetering van onderwijs. Met de uitkomsten van eindtoets en LVS hebben zowel de scholen als de overheid maar zeer beperkt zicht op wat er in het onderwijs anders en beter kan. In de wetenschap dat de groep waaraan lesgegeven moet worden in de wijk waarin de school staat een gegeven is waar een school of een leerkracht niet veel aan kan veranderen, worden wel verwachtingen neergelegd in termen van continue verbetering van scores. Met grote gevolgen bij het niet realiseren ervan. De relatieve meting en bijbehorende indeling in categorieën werken wel als een vorm van verwachtingsmanagement voor de leerlingen en hun ouders ten aanzien van de uitkomsten van de eindtoets. Daarmee ontstaat steeds vroeger een vorm van etikettering. Die etikettering heeft gevolgen voor hoe de leerling naar zichzelf kijkt en hoe de omgeving naar het kind kijkt. De vraag naar de rechtvaardigheid van deze manier van meten voor het gevoel van eigenwaarde van winnaars en verliezers dringt zich op. Dit is onderwerp van onderzoek in hoofdstuk 8. Eerst wordt in het volgende hoofdstukken nader ingegaan op de achtergrond van dit competitieve model in relatie tot maatschappelijke ordening die er mede mee tot stand komt.





MERITOCRACY



IMPERIUM  
IN PARTE

AMSON

ELDER

SCIENTIA

PANTONE 337-CC

PANTONE 64-4-C

PANTONE 121-2-C

LOW

FIGURE

CONGRATULATE

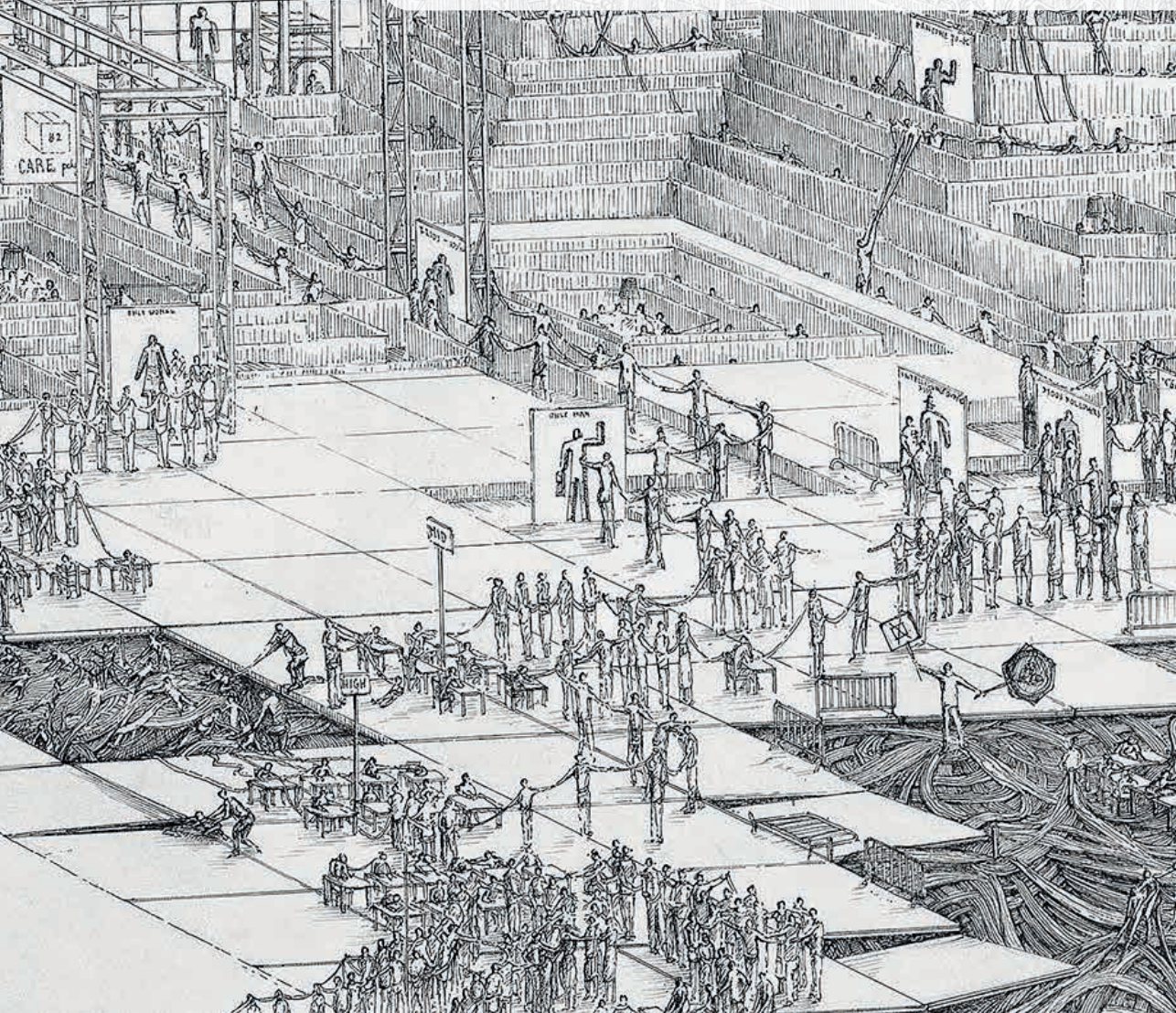
MICROSCOPIC

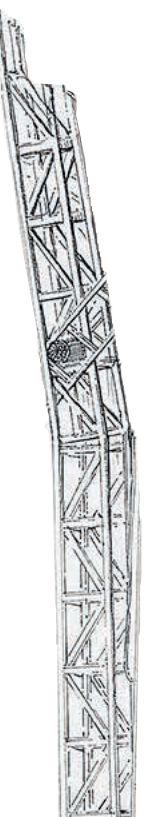




# HOOFDSTUK 6

Maatschappelijke ordening en testen





## Hoofdstuk 6 Maatschappelijke ordening en testen

*'Het individu wordt illusoire almacht aangepraat, terwijl in werkelijkheid zijn vrijheid almaar wordt ingeperkt. Zowel de illusie van het eerste, als het besef van het tweede, wakkert het onbehagen aan.'*

*Uit: Bas Heijne, 'Onbehagen' (2016)*

De totstandkoming en groei van de eindtoets in Nederland is geen autonoom proces geweest. Het past in bredere ontwikkelingen die de samenleving kenmerken en vormen zoals ook beschreven in de vorige hoofdstukken. Tests reflecteren keuzes die gemaakt worden over wat relevant geacht wordt te meten, hoe dat te operationaliseren en hoe daarmee te ordenen en te classificeren. Dat hangt samen met hoe samenlevingen zich willen vormen. Het streven naar gelijkheid en rechtvaardigheid heeft vanaf de negentiende eeuw het denken in westerse samenlevingen in hoge mate bepaald. Niet uitsluitend als na te streven, meer abstracte idealen, maar juist ook als uitgangspunten voor concreet handelen. Onderwijsdeelname wordt gezien als belangrijke factor om mee te dingen naar de belangrijke maatschappelijke posities. Tegen de achtergrond van een samenleving die volgens Amsing 'langzamerhand plaats maakte voor een samenleving waarin het beroep bepalend werd voor de sociale status en onder invloed van ideeën die voortkwamen uit de Verlichting bovendien de houding ten aanzien van sociale mobiliteit veranderde waarbij onderwijs werd gezien als middel om de kansen op sociale stijging te vergroten' (Amsing, 2002: 17). Dit vindt vooral weerklank aan de linkerkant van het politieke spectrum. Voor degenen die traditioneel de belangrijkste posities in handen hebben, wordt onderwijs de plek waar ze in hun posities bedreigd worden. De Vries spreekt in dit verband van 'schooldrang' en 'schooldwang': de drang van de lagere klassen die is wakker geroepen door de belofte van opwaartse sociale mobiliteit (De Vries, 1993: 57) en de dwang waaraan de hogere standen onderhevig zijn geraakt om niet verdrongen te worden door de opstomende lagere klassen (De Vries, 1993: 63). De offensieve onderwijsdeelname van de eerste groep en de defensieve onderwijsdeelname van de tweede groep zorgen volgens De Vries voor een ingebouwde klassenstrijd die de belangrijkste oorzaak is geworden van de voortgaande expansie van onderwijsdeelname (De Vries, 1993: 64). Het heeft de vraag naar onderwijs versterkt zonder opgelegd te hoeven worden: 'De verticale schooldwang die machtigen in de zestiende, zeventiende en achttiende eeuw uitoefenden over minder machtigen, is in de negentiende en twintigste eeuw geworden tot een zowel verticale als horizontale schooldwang die allen uitoefenen op allen' (De Vries, 1993: 68). De toename van onderwijsdeelname onder druk van maatschappelijke ontwikkelingen vraagt niet alleen om het opheffen van barrières voor onderwijsdeelname zodat ruimte kan worden gegeven aan 'schooldrang', maar ook om regie van de kant van de overheid om die 'schooldrang' in goede banen te leiden, daarbij rekening houdend met de belangen van de heersende elite. Kansengelijkheid ontwikkelt zich daarbij als begrip. Aan de hand van opeenvolgende, elkaar deels overlappende, beleidsperiodes wordt de ontwikkeling geschetst van een standenmaatschappij naar een samenleving op meritocratische leest geschoeid, in structuur en kennisdominantie (6.1). Vervolgens wordt de overgang van beleid in de periode na de Tweede Wereldoorlog naar de jaren zeventig onderzocht voor wat betreft de invulling van het uitgangspunt van meritocratie en hoe zich dat vertaald heeft in actief overheidsoptreden voor de samenleving als geheel en voor onderwijs in het bijzonder (6.2). De laatste decennia van de twintigste eeuw, waarbij veranderingen in economisch beleid, die in combinatie met de visie op mens en maatschappij bepalend

zijn geweest voor veranderingen in samenleving en onderwijs, worden vervolgens beschreven met daarbij een belangrijke rol voor tests en in het bijzonder voor de eindtoets en het LVS (6.3).

## 6.1 Maatschappelijke ordening en verdienste

De houdbaarheid van aristocratische uitgangspunten, waarbij afkomst en stand in hoge mate de levensloop bepalen, blijkt in de loop de 19e eeuw niet langer houdbaar. Er wordt gezocht naar een meer rechtvaardig principe voor verdelen. Het concept 'verdienste' komt meer tegemoet aan die gevoelens van rechtvaardigheid die al een lange historie in de samenleving hebben. Lang bestaande uitdrukkingen als 'je krijgt loon naar werken', 'boontje komt om zijn loontje', 'wie zaait zal oogsten' laten zien dat de waarde die wordt toegekend aan inspanningen en inzet ver teruggaat. Het goed gebruiken van talenten is een opdracht die in de christelijke traditie verankerd is. In de periode van de tanende macht van de aristocratie en de zoektocht naar een samenleving gebaseerd op meer rechtvaardige verdelingen dan afkomst en status is nadruk op een combinatie van begaafdheid en inzet als meer rechtvaardig indelingsprincipe dan ook een concept met een vertrouwde klank en een vertrouwde waarde. Maar tegelijkertijd kent het concept 'verdienste' geen concrete definiëring. En zeker ook geen neutrale definitie (Karabel, 2005: 3).

Afhankelijk van hoe het begrip gedefinieerd wordt, zullen sommige groepen er meer profijt van hebben dan andere. De definitie van 'verdienste' is fluïde en heeft de neiging de waarden en belangen te reflecteren van hen die de macht hebben die op te leggen. Karabel laat zien hoe dat heeft gewerkt aan de hand van de veranderingen die de toelating en uitsluiting van de drie elite-universiteiten Harvard, Yale en Princeton in de loop van de tijd hebben ondergaan (Karabel, 2005: 5).

Het vraagt onderzoek naar tijd en plaats om te zien welke invulling, waarde en belangen met het begrip verdienste samenhangen om te zien wat er feitelijk gewaardeerd wordt en welke belangen daarmee gediend worden. Bij het wegvallen van het denken over aanleg, geschiktheid en begaafdheid als een geschenk van god, is het begrip 'aptitude' volgens Bisseret vanaf de 18e eeuw samen gaan vallen met noties van gelijkheid, verdienste en individuele verantwoordelijkheid, als elementen van een egalitaire ideologie (Bisseret, 1979: 6). De egalitaire redenering van de bourgeoisie is in die tijd dat iedereen vrij is en voor de wet gelijk en dat de bestemming van mensen niet langer afhankelijk is van de gevestigde sociale orde maar gebaseerd wordt op individuele capaciteiten. Lang voor Darwin had de bourgeoisie volgens Bisseret daarmee een denken voor de rechtvaardiging van sociale ongelijkheden ontwikkeld gebaseerd op aannames over aangeboren geschiktheid en erfelijke karaktereigenschappen (Bisseret, 1979: 12). Intelligentie en studiegeschiktheid worden attributen van de kinderen van de bourgeoisie. Het begrip aptitude krijgt daarmee in de eerste helft van de 19e eeuw een connotatie van 'onveranderlijk', 'permanent', 'aangeboren' daarmee de bestemming van een individu bepalend (Bisseret, 1979: 13).

Bisseret stelt dat vanwege het ontbreken van een wetenschappelijke basis in de meeste psychologische theorieën het concept 'aptitude' zich niet verder heeft ontwikkeld met uitzondering van empirische praktijken van selectie bij onderwijs en arbeidsmarkt. Bisseret verklaart dit vanuit een sociale vraag naar deze toepassing waarbij de belangen buiten de wetenschap liggen (Bisseret, 1979: 7, 16). Hoe beter tests in staat blijken te zijn correcte voorspellingen te doen, hoe meer de werkelijkheid van wat gemeten is, de 'aptitudes', bewezen worden geacht (Bisseret, 1979: 18). Het begrip aptitude groeit in de wereld van tests uit tot een meetbare werkelijkheid waarmee de 'science of aptitudes' een onderbouwing vormt van de legitimatie van de sociale orde (Bisseret, 1979: 17).

Het denken gebaseerd op 'aptitude' als een individuele, vaststaande eigenschap, heeft gevolgen gehad voor het denken over kansengelijkheid in relatie tot onderwijs en de opvatting van kansengelijkheid die met onderwijs wordt nagestreefd. Door te stellen dat onderwijs gelijk is voor iedereen en dat onderwijs iedereen dezelfde kansen biedt, zijn de verschillen in individuele aanleg bepalend geworden voor succes. Het streven naar gelijkheid gaat volgens Frissen uit van een onderliggend ideologisch uitgangspunt dat ongelijkheid iets is wat in het kader van rechtvaardigheid van verdelingen tegengegaan moet en kan worden. Frissen spreekt in die context van een 'ideologie van de maakbaarheid' (Frissen, 2007: 56). Hij stelt dat in het ideaal van het streven naar gelijkheid een tweevoudige ambitie besloten ligt: het streven naar verheffing en een streven naar nivellering. Het streven naar verheffing is verbonden met het idee dat er sprake is van achterstand, gemis of tekort dat moet worden opgeheven maar met verheffing is ook het delen van het goede (leven) gemeind. Bij nivellering in relatie tot onderwijs gaat het met name om het herverdelen zodat de een niet veel meer heeft dan een ander (Frissen, 2007: 97). Als dit daadwerkelijk het geval is en gebeurt, blijft alleen dat nog over wat niet te nivelleren valt, namelijk intellectuele capaciteit en inzet. Die aspecten blijven dan over die er dan toe doen en op basis waarvan ongelijkheid tot stand zou mogen komen. Dit betekent volgens Frissen dat daarmee feitelijk alleen de school maatschappelijke ongelijkheid mag produceren. Omdat intellectuele capaciteit en inzet in hoge mate beïnvloedbaar zijn door overerving en opvoeding wordt ongelijkheid daarmee voor een belangrijk deel erfelijk. Dat betekent vervolgens dat de meest harde vorm van ongelijkheid – namelijk die welke niet te bestrijden is – als enige rechtvaardige vorm van ongelijkheid wordt gezien (Frissen, 2007: 59).

De acceptatie van verdienste als uitgangspunt en de meer concrete invulling die het begrip heeft gekregen in het Nederlandse onderwijs is op twee dimensies duidelijk zichtbaar geworden: in de structuur van het onderwijs die deels ontstaat en deels gestuurd wordt, en in de rol en concrete invulling die het concept 'intelligentie' krijgt in tests en in het onderwijs.

### 6.1.1 Stand en structuur

De Vijlder laat zien hoe in de 19e eeuw de vorming van de moderne, nationale economische ordening plaatsvindt 'met een bijna perfecte congruentie tussen institutionele structuren in het onderwijs en de economische productiestructuur', een dynamiek die vooral wordt veroorzaakt door pogingen om met de nationale economie de eigen natie op te stoten in de vaart der volkeren (De Vijlder, 1996: 14). In de praktijk leidt dat ertoe dat elke zich emanciperende groep zijn eigen onderwijssoort krijgt, zeker als dit een match heeft met een zich ontwikkelende economische behoefte. De Vijlder wijst erop dat het in eerste instantie vooral gaat om extensies van het primair onderwijs en de ontwikkeling en uitbreiding van het tertiair onderwijs (De Vijlder, 1996: 13). Halverwege de 19e eeuw wordt 'het gat' daartussen opgevuld met een concrete invulling van het tussenliggende onderwijs, het middelbaar onderwijs. In de formulering van Bartels: 'Dan omvat het tusschen beide inliggende gebied van het middelbaar onderwijs de vorming van die talrijke burgerij.' (Bartels, 1963: 6). De aristocratie behoudt daarnaast de Latijnse scholen en gymnasiën. Er is zo sprake van een naar standen gesegmenteerd onderwijssysteem. Onderwijs-historica Amsing spreekt van een onderwijs met 'parallele schooltypen die zowel verschilden in hun curriculum als in de sociale herkomst van de leerlingen. Binnen dit systeem kwamen leerlingen op basis van hun afkomst al op jonge leeftijd terecht in een bepaald segment van het onderwijs, waarna overstap naar een andere stroom moeilijk was.' (Amsing, 2002: 18). Het

onderwijsstelsel dat in de negentiende eeuw tot stand komt, biedt iedere groep de kans de mogelijkheden te benutten onderwijs te volgen passend binnen de eigen stand. Voor uitzonderlijk talent, in termen van inzet en intelligentie, biedt het stelsel tegelijkertijd kansen en mogelijkheden om hogerop te komen. Zo laat de emancipatie van verschillende groepen in de samenleving het stelsel in de praktijk meebewegen met de maatschappelijke ontwikkelingen. Dit vormt een verklaring voor het feit dat de veranderingen in de samenleving tot aan de invoering van de Mammoetwet 'opgevangen' kunnen worden met het onderwijsstelsel zoals dat in 1863 onder leiding van Thorbecke tot stand komt.

### 6.1.2 Intelligentie als dominant concept

De socioloog Matthijssen verklaart ontwikkelingen vanuit de idee dat een nieuwe dominantie altijd iets van een compromis in zich zal dragen met het verleden wat betekent dat het 'oude systeem zowel actief als passief in zal werken op het nieuwe regelsysteem' (Matthijssen, 1982: 102). De strijd om de macht van nieuwe en oude dominanties kan alleen gewonnen worden door bepaalde gedragsoriëntaties en zijnsinterpretaties van de oude intact te laten waarbij professionals van de oude macht daarnaast actief inwerken op het nieuwe regelsysteem door de eigen werkelijkheidsinterpretatie te vermommen in het nieuwe denken (Matthijssen, 1982: 102). Voor het onderwijs vormt het concept 'algemeen intellectuele vorming' die vermoming volgens Matthijssen. Door dit concept centraal te stellen als nieuwe kennisdefinitie blijft het oorspronkelijke onderscheid tussen theoretisch georiënteerd en praktisch georiënteerd onderwijs bestaan en behoudt het eerste de grootste waarde waarmee behoud van waarde van daarop gebaseerde vormen van onderwijs kunnen blijven bestaan: 'De theorie van de algemene intelligentie als maatstaf voor uitnemendheid houdt de wezenlijke kern van het aristocratische cultureel kapitaal staande, namelijk de verhevenheid van ideeën (theorie) boven de materiële werkelijkheid (praktijk).' (Matthijssen, 1982: 107). Het behoud van het belang van algemene intelligentie is volgens Matthijssen in die tijd niet rationeel omdat 'praktische bruikbaarheid' van scholing het centrale concept is in de tijd van Thorbecke en het veel meer voor de hand had gelegen als het realiseren van praktisch nut centraal was komen te staan in het onderwijs. Maar dit is niet gebeurd. Dit komt volgens Matthijssen omdat oude waarden terugkomen en daarmee een nieuwe 'mythe' realiseren die vanaf het begin is ingebouwd in het onderwijs als het nieuwe cultureel kapitaal: 'Het geloof dat de materiële werkelijkheid door leerlingen alleen doelmatig gekend kan worden via de omweg van de theorie.' (Matthijssen, 1982: 107). Er komt dus tegen de verwachting in geen onderwijsstelsel waarin praktische toepassingen centraal staan maar een waarin aan algemene vorming grote waarde wordt toegekend. Overgenomen vanuit de vermogenspsychologie die in de 19e eeuw hoogtij viert, wordt het trainen van het denkvermogen de hoogste prioriteit, waar men volgens Amsing wiskunde en Latijn het meest geschikt voor acht (Amsing, 2002: 36). Dit leidt tot een eenzijdige aandacht voor cognitie en de bijbehorende didactiek, die van het memoriseren (Amsing, 2002: 37). Intelligentie en bekwaamheid raken zo verweven met het behalen van succes in de samenleving wat voor een belangrijk deel bijdraagt aan de aantrekkingskracht van onderwijs als middel tot sociale stijging. Op die manier zijn volgens Matthijssen 'het ideaal van gelijke onderwijskansen en de meritocratische oriëntatie van de industriële samenleving volstrekt verenigbaar: uitbreiding van de onderwijskansen voor allen opent de bovenste lagen van de onderwijspiramide voor iedereen die de begaafdheid en de wil bezit om de vereiste vaardigheden te verwerven.' (Matthijssen, 1982: 121). De theorie van de algemene intelligentie is in de optiek van Matthijssen de basis van het cultureel kapitaal van het denken omstreeks 1900.

### 6.1.3 Meritocratie en samenleving

Als in 1958 het boek 'The Rise of the Meritocracy 1870-2033' van de Engelse socioloog en socialist Michael Young verschijnt, is de term 'meritocratie' nieuw, maar het achterliggende idee van verdienste als uitgangspunt voor het kunnen bereiken van maatschappelijke posities in de westerse wereld zeker niet. Young heeft het boek nadrukkelijk bedoeld als een satirische waarschuwing voor de gevolgen die hij in die tijd ziet bij samenlevingen die zich zijn gaan ordenen op basis van verdienste: 'There was nothing new in the proposition that  $IQ + effort = merit$ ; only in the way it was formulated.' (Young, 1994: xiii). Hij schetst dat er bij het vormgeven van een samenleving op basis van deze formule een periode zal zijn, de overgang, waarin getalenteerde kinderen uit de lagere sociale klassen kunnen opklimmen, maar dat die periode van op- en neerwaartse mobiliteit plaats zal maken voor juist een 'gesloten' samenleving, een gelaagde samenleving waarin maatschappelijke posities van generatie op generatie worden overgedragen, indirect via het IQ. Uiteindelijk, in 2034, zal het volgens Young leiden tot een opstand aangevoerd door hen die niet gehoord en gezien zijn met een crisis als gevolg. Hoewel Young meritocratie neerzet als een 'dystopie', als een waarschuwing aan samenlevingen die verdienste centraal stellen in hun maatschappelijke ordening, blijkt het begrip in de praktijk van samenlevingen zo aantrekkelijk dat het eerder is gaan fungeren als een 'utopie'.

Het meritocratisch concept heeft in de loop van de twintigste eeuw een invulling gekregen met als belangrijk uitgangspunt dat ontwikkelingsmogelijkheden voor iedereen toegankelijk moeten zijn en maatschappelijk succes afhankelijk moet zijn van de combinatie van talent en inspanning. Aan onderwijs wordt de taak toebedeeld de selectie van talent mogelijk te maken. Het onderwijs krijgt wat in sociologische termen de allocatie- en selectiefunctie heet: 'Prestaties in het onderwijs worden een maat voor geschiktheid voor maatschappelijke posities.' (Dekkers, 1999: 2). Het streven naar verdienste als uitgangspunt van sociale ordening en het bepalen van verdienste blijken goed samen te gaan met de meer economische invalshoek van het neoliberalisme in de jaren tachtig en negentig. De overheid wordt geacht de werking van de markt te faciliteren en geen belemmeringen op te leggen of zich te mengen in marktmechanismen. Het zorgen voor voldoende, goed opgeleid arbeidspotentieel past daarbij. Tonkens stelt dat het denken over accountability niet los kan worden gezien van de opkomst van het marktdenken binnen de publieke sector maar dat dit ook hoort bij een samenleving waar meritocratisering een prominente plek inneemt. Want inherent aan meritocratie is dat prestaties met elkaar moeten worden vergeleken zodat 'het mogelijk is te bepalen wat ieders relatieve verdienste is, wie dus de beste prestaties levert en wie daarom recht heeft op welke positie en beloning. Alleen dan kan ieder krijgen wat hij verdient en kan de juiste vrouw/man op de juiste plaats terecht komen. Meten moet neutraal en objectief, openbaar en controleerbaar zijn.' (Tonkens, 2008: 83). De rol van onderwijs is daarbij cruciaal omdat daar op basis van selectie en het verkrijgen van diploma's een eerste, belangrijke sortering voor de arbeidsmarkt plaatsvindt.

Volgens Van Pinxteren en De Beer bestaat het beeld dat een meritocratische samenleving een 'open' samenleving biedt, waarin iedereen gelijke (start)kansen heeft op maatschappelijk succes (Van Pinxteren & De Beer, 2016: 9). Zij wijzen erop dat er aanvankelijk bij de overgang van de traditionele klassensamenleving naar een meritocratische samenleving sprake zal zijn van een omvangrijke op- en neerwaartse sociale mobiliteit, omdat getalenteerde kinderen uit lagere sociale klassen de kans krijgen om op te klimmen op de ladder, terwijl 'domme' kinderen uit de hogere

klassen een stap omlaag moeten zetten. Zij stellen dat de aantrekkelijkheid van een meritocratische samenleving als na te streven ideaal zit in de combinatie van twee belangrijke aspecten van meritocratie: doelmatigheid en rechtvaardigheid. Een meritocratische samenleving wordt gezien als doelmatig omdat de talenten en capaciteiten van burgers optimaal worden benut. En als rechtvaardig omdat iedereen de mogelijkheid heeft om zijn/haar talenten te ontwikkelen en die op de (arbeids)markt te gelde te maken. Door verdienste te definiëren als de combinatie van intelligentie en inzet is het begrip goed te operationaliseren. Omdat onderwijs zich al bijna van nature aandient als de plaats waar intelligentie en inzet, al op jonge leeftijd, kunnen worden vastgesteld, heeft dat de doelmatigheid van het concept verder versterkt. Maar, stellen Van Pinxteren en De Beer, als eenmaal het proces van herordenen achter de rug is, treedt een nieuwe, anders geordende maar sterk gelaagde samenleving op waarin maatschappelijke posities van generatie op generatie worden overgedragen, indirect via het IQ. De sociale klasse waarin iemand terecht komt wordt meer en meer erfelijk (Van Pinxteren & De Beer, 2016: 10/11).

Bij het niet behoren tot het deel dat in de prijzen valt, kan in de perceptie van de verliezers een gevoel van uitsluiting ontstaan. Zeker omdat, volgens filosoof Koenis, mensen zich meer dan vroeger zijn gaan zien als 'bijna-gelijken' die denken dat ze in dezelfde wedstrijd spelen maar tegelijkertijd er de pest over in hebben dat ze de bal nauwelijks krijgen toegespeeld. Koenis stelt dat het heel wel mogelijk is dat de samenleving aan de ene kant steeds meritocratischer wordt, waardoor 'de kloof' tussen hoger en lager opgeleiden in termen van kansenongelijkheid groter wordt, terwijl aan de andere kant de hiërarchische verschillen tussen hoog- en laagopgeleiden afnemen waardoor mensen meer dan voorheen het gevoel hebben in dezelfde wedstrijd te spelen. Hij stelt dat de samenleving precies deze twee tendensen laat zien: niet de verschillen tussen hoog en laag zelf, maar de toenemende vergelijkbaarheid in combinatie met de toenemende verschillen in kansen zorgt voor boosheid, afgunst, rancune en frustratie (Koenis, 2016: 146). Koenis stelt dat kansengelijkheid in een meritocratische ideologie vooral betekent dat er gelijke kansen zijn op ongelijkheid. Voor allochtone kinderen leidt dat tot een dubbel probleem, omdat zij te maken hebben met nieuwe én oude vormen van ongelijkheid: 'Ze hebben niet alleen te maken met de harde markt van verdienste en talent waarin iedereen tegenwoordig zijn weg moet vinden. Daarbovenop worden zij ook nog belemmerd door die oudere vorm van ongelijkheid die verband houdt met hun afkomst.' (Koenis, 2016: 104).

Swierstra en Tonkens stellen dat verdienste niet simpel bepaald kan worden door te meten wat iemand tot stand brengt. Verdienste kan vooral inzichtelijk worden gemaakt in vergelijking met wat anderen presteren. Iedereen wordt vergeleken met iedereen in een universele competitie en samenleven is een wedren (Swierstra & Tonkens, 2008: 13). Een meritocratie vereist daarom een levendige meetcultuur. Het is noodzakelijk dat prestaties regelmatig worden vastgesteld en vergeleken. Dat vaststellen en vergelijken moet met regelmaat gebeuren om het presteren op het hoogste niveau te stimuleren en te controleren. Een meritocratie tendeeft in de optiek van Swierstra en Tonkens 'naar een doorgemeten en van wantrouwen doortrokken samenleving waarin niet alleen individuen maar ook instellingen en bedrijven met elkaar in voortdurende competitie zijn verwickeld. Meten en rangschikken van verdiensten versterken elkaar voortdurend. Voltooiing van de meritocratie vereist verfijning van meetsystemen.' (Swierstra & Tonkens, 2008: 136). Swierstra en Tonkens stellen dat dit meten transparant en objectief moet gebeuren. En de eis van transparantie en objectiviteit beperkt in wat gemeten kan worden. Daarmee biedt



transparantie volgens Swierstra en Tonkens geen neutraal doorkijkje op kwaliteit, maar selecteert actief wat als kwaliteit geldt, met eenvormigheid tot gevolg (Swierstra & Tonkens, 2008: 129). Het gebruik van een en dezelfde meetlat brengt met zich mee dat individuen, organisaties, praktijken zich aan moeten passen aan die meetlat, omdat ze alleen in termen van die meetlat bestaansrecht hebben (Swierstra & Tonkens, 2008: 127). Daarmee verdwijnt diversiteit en hebben afwijkingen geen bestaansrecht: 'Organisaties worden dus door de eis van transparantie tot eenheidsworst gemaakt.' (Swierstra&Tonkens, 2008: 128).

Of en in hoeverre een samenleving meritocratische hangt volgens Van Pinxteren en De Beer samen met het belang van IQ voor schoolprestaties en maatschappelijk succes: neemt dat belang toe en neemt het belang van andere factoren – in ieder geval in verhouding tot het IQ – af, dan is er sprake van meritocratisering (Van Pinxteren & De Beer, 2016: 13). Zij plaatsen belangrijke kanttekeningen bij de mate waarin zowel IQ als inspanning geheel aan het individu zijn toe te schrijven. Beide zijn maar beperkt te zien als verdienste van een individu. Vuyk wijst er op dat degenen die het begrip meritocratie verdedigen het centrale begrip 'verdiensite' niet goed doordenken. Want in de praktijk, wordt volgens Vuyk namelijk niet verdiensite beloond, maar vooral het bezit van talent, wat niet of nauwelijks samenhangt met hoe dat talent in de praktijk wordt ingezet. Omdat IQ voor een belangrijk deel door geboorte wordt doorgegeven stelt Vuyk dat 'we afstevenen op de terugkeer van een samenleving met aristocratische trekken' (Vuyk, 2017: 250). In de ontwikkeling van een samenleving waarin verdiensite wordt gezien als belangrijkste bepaler van maatschappelijk succes, heeft onderwijs als verdeler van kansen aan belang gewonnen.

Veel sociologisch, psychologisch en economisch onderzoek wijst volgens Mijs op de niet vervulde belofte van meritocratie in onderwijs. Mijs stelt dat er op basis van literatuur belangrijke lessen te leren zijn ten aanzien van meritocratie in onderwijs. Onderwijsinstellingen verstoren in de praktijk de ideale meritocratische processen door onder andere de streaming in onderwijsniveaus en de differentiatie binnen scholen en vakken, gebaseerd op docentoordeelen of (gestandaardiseerde) tests en selectieprocedures. Kansen gebaseerd op verdiensten zijn daarbij in hoge mate niet zelf 'verdiend'. En talenten zijn bij de start niet gelijk verdeeld. Daarmee start de meritocratische race niet voor iedereen gelijk. Hoe het begrip 'verdiensite' concreet wordt ingevuld, bevoordeelt sommige groepen en benadeelt juist andere groepen (Mijs, 2016:14).

Verhaeghe stelt dat de combinatie van 'vrijheid' en 'loon naar werken' aan meritocratie een nieuwe aantrekkingskracht hebben gegeven. De versmelting van beide elementen noemt Verhaeghe 'neoliberale meritocratie', die volgens hem gelijk een kantelpunt heeft gecreëerd. De oorzaak van de kanteling is volgens Verhaeghe dubbel: enerzijds is het een illusie te denken dat iedereen met gelijke kansen aan de start verschijnt en anderzijds zal het systeem na verloop van tijd een nieuwe elite installeren, die de deur zorgvuldig sluit voor wie na haar komt (Verhaeghe, 2012: 120). Dat leidt tot het stilvallen van de sociale mobiliteit en een groeiende kloof tussen onder- en bovengroep. Een nieuwe ongelijkheid ontstaat op deze manier met gevolgen voor de winnaars en verliezers.

In Nederland is het streven naar een samenleving geordend op basis van meritocratische gedachten uitgangspunt voor (onderwijs)beleid direct verbonden met het streven naar kansengelijkheid en de rol van onderwijs daarin.

## 6.2 Kansengelijkheid en onderwijs

In een brief aan de Tweede Kamer van maart 2019 geven de ministers van onderwijs aan dat zij kansengelijkheid in het onderwijs willen bevorderen en zich zorgen maken over de bevindingen van de Inspectie van het Onderwijs waaruit blijkt dat 'kinderen van ouders met een lager inkomen en minder diploma's lang niet altijd in staat worden gesteld het niveau te bereiken dat zij cognitief aankunnen' (OCW, 2019: 1). In die brief definiëren de ministers de bevordering van kansengelijkheid als volgt: 'Bij de bevordering van kansengelijkheid gaat het erom dat jongeren in staat worden gesteld – en worden gemotiveerd en waar nodig ondersteund – om hun talenten te ontwikkelen. Het schoolsucces van kinderen zou enkel afhankelijk moeten zijn van hun capaciteiten en inzet en niet van het inkomen of het opleidingsniveau van hun ouders.' (OCW, 2019: 1). Het denken over kansengelijkheid heeft zich nauw verbonden met meritocratische uitgangspunten. Onder de voorwaarde van meritocratie kan kansengelijkheid gerealiseerd worden. En daar waar wel capaciteiten en inzet als uitgangspunt genomen worden en geen sprake is van kansengelijkheid, is iets mis. De kansengelijkheid wordt daarbij afgemeten aan scores op toetsen voor vergelijkbare groepen. Deze nog altijd gangbare manier van interpreteren van kansengelijkheid vindt zijn oorsprong voor een belangrijk deel in de Verenigde Staten.

Sandel reconstrueert de ontwikkeling van meritocratie in de Verenigde Staten en laat zien dat het denken over onderwijs en meritocratie in hoge mate beïnvloed is door de president van Harvard in de jaren veertig, James Conant. Conant wilde een klasseloze samenleving waarin gelijke kansen zouden worden gerealiseerd middels onderwijs gebaseerd op meritocratische uitgangspunten, waarbij de besten de meeste kansen zouden krijgen op de hoogste vormen van onderwijs. Voor het onderwijs zag hij de taak om de 'abilities' vast te stellen, talent te ontwikkelen en ambities in goede banen te leiden. Scholen zouden de leerlingen selecteren en ordenen voor doorstroom naar hoger onderwijs waarmee de scholen de vijver zijn waaruit een nieuwe, meritocratische elite gevist wordt (Sandel, 2020: 159). Terugkijkend stelt Sandel dat er twee belangrijke kenmerken in een meritocratisch onderwijssysteem besloten zijn die bezwaarlijk zijn: het eerste is dat een meritocratische ordening weliswaar geïmplementeerd kan worden tegenover een erfelijk systeem, maar dat een meritocratische ordening zeker niet bedoeld is als middel tegen ongelijkheid. Het tweede punt is dat het een systeem is waarbij het belonen van de besten het omlaag halen van de rest betekent (Sandel, 2020: 161). De toelating tot hoger onderwijs is in de Verenigde Staten in hoge mate afhankelijk van de score op een intelligentietest, de SAT, geïmplementeerd als een 'aptitude test' waarvan bekend is dat de scores hierop in hoge mate correleren met het inkomen van de ouders (Sandel, 2020: 164). Dit is voor een belangrijk deel te verklaren door het feit dat een goede score op de SAT te trainen valt en ouders met hoge inkomens hierin investeren. Meritocratische ordening is dus geenszins bedoeld geweest om verschillen te verkleinen en kansengelijkheid te vergroten maar veel meer om uiteindelijke verschillen in de uitkomsten van de ordening te legitimeren, in de ogen van de winnaars maar ook in die van de verliezers.

In 1966 wijst Coleman in een paper dat onderdeel is van de samenvattingen van een grootschalig onderzoek dat in de Verenigde Staten onder zijn leiding heeft plaatsgevonden, op de veranderingen die zich daar dan voordoen bij de interpretatie van het begrip 'equality of educational opportunity'. Aanvankelijk was er in de Verenigde Staten geen aandacht voor het fenomeen onderwijskansen; het laten opgroeien van kinderen was de zorg van een familie. Aandacht hiervoor komt met de start van openbaar onderwijs: onderwijskansen worden direct verbonden met het

meer algemene concept van gelijkheid van kansen (Coleman, 1966: 5). Het concept van gelijke onderwijskansen heeft aanvankelijk betrekking op het bieden van gelijke leermogelijkheden voor allen los van het individu en zijn resultaten, een inputgerichte benadering. Dit verandert nadat in 1954 het Supreme Court de uitspraak doet dat rassenscheiding op scholen in strijd is met de grondwet van de Verenigde Staten. Die uitspraak gaat volgens Coleman verder dan alleen het regelen van de toegang; het concept van gelijke onderwijskansen wordt verbonden met het concept van onderwijsresultaten. Vanaf dat moment gaat gelijkheid van onderwijskansen over een actieve verantwoordelijkheid voor een onderwijskundig programma dat moet zorgen dat individuele handicaps die samenhangen met de omgeving van de leerling moeten worden overwonnen. Dit moet door de school aangetoond worden met resultaten (Coleman, 1966: 20). Het onderzoek van Coleman was het eerste grote survey in de Verenigde Staten naar onderwijs en de factoren die bepalend blijken voor succes voor de verschillende groepen leerlingen (zie ook hoofdstuk 8).

Deze ontwikkelingen positioneert Jacobs als ontwikkelingen van het begrip 'fairness'. In eerste instantie ligt de aandacht bij de vraag of de wedstrijd 'fair' is. Jacobs stelt dat het hierbij gaat om wat hij 'procedurele fairness' noemt, waarmee bedoeld wordt dat iedereen die deelneemt aan het onderwijs mee kan doen aan de wedstrijd. In de jaren zestig wordt daar in de Verenigde Staten het concept 'background fairness' aan toegevoegd. Op het moment dat het Supreme Court bepaalt dat het onderwijs voor blank en zwart even toegankelijk moet zijn, komt er aandacht voor de uitkomsten van het onderwijs: zijn de uitkomsten van onderwijs gelijk, ongeacht de achtergrond van de onderwijsdeelnemers? Het onderzoek van Coleman heeft volgens Jacobs gevolgen voor hoe het concept van gelijkheid van onderwijskansen vanaf die tijd gepercipieerd wordt: 'It crystallized the concept of equality of educational opportunity as being fundamentally about how education provides competitive ladders of opportunity and enables upward mobility for socially disadvantaged students.' (Jacobs, 2016: 319).

In landen waar er van oudsher een sterk gesegregeerd onderwijssysteem was, gebaseerd op standen van de samenleving, zoals in het Verenigd Koninkrijk, blijft kansengelijkheid lange tijd geen issue; er is onderwijs om te voldoen aan de behoefte van training voor de lager opgeleiden en er is onderwijs gericht op de individuele ontplooiing van het kind van de midden- of hogere klasse en gezamenlijk zorgt het geheel van onderwijs voor het in stand houden van de bestaande sociale orde. Alle onderwijsveranderingen die zich in de tweede helft van de vorige eeuw in Groot-Brittannië hebben voorgedaan, hebben weinig effect gehad op de sociale mobiliteit, zoals blijkt uit onderzoek van Bukodi en Goldthorpe (2018). Zij komen tot de conclusie dat mannen en vrouwen geboren in de jaren zeventig en tachtig net zoveel kans hebben gehad een andere maatschappelijke positie te bereiken als zij die in de jaren veertig geboren zijn (Bukodi & Goldthorpe, 2018: 207). De totale mobiliteit laat volgens deze onderzoekers gedurende een lange termijn stabiliteit zien. Dit geldt ook voor de relatieve mobiliteit. Ook die is stabiel (Goldthorpe, 2016: 89). Bukodi en Goldthorpe spreken over sociale mobiliteit in Groot-Brittannië als een 'zero-sum game' waarmee ze bedoelen dat elke verbetering van de relatieve mobiliteitskansen van kinderen van de minder bevoorrechte klasse alleen kan plaats vinden ten koste van de kansen van kinderen van de meer bevoorrechte klasse, die zich hier zo goed mogelijk tegen zal verzetten (Bukodi & Goldthorpe, 2018: 215). Bukodi & Goldthorpe komen tot de slotsom dat we af moeten van het idee dat onderwijs 'the great equalizer' is en dat een op onderwijs gebaseerde meritocratie

een wenselijke status is (Bukodi & Goldthorpe, 2018: 219). Het is volgens hen in het licht van alle historische bewijzen misleidend dat onderwijsbeleid zich bezig blijft houden met het streven naar meer gelijke mobiliteitskansen (Bukodi & Goldthorpe, 2018: 223). De belangrijkste bron van ongelijkheid van onderwijskansen is volgens Goldthorpe de ongelijkheid van omstandigheden die er is tussen kinderen van verschillende afkomst (Goldthorpe, 2016: 107). Kijkend naar de historische gegevens is het volgens Goldthorpe dan een ondoenlijke en ook unfaire last om van het onderwijs te verwachten dat daar de ongelijkheid wordt opgelost (Goldthorpe, 2016: 107).

Veel westerse landen hebben in de tweede helft van de vorige eeuw gekozen voor een onderwijsstelsel waarbij de toewijzing aan vormen van onderwijs, tracks, wordt uitgesteld (zie hoofdstuk 3). De gedachte hierbij was dat uitstel van selectie de onderwijskansen van kinderen van lager opgeleiden zou vergroten. Landenvergelijkend onderzoek in 21 Europese landen naar de invloed van uitgestelde selectie door socioloog Van de Werfhorst laat zien dat het verband tussen de vooropleiding van ouders en die van hun kinderen nauwelijks veranderd is door uitstel van selectie. Wel concludeert Van de Werfhorst dat vroeger selecterende onderwijssystemen nadeliger zijn voor de kansen van kinderen van lager opgeleiden (Van de Werfhorst, 2019). Het gebruik van tests heeft volgens Van Pinxteren en De Beer in het onderwijs het effect van de opleiding van ouders op de opleiding van het kind nauwelijks veranderd (Van Pinxteren & De Beer, 2016: 41). Zij stellen dat, uitgaande van onderwijs als een positioneel goed, de stratificatiepatronen nog even sterk zijn als vroeger (Van Pinxteren & De Beer, 2016: 47).

In de Verenigde Staten heeft het tijdperk van meritocratie niet gefungeerd als een motor voor sociale mobiliteit. Een kind van een familie die wat inkomen betreft behoort tot de één procent rijkste families, heeft 77 keer meer kans tot een van de top drie universiteiten toegelaten te worden (Sandel, 2020: 165). De kans op toelating tot die drie universiteiten is voor een kind van ouders uit de werkende klasse nog even groot als in 1954 (Karabel, 2006: 547). Markovits spreekt over meritocratie als een 'sham', een schijnvertoning. Meritocratie zorgt voor ongelijkheid. En die meritocratische ongelijkheid is schadelijk, zowel voor kansen als voor uitkomsten van het overgrote deel van de bevolking. Want ook voor de middenklasse in de Verenigde Staten geldt dat de ongelijkheid toeneemt en dat meritocratie geen belofte voor hen inhoudt. De afname van kansen die zich aftekenen voor deze groep treft ook hun kinderen die nauwelijks meer kansen hebben op toelating tot de elitescholen dan kinderen uit arme gezinnen (Markovits: 2019: 24). Markovits wijst ook op de schadelijke gevolgen van de meritocratie voor de elite: 'Meritocracy entices an anxious and inauthentic elite into a pitiless, lifelong contest to secure income and status through its own excessive industry.' (Markovits, 2019: X). Daarbij worden de middenklasse en de elite uit elkaar gedreven. Meritocratie is volgens Markovits een mechanisme geworden waarmee rijkdom en privileges van generatie op generatie worden overgedragen, precies dat wat de bedoeling was tegen te gaan (Markovits, 2019: X).

Sandel stelt dat vormen van meritocratische selectie verbonden zijn geraakt met oordelen over 'verdienen' en 'niet verdienen'. Meritocratische selectie is onvermijdelijk een publiekelijke veroordeling geworden van welke talenten en prestaties waardevol genoeg zijn om erkend en herkend te worden (Sandel, 2020: 173). Kansengelijkheid is volgens Sandel ooit bedoeld als een morele correctie op onrechtvaardigheid. Maar was daarbij bedoeld als een remediërend principe, en niet als een adequaat ideaal voor een goede samenleving. Het streven naar sociale mobiliteit is

geen antwoord op kansengelijkheid: 'A good society cannot be premised only on the promise of escape.' (Sandel, 2020: 224).

### 6.2.1 Kansengelijkheid en onderwijs in Nederland

In de jaren na de Tweede Wereldoorlog krijgt het concept 'kansengelijkheid' in relatie tot onderwijs eerst vooral betekenis in beleid gericht op het opheffen van barrières en belemmeringen, zodat iedere leerling de kans krijgt aan het onderwijs deel te nemen mits voldoende geschikt. Dit wordt met name ingegeven vanuit de zoektocht naar talent. Dit betekent dat het voor slimme, hardwerkende leerlingen mogelijk moet zijn deel te nemen aan die vormen van onderwijs die voorheen vooral zijn voorbehouden aan de hoogste standen. Beleid wordt gericht op juridische en administratieve maatregelen om de toegankelijkheid van het onderwijs te garanderen voor iedereen. Tot de jaren zeventig is het debat over kansengelijkheid en onderwijs volgens Matthijssen vooral te typeren als een zoektocht naar aanpassingen waarbij het onderwijs zelf gezien wordt als een afspiegeling van de maatschappelijke orde en de maatschappelijke orde een gegeven. Het realiseren van gelijke kansen heeft aanvankelijk daardoor vooral betrekking op a) de toelating tot onderwijs, b) selectie op basis van rationele/cognitieve capaciteiten en c) het bieden van de meeste kansen aan de meest begaafden (Matthijssen, 1971:442).

De democratiseringsgolf van de jaren zestig streeft naar een meer egalitaire samenleving waarin maatschappelijke posities worden bekleed door de meest competente individuen, ongeacht sekse, geloof en afkomst. Dit streven voegt zich eenvoudig in het gedachtegoed van een meer meritocratisch geordende samenleving. Onderwijs wordt cruciaal in deze ontwikkeling. Belemmeringen en barrières voor de onderwijsdeelname van meisjes waren voor een belangrijk deel al voor de democratiseringsgolf geslecht. Die deelname is al voor de invoering van de Mammoetwet bijna gelijk aan de onderwijsdeelname van jongens. Voor de ULO is dat in de jaren vijftig al het geval maar voor het v.h.m.o. volgt dat later; in 1966 stroomt 21% van de jongens in deze vorm van onderwijs in en 19% van de meisjes (Ministerie Onderwijs en Wetenschappen, 1968: 13). Voor kinderen uit arbeidersgezinnen blijkt de instroom in het v.h.m.o. in diezelfde periode wel sterk achter te blijven bij de kinderen van de hogere maatschappelijke regionen: 'Waar 85 procent van de twaalfjarigen uit de hoogste maatschappelijke regionen doorstroomde naar het v.h.m.o., kwamen arbeiderskinderen niet verder dan vier à vijf procent' (Mellink, 2014: 137). In het licht van de democratisering gaat deze beperkte doorstroom naar vervolgonderwijs van grote delen van de bevolking leiden tot een besef van onrecht. Met een stelselwijziging in de vorm van de invoering van de Mammoetwet was, zoals De Rooy het stelt, 'oppervlakkig gezien het Nederlandse onderwijsbeleid na de oorlog gedemocratiseerd, in die zin dat alom lippendienst werd bewezen aan de gedachte dat ieder kind het onderwijsgebouw zou mogen betreden en naar zijn of haar vermogen opklimmen naar de verschillende verdiepingen. Dit was goed voor de samenleving en het was rechtvaardig voor de betrokken leerlingen. Het onderwijs was dus niet langer naar stand ingericht, maar naar verdienste.' (De Rooy, 2018: 183).

Geleidelijk krijgt het denken over gelijke kansen en de rol van de overheid daarbij een ander karakter. Het handelen van de overheid moet niet alleen passief zijn maar er wordt actief ingrijpen nodig geacht. De politiek van gelijke kansen zoals die in de tweede helft van de 20e eeuw gestalte krijgt, is een politiek van 'verheffing'. Mensen die belemmeringen ondervinden die niet aan henzelf te wijten zijn, moeten de kans krijgen hun talenten te ontplooiën. Sociale hinder-

nissen zijn er om op te ruimen, zodat iedereen gelijke kansen krijgt. Zo stelt staatssecretaris Grosheide in een lezing in 1971: 'Het motief is elke leerling, elke student alle kansen geven. Dit is dringend nodig. Want iedereen weet, dat slechts hij in deze maatschappij slaagt die goed onderwijs heeft genoten. Het is daarom onze plicht, om iedere leerling van het kleuteronderwijs af alle ontwikkelingsmogelijkheden te bieden.' (Grosheide, 1971: 23). Om achterstandsgroepen kansen op gelijkwaardige deelname te kunnen bieden, is meer nodig dan het opheffen van belemmeringen. De overheid dient de maatschappelijke verandering en dient dat op een manier te doen die velen bij het beleid betreft (Meijnen & Van Wieringen, 2000: 32). Deze verandering in denken leidt tot experimenten en initiatieven op lokaal en schoolniveau, aangemoedigd door een overheid die onderwijs ziet als belangrijke hefboom voor de bevordering van kansen van kinderen uit achterstandsposities. Er komen compensatieprogramma's, een aanpak overgewaaid uit de Verenigde Staten, activeringsprogramma's en speciale projecten om achterstanden weg te werken. Uit onderzoek is duidelijk gebleken dat achterstanden en verschillen tussen kinderen uit lagere en hogere sociale milieus al vroeg worden opgelopen en dat onderwijs in de lagerschoolfase moet zijn gericht op het wegwerken van achterstanden om kansengelijkheid te kunnen bevorderen, het 'onderwijsstimuleringsbeleid' genoemd. Het aanvankelijke enthousiasme voor het voeren van onderwijsachterstandenbeleid wordt al snel getemperd. Het blijkt uitermate lastig om achterstanden effectief weg te werken en te zorgen voor gelijke kansen bij de start van het vervolgonderwijs. Dit maakt dat de aandacht verschuift naar voorschoolse educatie: 'hoe jonger hoe effectiever' is daarbij het motto, waarmee het probleem feitelijk buiten de scope van het primair onderwijs geplaatst wordt. Politiek gezien brokkelt het draagvlak af voor een vorm van stimulering omdat de gelden voor het onderwijsachterstandenbeleid niet zozeer blijken te worden besteed aan het wegwerken van achterstanden van de kinderen van laagopgeleide ouders, maar vooral aan extra aandacht voor onderwijs aan allochtone leerlingen wat de solidariteit voor dit beleid ondermijnt, zeker onder invloed van de meer populistische winden die opsteken. Er ontstaat breder een meer algehele twijfel of een overheid wel in staat is om zulke grote maatschappelijke problemen als onderwijsachterstanden aan te pakken, of het wenselijk is en tot welke prijs. Daarmee wordt kansengelijkheid meer een opdracht aan het individu, ingepast in de trend van individualisering die zich vanaf de jaren zestig aftekent.

### 6.2.2 Individualisering

Mellink positioneert de democratiseringsgolf in Nederland vooral in het perspectief van het streven naar de vrijheid van het individu dat met de die golf gepaard gaat en waar dan nog onvoldoende ruimte voor is: 'Dat kinderen uit lagere sociale milieus geen eerlijke kans kregen om zich een weg naar boven te banen, werd onder invloed van het individuele vrijheidsideaal een ernstig maatschappelijk onrecht. Hoe kon een onderwijssysteem dat zich tot doel stelde om vrije, verantwoordelijke en zelfstandige burgers af te leveren, het zich permitteren om niet iedereen daartoe een eerlijke kans te geven?' (Mellink, 2014: 136). Het economisch belang van de zoektocht naar talent gaat zo naadloos samen met de meer sociaal maatschappelijke behoefte aan rechtvaardigheid om allen de kans te bieden op het onderwijs dat het beste bij hen past. Bij dat laatste past dat ook het onderwijs zelf ter discussie wordt gesteld. Wat kan er aan het onderwijs gebeuren om de aantrekkelijkheid ervan voor de leerlingen van nu te vergroten? De aanval tegen de uitval en het zittenblijven wordt nu benaderd vanuit het perspectief van de leerling en diens vermeende behoefte aan onderwijs dat het best passend voor hem is. In de jaren zeventig wordt het van belang geacht dat 'het onderwijsleerproces zich naar de mogelijkheden van de

leerlingen zal moeten richten, zowel naar niveau als naar tempo' (Ministerie O&W, 1975: 49). Daar hoort bij de 'noodzaak het onderwijs meer te doen aansluiten bij de individuele verschillen in belangstelling, begaafdheid, sociale ervaring en ontwikkelingsfase' (Ministerie O&W, 1975: 15). Het leerstof-jaarklassensysteem voldoet niet goed omdat 'in zo'n opzet met individuele verschillen weinig rekening kan worden gehouden' (Ministerie O&W, 1975: 15). Het moet er niet langer om gaan leerlingen te vergelijken met uniforme schoolstandaarden maar 'vast te stellen waartoe de leerlingen in staat zijn en hen naar een verwezenlijking van hun mogelijkheden te begeleiden' (Ministerie O&W, 1975: 49). Van Kemenade wil daarbij nadrukkelijk streven naar wat hij noemt 'interne differentiatie', waarmee hij bedoelt dat leerlingen les krijgen op het niveau en in het tempo dat bij hen past. Hij heeft hierbij ook oog voor het gevaar van 'streaming', het opdelen in homogene groepen, dat daarbij op de loer ligt: 'In hoeverre zich in de loop van de middenschool 'setting' en 'streaming' zullen gaan voordoen, blijft de vraag. In ieder geval moet gepoogd worden deze verschijnselen zo lang mogelijk uit te stellen.' (Ministerie O&W, 1975: 53).

De jaren tachtig laten zich karakteriseren als een overgang van onderwijsbeleid, van een overheid die zich inspant om actief bij te dragen aan meer maatschappelijke gelijkheid naar een overheid die het meer overlaat aan het individu zich te emanciperen en zich een weg naar boven te banen. Het 'geloof' in het meritocratisch gedachtegoed in combinatie met een afnemend vertrouwen in de rol die onderwijs kan vervullen in het vormen van een meer egalitaire samenleving, leidt tot meer nadruk op de rol van het individu. Na jaren van investeren in onderwijs om onderwijsachterstanden weg te werken, laat de overheid het aan het individu zich te emanciperen en zich een weg naar boven te banen. Emancipatie gaat daarbij volgens Mellink steeds vaker zelfemancipatie betekenen. Emancipatie is volgens Mellink geworden tot een opdracht aan het individu dat in een vrije, reeds geëmancipeerde samenleving zijn eigen vrijheid en zelfstandigheid moet grijpen (Mellink, 2014: 229). En emancipatie wordt niet langer gezien als een collectief proces dat de gehele samenleving raakt maar wordt een eigenschap die aan individuele burgers wordt toegekend (Mellink, 2014: 230). Het concept gelijke kansen leidt volgens De Rooy 'niet alleen tot een individualisering, maar impliceerde een competitie tussen leerlingen'. Een competitie waarbij deelnemers zelf verantwoordelijk worden geacht voor het bereiken van hun plek in de samenleving. Daarmee zijn succes en falen geïndividualiseerd. De Vries wijst op de statusonzekerheden die hierachter schuilgaan. De angst om zelf te mislukken vermindert daarbij de morele solidariteit met de verliezers en scheidt psychische afstand in plaats van identificatie (De Vries, 1993: 159). De Rooy vat de keerzijde van het streven in deze periode naar meer 'gelijkheid' samen door te stellen 'dat dit kon betekenen dat de 'vrijheid' toenam om de minderbegaafden achter te laten' (De Rooy, 2018: 229).

Individualisering wordt aan het eind van de twintigste eeuw een breder na te streven doel. Politici zetten zich aan de taak structuren van de samenleving aan het individu aan te passen. Die focus op individuen maakt volgens Vuyk dat de meritocratie niet ziet wat ze doet met de samenleving als geheel (Vuyk, 2019:60). Eigen verantwoordelijkheid, ondernemerschap en zelfbeschikking worden de richtinggevendende idealen in de jaren negentig. De leefsituatie van intelligente en/of goed opgeleide mensen wordt zo tot model voor alle burgers. Maar, vraagt Vuyk zich af, 'is deze norm van individualiteit wel zo nastrevenswaardig?' (Vuyk, 2017: 272). De huidige politieke uitwerking van de idee van gelijkheid, de politiek van gelijke kansen, veronderstelt bij mensen kwaliteiten die ze niet bezitten en dat maakt hen, volgens Vuyk, uiteindelijk meer ongelijk en ongelukkig (Vuyk, 2019: 157).

De individualisering heeft daarbij tot gevolg dat niemand meer ergens vanzelf bij hoort. Ergens bij horen moet in onze samenleving worden verdiend. Je moet dus prestaties leveren, op alle mogelijke terreinen. Dat maakt onze samenleving buitensporig competitief. De paradoxale uitkomst is dat ondanks al het individualisme de samenleving steeds ongelijker wordt. Typierend voor de huidige verhoudingen is volgens Vuyk dat deze winnaars geen band meer hebben met de losers (Vuyk, 2019: 104). Voor met name nieuwkomers in onze samenleving is dit een extra hindernis.

### 6.2.3 Nieuwe doelgroepen

In de jaren zeventig gaan de ambities verder dan alleen de 'openstelling van onderwijs voor iedereen mits geschikt'. De overheid dicht zichzelf een actieve rol toe in het vergroten van de kansen voor kinderen uit achterstandsposities. Uit onderzoek is duidelijk gebleken dat achterstanden en verschillen tussen kinderen uit lagere en hogere sociale milieus al vroeg worden opgelopen en dat onderwijs bij uitstek in de gelegenheid is zich in de lagerschoolfase te richten op het wegwerken van achterstanden om kansgelijkheid te kunnen bevorderen, het onderwijsstimuleringsbeleid genoemd. Er wordt veel geëxperimenteerd en geld geïnvesteerd in tal van initiatieven op lokaal en schoolniveau, aangemoedigd door een overheid die onderwijs ziet als belangrijke hefboom voor de bevordering van kansen van kinderen uit achterstandsposities die niet langer als onveranderlijk worden beschouwd. Evenmin als intelligentie. Met de juiste aanpak is intelligentie in hoge mate ontwikkelbaar gebleken (zie hoofdstuk 8). De klassikale school met gelijk onderwijs voor iedereen komt onder druk omdat dergelijk onderwijs geen oog zou hebben voor de ongelijke kwaliteiten en behoeften van kinderen; ongelijke behandeling om gelijke kansen te bevorderen. De overheid voelt zich geroepen om met actief ingrijpen de kansen van kinderen uit achterstandsgroepen te vergroten. In eerste instantie is dit 'onderwijsstimuleringsbeleid' gericht op autochtone kinderen uit de arbeidersklasse. Door de toestroom van kinderen van 'gastarbeiders' die niet te gast zijn maar hier permanent blijken te blijven, staat de overheid voor de opgave hun middels onderwijs in het kader van een goede participatie en emancipatie kansen te bieden. Dit beleid krijgt een eigen naam: het 'onderwijsachterstandenbeleid'. Mellink wijst op het feit dat de principes van individuele vrijheid en verantwoordelijkheid, die in de jaren zeventig komen bovendrijven als algemene waarden in het onderwijs, van meet af aan als uitgangspunt gelden voor de opname van buitenlanders in het onderwijs (Mellink, 2014: 172). Opvallend is daarbij volgens Mellink dat de 'sturende werking van de politiek in de middelschooldiscussie werd afgewezen' maar dat 'zij in relatie tot nieuwkomers zo blijmoedig werd omarmd' (Mellink, 2014: 172). Culturele minderheden, zoals de term in die tijd was, zijn voorwerp van overheidszorg; ze moeten worden opgevangen en op de Nederlandse samenleving worden voorbereid. Emancipatie van nieuwkomers wordt daarmee een middel om migranten door Nederlanders gewenst gedrag voor te schrijven (Mellink, 2014: 183). Er is ruimte voor behoud van de eigen identiteit maar alleen als die identiteit strookt met de diepere waarden van de geïndividualiseerde samenleving (Mellink, 2014: 184).

In een studie op verzoek van de Onderwijsraad over de toegankelijkheid van het Nederlandse onderwijs stelt Van Dyck dat 'het op achterstandsgroepen gerichte toegankelijkheidsbeleid, achteraf gezien, over de gehele linie als weinig specifiek en niet-inhoudelijk blijkt te karakteriseren'. Zij wijst erop dat vanaf de jaren zeventig het streven naar 'gelijke kansen ongeacht achtergrond, herkomst of ascriptieve persoonskenmerken als etniciteit of sekse' als meritocratisch ijkpunt aan het toegankelijkheidsbeleid is toegevoegd (Van Dyck, 1997: 28). Als criterium voor selectie en



toewijzing stelt Van Dyck dat afkomst plaats heeft gemaakt voor verdienste, waarbij zij stelt dat verdienste te zien is 'in termen van geschiktheid en in termen van prestaties tijdens de onderwijs- en beroepsloopbaan'. Dit meritocratisch principe is 'stevig verankerd in onderwijsbeleid en -wetgeving. Het onderwijs staat daarbij voor de opdracht leerlingen zo goed als mogelijk kansen te bieden hun geschiktheid in prestaties tot uitdrukking te laten komen.' (Van Dyck, 1997: 30). Maar tegelijkertijd blijkt er bezorgdheid over de toegankelijkheid van het onderwijs die wel in strikte zin gewaarborgd is, maar in de praktijk moeilijk is te realiseren; veel kinderen krijgen niet wat hun op grond van specifieke behoeften toekomt. Van Dyck ziet barrières voor bijvoorbeeld allochtone kinderen en hun ouders die bijdragen aan het ontstaan en verergeren, c.q. niet tijdig inlopen van achterstand. De verschillen bij de start van het onderwijs 'stellen veel schoolteams voor bijkans onmogelijke opgaven' (Van Dyck, 1997: 82).

### 6.3 Meritocratie en onderwijs in de praktijk

De overgang van een standensamenleving naar een ordening op meer meritocratische leest betekent voor het onderwijs dat niet langer op voorhand vastligt wie welke vorm van onderwijs volgt. Omdat verdienste daarvoor bepalend wordt, heeft onderwijs daarmee een nieuwe functie gekregen, die van plaatsbepaler. Zoals Idenburg in 1960 stelt: 'De school selecteert. Bij het wegvallen van de traditionele criteria der maatschappelijke selectie, zoals afkomst en bezit, heeft zij de taak van de beheersing der maatschappelijke opstijging op zich genomen.' (Idenburg, 1960: 511). Diploma's veranderen van karakter. Het zijn volgens De Vries niet langer alleen bewijzen dat de bezitters ervan iets nuttigs hebben geleerd, maar het zijn in de eerste plaats graadmeters van aanvankelijke bekwaamheden; intelligentie, doorzettingsvermogen, sociale vaardigheden. Deze bekwaamheden beloven volgens De Vries productiviteit op het werk en daarom worden diploma's door werkgevers gevraagd en gehonoreerd. Onderwijs fungeert als doorlichtingsinstrument doordat het individuen met verschillende bekwaamheden rangschikt en daarmee informatie verschaft aan de kopers van arbeidskracht (De Vries, 1993: 13). Dit betekent dat de inhoud van het doorlopen curriculum voor het te verrichten werk niet relevant hoeft te zijn. Het verklaart volgens De Vries ook waarom algemeen vormende opleidingen zoveel waard zijn voor een markt waar ze niet voor opleiden, de arbeidsmarkt, en waarom algemeen vormende schooltypen beroepsopleidingen weten te deklasseren (De Vries, 1993: 47). Het is steeds moeilijker geworden om zonder diploma's toegang te krijgen tot de arbeidsmarkt of om op te klimmen binnen de arbeidsmarkt, zeker waar het de meer aantrekkelijke posities betreft. Hierbij komt dat een diploma volgens De Vries de meeste waarde heeft als het bezit ervan beperkt is. Het omslagpunt ligt volgens hem rond de vijftien procent, daarbij Martin Trow citerend die stelt dat dit het omslagpunt is van elite-onderwijs naar massa-onderwijs (De Vries, 1993: 48). Dit sluit aan bij de theorie van Hirsch over 'positionele goederen'. Hirsch stelt dat de waarde ervan mede afhankelijk is van wat de rest ervan heeft (Hirsch, 1977). Voor onderwijs betekent het dat de wedstrijd om de beste onderwijsplekken, in het kader van kansengelijkheid formeel opengesteld voor allen, maar door een beperkte groep moet kunnen worden gewonnen om een diploma van waarde te kunnen laten zijn. Daarbij staan twee opties open: of de toelatingseisen moeten worden verhoogd waarmee de selectie wordt verscherpt of de onderwijskwalificaties moeten worden opgestapeld zodat degenen die er het meeste van hebben, de meeste kansen maken. De Vries spreekt van een paradox: Het is in het belang van allen om onderwijsexpansie te beteuigen, maar ieder heeft er individueel belang bij om daaraan niet mee te doen (De Vries, 1993: 72). Vuyk typeert het huidige onderwijs als de 'meest basale' van de competities die start op vierja-

rige leeftijd en waarbij op 12-jarige leeftijd 'de bokken van de schapen worden gescheiden'. Het winnen zelf komt volgens Vuyk daarbij voorop te staan (Vuyk, 2017: 167). De winnaars mogen, aldus Vuyk, door naar schooltypen die hen in het hoger onderwijs zullen brengen, waarvan het diploma tegenwoordig een minimumvoorwaarde is om na het onderwijs toegelaten te worden tot de competities voor de meer profijtelijke functies en machtiger posities. De afvallers moeten met minder genoegen nemen, maar ook voor hen houdt de wedstrijd niet op. Zij zullen nog heel vaak hun talenten kunnen, maar vooral ook moeten bewijzen (Vuyk, 2017: 168).

### 6.3.1 Meritocratie als wedstrijd

Het beschouwen van onderwijs als een competitie van winnaars en verliezers roept de vraag op wie de spelregels bepaalt en in wiens voordeel de spelregels zijn. Omdat een concrete, breed gedragen invulling van het begrip 'verdienste' niet bestaat, is het niet altijd duidelijk welke criteria er bij een competitie gelden. Een selectiesystematiek kent parameters die op basis van bewust, menselijk handelen worden afgesteld. Hierbij worden keuzes gemaakt en zijn belangen in het geding. Keuzes worden door maatschappelijke ontwikkelingen beïnvloed en door de heersende orde bewaakt.

In de Verenigde Staten bijvoorbeeld heeft de democratisering geleid tot druk op de toelating tot de meest prestigieuze universiteiten voor de toelating van anderen dan leden van de 'WASP' (White, Anglo-Saxon, Protestant men) zoals vrouwelijke, Joodse en Afro-Amerikaanse studenten. Karabel laat zien dat een continuering van de uitsluiting van deze groepen de ondermijning van de legitimiteit van deze universiteiten tot gevolg zou hebben gehad. Dit heeft ertoe geleid dat deze universiteiten zich hebben moeten openstellen voor de getalenteerden uit opkomende sociale groepen. Dit is echter wel zodanig gebeurd dat de belangen van de van oudsher dominante sociale groepen niet in het geding zijn gekomen. Zo is de toelating verruimd en is ras of geloof niet meer bepalend voor het al dan niet toelaten, maar is in 2005 de kans dat een kind van laagopgeleide ouders een plek krijgt binnen een van de drie elite-universiteiten in de VS nog dezelfde als in 1954 (Karabel, 2005: 547). De manier waarop 'verdienste' bij de toelating wordt gedefinieerd maakt het op het eerste gezicht open voor iedereen, maar in de praktijk blijken vooral de jonge mannen en vrouwen kans te maken die uit families komen die de kinderen hebben laten opgroeien met het type cultureel kapitaal dat impliciet vereist is voor toelating (Karabel, 2005: 549). Karabel verklaart dit op basis van het feit dat bestaande elites die macht hebben te definiëren hoe 'verdienste' ingevuld wordt daar altijd zelf meer van hebben en degenen die hier grotere bronnen van hebben – cultureel, economisch en sociaal – zullen ervoor zorgen dat hun kinderen via het onderwijs als 'more meritorious' tevoorschijn komen (Karabel, 2005: 550). Hoewel dit 'quasi-meritocratische' systeem volgens Karabel te beschouwen is als een grote verbetering ten opzichte van het meer openlijke discriminerende en erfelijk systeem van daarvoor, zijn kinderen van laagopgeleiden ook nu bij voorbaat kansloos.

Op dit fenomeen wijst ook Vuyk voor Nederland. Hij stelt dat uit de kloof die ontstaat tussen hoger en lager opgeleiden blijkt dat de samenleving verdeeld raakt langs de lijnen van intelligentie. Die verdeling betreft volgens Vuyk meer aspecten dan alleen het inkomen. Intelligentie en opleiding hebben ook effect op zaken als gezondheid, levensverwachting, politieke invloed, culturele participatie en het gevoel greep te hebben op het eigen leven. Die bijkomende effecten maken de kloof groter en moeilijker te overbruggen (Vuyk, 2019: 61). Bovens ziet in de huidige 'diplo-

mademocratie', zoals hij de kloof noemt, een relatie met de mate van politieke betrokkenheid van burgers. Hij stelt dat burgers beduidend meer politieke invloed hebben naarmate hun opleiding hoger is en dat dit feit de samenleving heeft veranderd: 'In de moderne informatiemaatschappij zijn kennis en informatie de belangrijkste kapitaalgoederen en bestaat de politieke oligarchie niet meer uit adel, patriciaat of industriëlen, maar uit hoogopgeleide professionals.' (Bovens, 2006).

Swierstra en Tonkens wijzen met betrekking tot de problemen die samenhangen met het uitgangspunt van verdienste voor de selectie voor onderwijs en arbeidsmarkt op het feit dat dit in de praktijk een relatief begrip blijkt te zijn. 'Zij wordt niet simpel bepaald door te meten wat iemand tot stand brengt, maar door die prestatie vervolgens te vergelijken met de prestaties van anderen. Omdat presteren wordt beloond met macht, status en inkomen, is het voor iedereen van belang beter te presteren dan de omgeving.' (Swierstra & Tonkens, 2008: 13). Daarmee doet meritocratie een aanslag op het zelfrespect van de verliezers, van degenen die laag op de ladder eindigen: 'De belangrijkste keerzijde van de meritocratie is dat het zelfrespect van de verliezers systematisch wordt ondermijnd.' (Swierstra & Tonkens, 2008: 62). Verliezen staat in het meritocratisch gedachtegoed gelijk aan een gebrek aan IQ en/of inzet. De bal ligt dus geheel bij het individu als hij/zij onderaan de maatschappelijke ladder terecht is gekomen. Dit effect is het grootst als het vergelijken eerlijk gebeurt omdat dan de waardering van en voor de winnaars groot is. Meritocratie leidt op deze manier volgens Swierstra en Tonkens tot een wedstrijd waarbij iedereen er steeds toe aangezet wordt beter te presteren dan de anderen uit de competitie, op basis van een eerlijke wedstrijd met een eerlijke, onpartijdige, controleerbare scheidsrechter (Swierstra & Tonkens, 2008: 119). Daarmee leidt het meritocratisch ideaal tot een gerichtheid op diploma's, met gevolgen voor de classificatie van burgers en hun levenskansen.

Het meritocratisch denken tezamen met de moeizame ervaringen van onderwijs als 'kansmaschine' hebben het debat over kansengelijkheid in Nederland in een bepaalde richting gestuurd. Het gaat er ten aanzien van kansengelijkheid in een meritocratische context tegenwoordig vooral om dat iedereen de kans krijgt om te mee te doen aan de wedstrijd. Onder kansengelijkheid wordt niet begrepen dat gezorgd moet worden dat iedereen gelijk aan de start verschijnt of dat de uitkomsten voor iedereen gelijk zouden moeten zijn. Juist niet. Zolang verschillen vanuit meritocratisch perspectief het gevolg zijn van keuzes die mensen zelf kunnen maken, worden verschillende uitkomsten als gevolg van de gemaakte keuzes door de meritocraat als rechtvaardig gezien. Dit soort keuzes vallen daarmee niet onder de zorg van de staat omdat het wordt gezien als de eigen verantwoordelijkheid van mensen. Door iedereen toe te laten tot onderwijs, is van onrechtvaardigheid van sociale ongelijkheid geen sprake, zo is de redenering. Iedereen heeft immers de mogelijkheid andere keuzes te maken. In het licht van de 'procedurele fairness' is het van belang de competitie eerlijk te laten verlopen. Tests hebben vanuit dat perspectief een belangrijke waarde.

### 6.3.2 Meritocratie en testen

De invoering van de brugklas en de discussie over de invoering van de Middenschool hebben de discussie over selectie voor het voortgezet onderwijs in de jaren zeventig en tachtig wat naar de achtergrond geschoven. Dat geldt niet in zijn algemeenheid voor de aandacht voor testen in het onderwijs. Tests in het kader van het vaststellen van capaciteiten en het (systematisch) volgen van vorderingen en resultaten hebben veel aandacht gekregen. Deels met het argument om tijdig

eerst talent en later risicoleerlingen te kunnen opsporen, deels vanuit een zoektocht naar effectieve aanpakken voor 'goed' onderwijs maar vooral later ook vanuit de wens om grip op niveau en daarmee op kwaliteit van onderwijs te krijgen en te houden. De verschillende invalshoeken samen hebben bijgedragen aan het belang en de waarde die ook aan tests meer in het algemeen worden toegekend. En hoewel selectie een tijdlang niet openlijk een na te streven doel is en aan waarde dreigde te verliezen door het zoeken naar uitstel van selectie, heeft de ontwikkeling van tests in het kader van het individueel volgen van leerlingen kunnen zorgen voor een toename van vertrouwen in tests en de uitkomsten ervan.

De vroege selectie, kenmerkend voor het Nederlandse onderwijsstelsel, is door de jaren heen steeds gelegitimeerd door te wijzen op het belang van differentiatie of onderwijssegregatie in het licht van het optimaal benutten van talent. Het vergelijken van leerlingen met elkaar is echter volgens Wilbrink 'een competitief systeem, waarvan de parameters zo zijn afgesteld dat er altijd verliezers zullen zijn, ongeacht hoe men zich inspant' (Wilbrink, 1997: 357). Het beoordelen van leerlingen ten opzichte van elkaar noemt Wilbrink 'bepaald geen meritocratische vorm van beoordelen' omdat 'het nulsomkarakter van de relatieve beoordeling van meet af aan ongelijke kansen scheidt'. De vergrote onderwijsdeelname en de toenemende duur van het onderwijs zouden, volgens Wilbrink, de oorzaak ervan zijn dat het wel lijkt alsof kansen voor havo of vwo meer gelijk zijn geworden, terwijl in werkelijkheid de verhouding van de kansen van leerlingen uit verschillende milieus gelijk is gebleven (Wilbrink, 1997: 363). Het gebruik van objectieve toetsen bij de overgang naar het vo lijkt meritocratiserend, maar is bij nadere beschouwing vanuit meritocratisch oogpunt volgens Wilbrink schadelijk (Wilbrink, 1997: 365). Leerlingen krijgen daarmee niet het onderwijs dat ze nodig hebben vanuit de visie dat onderwijs wordt gezien als iets dat zich moet aanpassen aan de leerlingen die het krijgt, maar leerlingen worden 'geplaatst' in een traditioneel gegeven structuur met alle fouten die bij dat 'plaatsen' onvermijdelijk worden gemaakt (Wilbrink, 1997: 367). Wilbrink spreekt over een 'proto-meritocratisch' ideaal waarmee hij bedoelt dat het oorspronkelijke meritocratische ideaal van gelijke kansen bij gelijke capaciteiten zich vertaald heeft in een situatie waarin 'leerlingen, studenten en sollicitanten ongeacht hun afkomst of andere persoonlijke kenmerken moeten worden afgerekend op verschillen in objectief en in meerdere decimalen nauwkeurig gemeten prestaties' (Wilbrink, 1997: 341) terwijl er een wereld van verschil kan liggen tussen wat verstaan kan worden onder 'capaciteiten' en wat onder 'prestaties'. En bovenop die kloof komt nog de volgens Wilbrink 'ongelukkige combinatie van de aantoonbaar grote feilbaarheid van de instrumenten waarmee prestaties worden gemeten' (Wilbrink, 1997: 379).

Het vergelijkend meten kan door leerlingen met elkaar te vergelijken op basis van ordening van laag naar hoog scorend of door prestaties te vergelijken met een onderwijsinhoudelijke norm waar leerlingen wel of niet aan kunnen voldoen. Het vergelijken van leerlingen met elkaar is de vorm van meten waar psychometristen van het eerste uur zich mee bezig hebben gehouden. L. J. Cronbach, bekend van de nog altijd belangrijke en veel gebruikte betrouwbaarheidscoëfficiënt Cronbachs alpha, ziet dit vergelijken als basis van de definitie van het begrip 'test': 'A test is a systematic procedure for comparing the behavior of two or more persons.' (Cronbach, 1960: 26). Het ordenen en vergelijken van grote groepen individuen met elkaar op basis van scores op een serie opgaven is een praktijk die zich heeft ontwikkeld in relatie tot psychologische tests, met name intelligentietests; de relatieve positie van de score die iemand heeft behaald ten opzichte

van de rest van de populatie, is bepalend voor zijn IQ. In de meritocratische definitie is intelligentie een belangrijke bepalende factor, aangevuld met inzet en/of motivatie. Inzet en motivatie zijn als concept lastig objectief, dat wil zeggen, eenduidig te scoren. Dat heeft gemaakt dat metingen vanuit meritocratisch perspectief zwaar zijn gaan leunen op het willen vaststellen van intelligentie met behulp van intelligentietests waarmee vanaf de eerste Wereldoorlog op grote schaal gewerkt en geëxperimenteerd wordt. De antropoloog Hanson wijst op een ogenschijnlijke tegenstrijdigheid die zich daarbij heeft voorgedaan. In een samenleving die steeds complexer wordt zou een complex systeem van testen en meten worden verwacht om geschiktheid voor een in die samenleving passende rol te vinden. Maar dit is juist niet gebeurd. Er is één aspect boven komen drijven dat succes zou garanderen boven alle anderen: algemene intelligentie. Hanson verklaart dat omdat het technologisch en praktisch mogelijk was dit op grote schaal vast te stellen met de bestaande intelligentietests. Mede dankzij de introductie van de meerkeuzevraag kon goedkoop en snel de intelligentie van miljoenen mensen worden vastgesteld. Dit vereiste wel dat men op grote schaal ging accepteren dat wat uit de test kwam aan intelligentie-score ook gezien zou worden als belangrijk. Het is dus niet zo, stelt Hanson, dat intelligentie daarmee daadwerkelijk bestaat en dat een intelligentietoets dat meet. Het proces van betekenis geven aan intelligentie geeft geen inhoud aan het begrip maar produceert het vooral. In dat kader stelt Hanson dat je kunt spreken van 'kunstmatige' intelligentie (Hanson, 1993: 295).

### 6.3.3 Eindtoets en LVS in meritocratisch perspectief

'Talent moet worden opgespoord' is bij de invoering van de verplichting tot het gebruik van een eindtoets en een leerlingvolgsysteem dan al vijftig jaar een belangrijk motto voor onderwijsbeleid. Talent wordt daarbij gedefinieerd als een te definiëren en vast te stellen, met name cognitieve capaciteit. Vanaf het begin is de eindtoets te typeren als een instrument bedoeld om de 'cognitief besten' op te sporen. Dat was de intentie van het Koninklijk Besluit van 1965, de intentie van de toetsmakers van het eerste uur en van de toetsontwikkelaars die daarna kwamen. In de laatste verantwoording van de eindtoets van Cito uit 2011 wordt de keus voor een nadruk op de cognitieve elementen taal en rekenen gelegitimeerd door te stellen dat 'verschillen in leerprestaties op de onderdelen die in de eindtoets vertegenwoordigd zijn, vermoedelijk voor een belangrijk deel kunnen worden verklaard door verschillen in eigenschappen als 'intelligentie', 'concentratie', 'motivatie' en 'doorzettingsvermogen'. Via de eindtoets worden deze eigenschappen indirect en afgewogen gemeten.' (Van Boxtel et al., 2011: 10). In deze tekst is de formule van meritocratie van Young, 'merit = I.Q. + effort', duidelijk herkenbaar. In de verantwoording voegt Cito daaraan toe, mede op basis van eigen correlatieel onderzoek met bestaande IQ-tests: 'Dat leervorderingstoetsen [bedoeld wordt toetsen zoals de eindtoets] in dit opzicht superieur zijn is vermoedelijk te danken aan het feit dat ze niet alleen iets zeggen over de intelligentie van de leerling maar indirect ook over persoonlijke eigenschappen die het leereffect mede bepalen.' (Van Boxtel et al., 2011: 11). In de eerste publiekelijke verantwoording van CvTE uit 2015 drukt CvTE zich in vergelijkbare bewoordingen uit over de eindtoets (zie ook hoofdstuk 3).

De eindtoets is, volgens de oorspronkelijke makers uit 1967 en nog altijd, te zien als een vorm van selectie op basis van meritocratische uitgangspunten. De toets wordt zo geconstrueerd dat de toets onderscheid kan maken, kan differentiëren, tussen deelnemende leerlingen om de besten te kunnen selecteren. Er wordt voor het verkrijgen van een goede differentiatie gestreefd naar een scoreverdeling die het karakter heeft van een normaalverdeling. Daartoe bevat de toets

opgaven met een spreiding in moeilijkheid waarbij té gemakkelijke items en té moeilijke items alleen worden opgenomen mits ze bijdragen aan een goede differentiatie tussen individuen. Om de resultaten ook jaarlijks met elkaar te kunnen vergelijken, wordt gewerkt met standaardscores. Deze manier van werken is bekend van het gebruik bij intelligentietests en in de Verenigde Staten in de tweede helft van de vorige eeuw ook overgenomen bij andere vormen van testen. De eindtoets is uitgegroeid tot meer dan een instrument om de besten te selecteren maar heeft, zeker bij de huidige wetgeving, de taak om iedereen een plek toe te wijzen in acht onderscheiden categorieën van het voortgezet onderwijs. Dit gebeurt nu op basis van één schaal met één ranking waarbij een score op een combinatie van cognitieve elementen bepalend is voor plaatsing in het vwo (bij een hoge score), voor plaatsing in de basisberoepsgerichte leerweg in het vmbo (bij een lage score) en alle onderwijsvormen daartussenin. Met behulp van één schaal worden alle leerlingen gedetermineerd in de mate waarin één kenmerk, een cognitieve inschatting, aanwezig is en op basis waarvan verdeling over verschillende vormen van vervolgonderwijs wordt gerealiseerd. De verschillende opleidingen vinden zo hun legitimering niet in de verschillen in inhoudelijke kenmerken van de verschillende voortgezet onderwijs-trajecten maar vooral in de verschillen in scores zoals gemeten met de eindtoets. Daarmee is niet de geschiktheid en de passendheid bij de verschillende vormen van het vmbo bepalend maar de vermeende ongeschiktheid, in termen van lager dan gemiddeld cognitief scorend, voor de algemeen vormende vormen van voortgezet onderwijs. De pedagoog Terwel stelt dat de selectiepraktijk in Nederland is gebaseerd op een fundamenteel verkeerd idee over de verdeling van begaafdheden bij leerlingen, namelijk 'wie niet theoretisch is moet wel praktisch zijn' (Terwel, 2006: 11). Deze assumptie is volgens hem eerder een afgeleide van het bestaande schoolsysteem dan dat het gebaseerd is op een logische redenering of op feitelijk aanwezige talenten van leerlingen. Als dergelijke categorieën eenmaal geïnstitutionaliseerd zijn in 'leerwegen' of schooltypen bevestigen zij zichzelf: de hokjes moeten wel gevuld worden (Terwel, 2006: 11). De eindtoets is geen 'talentscout' voor het brede palet aan opties voor het voortgezet onderwijs. Uitgaande van de stand van beheersing van (bepaalde) taal- en rekenvaardigheden met opgaven die bewust oplopen in moeilijkheidsgraad wordt gezocht naar de 'best scorenden' zodat het feitelijk een afvalrace is op basis van een bepaalde combinatie van gekozen inhoud waarvan gesteld wordt dat die een combinatie van IQ en inzet vaststellen. Zo bepaalt een en dezelfde schaal naar welke vorm van voortgezet onderwijs iemand toe 'mag'.

Voor de nadere onderbouwing van het schooladvies wordt scholen geadviseerd de uitkomsten van het LVS een belangrijkere rol toe te kennen. Dit omdat de uitkomsten van de eindtoets sinds de wetwijziging van 2011 zijn verplaatst naar de periode nádat het schooladvies is gegeven. Om schooladvies en advies van de eindtoets zoveel mogelijk in lijn te laten zijn met elkaar, is het belang van het LVS versterkt (zie ook hoofdstuk 3). De toetsen van het LVS hanteren voor een substantieel deel een zelfde combinatie van inhoudelijke alsook psychometrische uitgangspunten als de eindtoets; het is een vergelijkende meting van leerlingen met elkaar. Kijkend naar de combinatie van scores op taal (met name begrijpend lezen) en rekenen wordt scholen aangeraden hun advies vorm te geven. Zo wordt ook het schooladvies, los van de uitkomsten van de eindtoets, meer en meer gebaseerd op het meritocratisch uitgangspunt van intelligentie, dat ontleend wordt aan de scores op begrijpend lezen en rekenen, als relevante factor voor plaatsing in het vervolgonderwijs, eventueel aangevuld met een inschatting van inzet. De aanvulling met gegevens uit het LVS dient daarmee niet om een breder palet aan individuele kenmerken mee te laten wegen in het schooladvies, maar om het meritocratisch oordeel op basis van een cognitieve

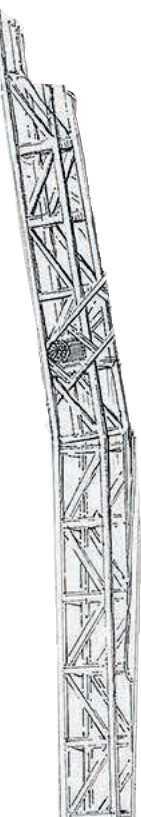
score gebaseerd op vergelijkbare inhouden en schalen te versterken.

Met de verplichte afname van de eindtoets voor alle leerlingen in het basisonderwijs, heeft de overheid via (gecombineerde) scores op een cognitieve meting de regie in handen voor de selectie voor het voortgezet onderwijs. De huidige aanpak is een vervolgstap van een overheid die blijkt geeft van een groeiend vertrouwen in resultaten van op als meritocratisch te typeren tests als instrument voor het verdelen van onderwijskansen.

## 6.4 Samengevat

De meritocratische manier van denken en handelen in onderwijs en samenleving is mede bepalend geweest voor hoe toetsing in Nederland zich heeft ontwikkeld en omgekeerd. De toelating, voortgang en afsluiting van onderwijs zijn verbonden geraakt met het concept van 'verdienste'. Wie het hoogst scoort op een toets die gebaseerd is op de 'communale' onderwijsdoelstellingen, krijgt de grootste beloning. Dit wordt gelegitimeerd door te stellen dat resultaten uit het verleden gezien kunnen worden als voorspellers van toekomstig succes. Kennis is zo onlosmakelijk verbonden geraakt met kunde waarbij kunde meer en meer geassocieerd is geraakt met geschiktheid, 'ability'. In 1993 stelt De Groot, terugblikkend op het werk van Cito, dat Cito niet voldoende geijverd heeft voor het 'verspreiden van verworven algemene inzichten. Ik denk hierbij aan inzicht in de grote verschillen in prestatievermogens van leerlingen die telkens weer uit scoreverdelingen naar voren komen. En aan de onmiskenbare hoofdoorzaak daarvan: grote verschillen in leervermogen, in aanleg.' (Cito, 1993a: 16). De eindtoets is vanaf het begin gericht geweest op het willen vaststellen van leervermogen, van aanleg, van 'ability'. In dat kader wekt het geen verbazing dat bij de ontwikkeling van de eindtoets geleund is op vormen van ability-tests zoals bestaande IQ-tests. De vergelijkende psychometrie die daarvoor beschikbaar was, is overgenomen bij de start van de ontwikkeling van de eindtoets. Het psychometrisch model van de eindtoets is vanaf het begin gericht geweest op differentiatie, in lijn met de bestaande maatschappelijke ordening. De combinatie van inhoud en psychometrisch model is zodanig gekozen dat de ordening die ermee wordt gerealiseerd, een bewuste en gestuurde ordening is. In het geval van de eindtoets is gestuurd naar een ordening van de samenleving zoals die halverwege de jaren zestig in Nederland gebruikelijk was. Daarmee heeft de eerste versie van de eindtoets niet het doel gehad te komen tot een andere manier van ordenen en selecteren om daarmee kansen te bieden aan andere, nieuwe groepen kinderen uit andere lagen van de samenleving. Ook bij latere vormen van valideren is de bestaande onderwijspraktijk als uitgangspunt genomen voor het aantonen van de waarde van de eindtoets. De eindtoets is daarmee eerder te zien als volgend, en als een bevestiging van de bestaande maatschappelijke ordening. De toets in deze vorm en opzet kan geen andere maatschappelijke ordening tot stand brengen, maar is uitsluitend gericht op het reproduceren van de bestaande. De eindtoets is nooit bedoeld geweest en ook nu maar zeer beperkt in staat kansen te bieden aan leerlingen die vanuit een achterstandspositie het onderwijs binnenstromen. Het gebruik van de eindtoets, en zeker in combinatie met het gebruik van toetsen uit het LVS, vormt eerder een verklaring voor het feit dat decennialang vastgesteld wordt dat relatieve posities in de sociale ordening stabiel blijven en door onderwijs maar moeizaam te beïnvloeden zijn. Het ogenschijnlijk meritocratische karakter van de eindtoets schuilt in het feit dat iedereen op hetzelfde moment dezelfde toets aflegt en daarmee het gevoel heeft deel te nemen aan dezelfde wedstrijd. De wedstrijd is daarmee echter niet gelijk en al helemaal niet even kansrijk voor alle deelnemers. Eenmaal opgelopen achterstanden blijken moeilijk in te halen. Het streven van de overheid naar zowel gelijkheid van kansen als naar een meritocrati-

sche ordening staan op gespannen voet met elkaar. De ordening op basis van de eindtoets, in combinatie met LVS, heeft de afgelopen decennia niet kunnen zorgen voor meer gelijkheid van onderwijskansen. De eindtoets meet geen 'ability' maar eerder wat de vruchten zijn die kinderen op 12-jarige leeftijd en in de periode daarvoor hebben weten te plukken van hun opvoeding, van de positie van hun ouders op de maatschappelijke ladder en van het onderwijs dat zij hebben kunnen en mogen genieten. De rechtvaardigheid van de uitkomsten van deze realiteit staat centraal in het volgende hoofdstuk.





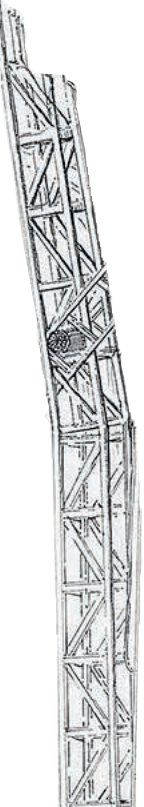






# HOOFDSTUK 7

Rechtvaardigheid en testen



## Hoofdstuk 7 Rechtvaardigheid en testen

*'Ach heel 't leven is één keten van absurde  
coïncidenties en je ouderlijk huis  
gewoon de plek waar een van die triljarden,  
triljarden dobbelstenen van het toeval  
even liggen bleef en toevallig met jou er in.'*

*Uit: Cees Buddingh, 'Het houdt op met zachtjes regenen' (1976)*

De ordening van samenlevingen op basis van het concept van verdienste die zich uitkristalliseert in de loop van de 20e eeuw, roept vragen op naar de mate van rechtvaardigheid die deze nieuwe ordening in de praktijk met zich brengt. Langzaam wordt duidelijk hoe een ideaal van een klas-seloze samenleving met kansen op sociale stijging voor iedereen met daarbij een belangrijke rol voor onderwijs uitwerkt. Grote verschillen tussen winnaars en verliezers tekenen zich af terwijl de sociale mobiliteit nauwelijks van de grond komt in de meeste landen. Het roept de vraag op naar de rechtvaardigheid van deze ordening en de manier waarop deze tot stand komt. En naar de reden waarom er op bepaalde terreinen sprake is van schaarste waarbij er verdeeld moet worden. Zoals in het geval van onderwijs: wat maakt dat onderwijs wordt beschouwd als zodanig schaars dat een verdeling nodig is? Schaarste wordt volgens filosoof Achterhuis gezien als een soort natuurfeit terwijl dat nergens systematisch wordt geproblematiseerd of als mogelijkheden-voorwaarde wordt bekritiseerd. Hij stelt dat de moderne maatschappij de schaarste niet heeft overwonnen maar juist heeft gecreëerd (Achterhuis, 1988: 13). Volgens Achterhuis lijkt schaarste 'het meest verborgen uitgangspunt van de moderne maatschappij' te zijn (Achterhuis, 1988: 12). Frissen typeert schaarste als een fundamentele assumptie van de moderniteit. Daarmee heeft volgens Frissen vrijwel elke opvatting over gelijkheid en ongelijkheid betrekking op verdeling en criteria voor verdeling. Het lijkt er volgens hem voortdurend om te gaan welke mate van ongelijke verdeling rechtvaardig is en welke niet. Daarom is de waarde van gelijkheid ook alleen maar relevant onder de voorwaarde van schaarste (Frissen, 2007: 55). Frissen benoemt de creatie van gelijkheid het belangrijkste politieke programma van de moderniteit dat tegelijkertijd tot mislukken gedoemd is, omdat het schaarste moet veronderstellen en dus niet aan ongelijke verdeling zal kunnen ontsnappen (Frissen, 2007: 56). Schaarste in combinatie met dynamische behoeften van de moderne mens is de motor van de vooruitgang. De samenleving profiteert ervan. En in het licht van die vooruitgang is het van belang dat schaarste verborgen blijft omdat deze anders ambities doodt (Frissen, 2007: 55). We weten dat schaarste tot ongelijkheid leidt en juist daarom willen we deze bestrijden en we weten dat ongelijkheid onvermijdelijk is, maar kunnen dat niet met onze wens tot rechtvaardigheid in overeenstemming brengen (Frissen, 2007: 55). In institutionele arrangementen van de verzorgingsstaat ziet Frissen drie vormen van gelijkheid: gelijkheid van startposities, gelijkheid van aanspraken, rechten en voorzieningen en gelijkheid van uitkomsten. Omdat de gelijkheid waarnaar gestreefd wordt niet absoluut is maar relatief, is er volgens Frissen altijd sprake van ongelijke verdeling die meer of minder rechtvaardig is (Frissen, 2007: 61). Als verdelingen leiden tot uitkomsten waarbij besloten wordt wie wel mee mag doen en wie niet, doen zich vraagstukken voor die samenhangen met vormen van sociale uitsluiting. In de kern komt sociale uitsluiting volgens socioloog Schuyt neer op ergens niet meer bij horen. De oorzaak van het er niet bij horen kan behoorlijk verschillen. Vandaar dat Schuyt 'uitsluiting'

bij voorkeur omschrijft als: 'Er niet bij mogen horen; er niet bij kunnen horen; er niet (meer) bij willen horen.' (Schuyt, 2009: 69). Alle drie vormen van uitsluiting acht Schuyt van belang. 'Het er niet bij mogen horen slaat op groepen die opzettelijk worden gediscrimineerd of er buiten worden gehouden. Ze worden op meer of minder subtiele wijze naar de marge van de samenleving geschoven. De geschiedenis is vol van voorbeelden van opzettelijke discriminatie: joden, katholieken, negers, zigeuners, homoseksuelen. Maar omdat de moderne samenleving – juist op grond van die historische ervaringen – een ideologie heeft opgebouwd van gelijke rechten en van democratische participatie van alle burgers, voelt men aan dat het niet helemaal juist is als men mensen aan de kant schuift. Er treedt een ambivalentie op in de houding van de machtigen in een samenleving ten opzichte van de potentieel uitgesloten, die dit ook haarfijn aanvoelen. Ze horen erbij, maar tegelijk horen ze er toch niet bij.' (Schuyt, 2009: 69).

In dit hoofdstuk wordt de rol van tests nader beschouwd in het licht van rechtvaardigheid van verdelingen. Allereerst worden de meer algemene principes van rechtvaardigheid van verdelingen onderzocht, worden algemene verdelingsprincipes beschreven en wordt gekeken naar de rol van politiek daarbij (7.1). Vervolgens wordt ingegaan op hoe verdelingen hun beslag krijgen in onderwijs (7.2) en hoe tests internationaal en nationaal, zoals LVS en de eindtoets, daarbij hun rol hebben gekregen (7.3). Tot slot wordt stilgestaan bij de eindtoets en de functie ervan: Is er sprake van selectie, toelating of plaatsing? (7.4).

## 7.1 Rechtvaardigheid en verdeling

In het voorwoord bij een nieuwe editie van 'The rise of the meritocracy' uit 1994 stelt Young dat Rawls een van de schrijvers is die gezien heeft dat het streven naar faire kansen kan leiden tot een harde en hardvochtige meritocratische samenleving. En dat een vorm van correctie op die hardheid gewenst is. Het 'difference principle', een belangrijk element uit de rechtvaardigheidstheorie van Rawls die in 1972 verschijnt, is volgens Young bedoeld om het persoonlijke en sociale leven van alle burgers, ook van hen die minder bedeeld zijn, van waarde te laten zijn (Young, 1994: XVI). Met zijn boek 'A Theory of Justice' heeft Rawls het denken over 'rechtvaardigheid' en 'verdelen' bepalend weten te beïnvloeden en te veranderen. Het 'Handbook of Social Justice Theory' (2016) start met een samenvatting van het denken van Rawls. Rawls zorgt dat het heersende, utilitaire denken plaatsmaakt voor denken vanuit meer fundamentele opvattingen over rechtvaardigheid. Hij hanteert twee belangrijke principes ten aanzien van rechtvaardigheid. Hij richt de vraag naar rechtvaardigheid op de beschikbaarheid van zogenaamde 'primaire sociale goederen' waaronder hij een aantal fundamentele vrijheidsrechten verstaat, zoals vrijheid van meningsuiting, vrijheid van vereniging, vrijheid van beweging en beroepskeuze, gelijke toegang tot macht en verantwoordelijke posities, inkomen en vermogen, de mogelijkheid om je zelfrespect te handhaven. Ongeacht achtergrond, ideologie, levensbeschouwing of overtuiging, heeft een mens behoefte aan deze

**(a) "Each person has the same infeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all; and**

**(b) "Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least-advantaged members of society (the difference principle)"**

*Uit: Sabag & Smitt, M. (Eds). (2016) Handbook of Social Justice Theory and Research, p. 41-42*

goederen en Rawls acht ze nodig om in vrijheid een eigen levensplan te kunnen uitvoeren. Hij gaat ervan uit dat iedereen voor zichzelf zoveel mogelijk vrijheid zal willen garanderen, het eerste principe (a). Een tweede belangrijke waarde is die van (on)gelijke kansen (b) waarbij Rawls stelt dat sociale en economische ongelijkheden onder twee voorwaarden acceptabel zijn. Ten eerste moet de toegankelijkheid ervan zodanig geregeld zijn dat er sprake is van kansengelijkheid en ten tweede dat ongelijkheid van positie en inkomen mag, maar alleen als dat de positie van de zwaksten in de samenleving ten goede komt.

Verschillen tussen mensen zijn volgens Rawls niet te beschouwen als verdiensten: ze zijn in hoge mate toe te schrijven aan toeval en geluk en mogen geen aanleiding geven tot sociale ongelijkheid. Dit geldt niet voor keuzes die mensen maken. Ongelijkheid die daardoor ontstaat, heeft geen compensatie nodig in de ogen van Rawls. Met de beide principes, in volgorde van toe te passen, komt hij tegemoet aan het verlangen naar vrijheid en het verlangen naar gelijkheid (Sabagh & Schmitt, 2016: 42). Het denken van Rawls heeft bijgedragen aan een belangrijke nuancering van het streven naar gelijke toegankelijkheid door te stellen dat ongelijkheden in hulpbronnen alleen maar zijn toegestaan als die uiteindelijk ten goede komen aan de minst bedeelde groep. Econoom en filosoof Sen heeft als kritiek op Rawls gewezen op het feit dat sociaal-economische ongelijkheden niet gemeten moeten worden in termen van hulpbronnen (zoals inkomen, vermogen en gelijke kansen in onderwijs en op de arbeidsmarkt) maar in termen van de door hem zo genoemde 'capabilities' van mensen (Sen, 2013: 13). Dit zijn de reële mogelijkheden die mensen hebben om na weloverwogen reflectie te doen wat ze willen doen, en te zijn wie ze willen zijn in een samenleving die alle effectieve hindernissen moet wegnemen (Sen, 2013: 13). Sen en filosoof Martha Nussbaum worden beschouwd als de belangrijkste vertegenwoordigers van de 'Capability approach'. Deze benadering is een poging tot comparatieve inschatting van levenskwaliteit en tot het theoretiseren over fundamentele sociale rechtvaardigheid (Nussbaum, 2012:38). Het is niet bedoeld als een rechtvaardigheidstheorie maar is vooral gericht op het stellen van de vraag hoe we onrechtvaardigheden kunnen verminderen (Sen, 2013: 16). Het idee van capabilities richt zich op vrijheid en kansen, dat wil zeggen op het feitelijke vermogen van mensen om binnen hun mogelijkheden verschillende soorten levens te leiden: van resultaten naar kansen! (Sen, 2013: 258). Aanhangers van de 'Capability approach' vinden dat samenlevingen mensen kansen of substantiële vrijheden moeten bieden waarvan mensen vervolgens in hun handelen al dan niet gebruik kunnen maken: de keuze daartoe is aan hen. Daarmee wordt deze benadering volgens Nussbaum gekenmerkt door respect voor het vermogen tot zelfbepaling van mensen. Capabilities die wezenlijk zijn voor mensen kunnen niet alleen kwantitatief van elkaar verschillen maar ook kwalitatief waarmee ze niet zonder vervorming te reduceren zijn tot punten op één numerieke schaal (Nussbaum, 2012: 38). Samenlevingen die menselijke capabilities willen bevorderen, richten zich op het creëren van mogelijkheden voor de ontwikkeling van capabilities, onder andere door middel van onderwijs. Want iedereen dient boven een bepaald drempelniveau van gecombineerde capabilities te komen. Dat zal volgens Nussbaum altijd inhouden dat mensen met gelijk respect behandeld dienen te worden: 'De opstelling ten opzichte van 'elementaire capabilities' van mensen is dus niet meritocratisch – dat wil zeggen dat mensen die over meer aangeboren vaardigheden beschikken een betere behandeling krijgen – maar eerder het tegenovergestelde: dat degenen die meer hulp nodig hebben om boven de drempelwaarde uit te komen, meer hulp krijgen.' (Nussbaum, 2012: 45). Toegepast op de context van onderwijs streven aanhangers van de capability-benadering volgens Dewan ernaar dat onderwijs zo

wordt ingericht dat het ieder kind het bewustzijn, de kansen en de capabilities geeft om zijn of haar eigen pad te kiezen en te zijn wat hij of zij wil zijn (Dewan, 2020: 10). Daarmee staat deze benadering haaks op de invulling van onderwijs in de geest van de human capital theorie waar meritocratische uitgangspunten de belangrijkste criteria vormen voor toelating, selectie en vormen van uitsluiting. En waar mensen gezien worden als 'resources' bedoeld voor de ontwikkeling van de natie, en meer concreet de economie en de markten. Deze manier van kijken heeft volgens Dewan gevolgen voor de doelen van onderwijs en de activiteiten van onderwijs die meer en meer in dienst zijn komen te staan van economische groei en minder in dienst van andere ethische en morele principes (Dewan, 2020: 10). Deze analyse van Dewan sluit aan op de ontwikkelingen die Nussbaum signaleert als zij spreekt over een crisis die 'grotendeels onopgemerkt voortwoekert [...]; een crisis die op lange termijn veel schadelijker kan zijn voor de toekomst van het democratische zelfbestuur: een wereldwijde crisis in het onderwijs.' (Nussbaum, 2011: 16). Zij doelt hierbij op de ontwikkeling richting een uitsluitend op groei gericht onderwijsmodel (zie ook hoofdstuk 3). De Amerikaanse filosoof Sandel stelt dat het bij rechtvaardigheid niet alleen gaat om de juiste manier om dingen te verdelen, maar ook om de juiste manier om dingen te waarderen (Sandel, 2010: 308). Een rechtvaardige samenleving houdt volgens Sandel in dat we met elkaar beraadslagen over de goede manier van leven wat vraagt om een passend politiek discours om ons daarheen op weg te helpen (Sandel, 2010; 308).

Het huidige, gangbare uitgangspunt dat onderwijs gericht moet zijn op het selecteren van de besten, die vervolgens de meeste kansen krijgen in het onderwijs op het behalen van de meest waardevolle diploma's, ontleent zijn waarde voor een belangrijk deel aan de voorwaarde van schaarste. Schaarste, ook gecreëerde schaarste, maakt dat verdelingen moeten worden aangebracht. Binnen samenlevingen gemaakte verdelingen bepalen wie er wel en wie er niet meedoen, wat relevant geacht wordt in dat verband en hoe de verdeling geoperationaliseerd wordt op basis van welke principes.

### 7.1.1 Schaarste en verdelingen

De politiek wetenschapper en filosoof Jon Elster gaat in zijn boek 'Local Justice' in op principes die voor verdelingen kunnen worden gehanteerd, verdelingen van goederen én van lasten. Hij maakt onderscheid in drie kenmerken die goederen en lasten kunnen hebben die in combinatie zorgen voor verschillende vraagstukken van verdelingen: het eerste kenmerk heeft betrekking op de hoeveelheid die ergens van is (al dan niet schaars zijn), het tweede kenmerk op de variabiliteit (wel of niet opdeelbaar zijn) en het derde kenmerk op de kwaliteit (de waarden van de te verdelen eenheden: homogeen versus heterogeen).

Op basis van deze drie kenmerken onderscheidt Elster vijf belangrijke combinaties waarop verdelingen betrekking kunnen hebben:

1. Het goed is niet schaars maar wel heterogeen op aspecten die door de ontvangers belangrijk gevonden worden (bijv. werkplekken)
2. Het goed is schaars, niet opdeelbaar en homogeen (bijv. toelating tot Harvard)
3. Het goed is schaars, niet opdeelbaar en heterogeen (bijv. hart, lever, longen voor transplantatie)
4. Het goed is schaars, opdeelbaar en homogeen (bijv. water, energie en veel consumptiegoederen)
5. Het goed is schaars, opdeelbaar en heterogeen (bijv. land)



Is er sprake van een goed of een last die beperkt is in hoeveelheid of varieert in kwaliteit dan is een procedure waarbij goed en afnemer in ogenschouw worden genomen van belang. Elster sluit zich in dat geval aan bij het eerder door de Nederlander Hofstee (Allocation by lot) gemaakte onderscheid tussen:

1. Selectie - individuen worden vergeleken met elkaar, meestal door een rankinglist te produceren waarbij de acceptatie topdown gaat totdat het goed op is.
2. Toelating - bij toelating worden individuen vergeleken met een norm en het goed wordt alleen gegeven aan hen die aan de norm voldoen.
3. Plaatsing - een plaatsingsprocedure is bedoeld om niet-schaarse, heterogene goederen toe te wijzen zodanig dat iedereen eindigt met een deel ervan.

Volgens Elster laten toelatingsprocedures zoals voor onderwijs zich slecht met schaarse, niet in hoeveelheid op te delen en rigide beperkte goederen verenigen. Hij onderscheidt twee soorten toelatingsprocedures: 'pure' toelatingsprocedures (de gehanteerde grens, de threshold, is passend voor de te nemen beslissing) en 'impure' toelatingsprocedures (hierbij wordt een proxy gehanteerd voor de te hanteren grens) waarbij de grens zo gekozen is dat er een redelijke match te maken is tussen het aantal mensen en de goederen. Toelatingsprocedures die gebaseerd zijn op een met behulp van een proxy te hanteren grens (impure) zijn volgens Elster acceptabel in twee gevallen: het te verdelen goed is makkelijk op te delen zodat feitelijk bijgesteld kan worden opdat iedere gekwalificeerde toegang ertoe kan krijgen en ten tweede als het te verdelen goed makkelijk uit te breiden is tot het matcht met het aantal mensen dat de grens gehaald heeft. Bij toelating is vaak sprake van een toelatingsprocedure die als 'impure' te karakteriseren is waarbij er, als er meer goederen beschikbaar komen, sprake is van een verlaging van de grens zodat het beschikbare aantal goederen wel verdeeld wordt of er voor eenieder die voldoet aan de grens, wel een goed beschikbaar komt. En zo is er bij toelating vaak sprake van een tweetrapsraket (Elster, 1992: 24-27).

Hoeveel van een goed er beschikbaar is, is volgens Elster aan de 'eerste orde' van actoren om te bepalen of om beschikbaar te stellen. De eerste orde bepaalt of er gegeven wordt, hoeveel en aan wie. Eerste orde beslissingen worden veelal genomen op het niveau van de overheid (Elster: 1992: 140). De overheid laat zich daarbij vooral leiden door zorgen over de efficiency maar ook door billijkheid omdat ze beducht is voor de publieke opinie (Elster, 1992: 180). Voor de feitelijke verdeling van het goed, het uitvoeren van procedures, is de tweede orde van actoren van belang. Zij bepalen 'wie waar naartoe gaat' of 'wie wat krijgt'. Dit gebeurt veelal op het niveau van lokale instanties. Ook op dit niveau is efficiency, niet noodzakelijkerwijs dezelfde efficiency als in de eerste orde, van belang. Dit geldt ook voor het argument van billijkheid. Eigenbelang speelt een rol als op dit niveau de beslissingsprocedures gevoelig zijn voor of te beïnvloeden zijn met corruptie of steekpenningen (Elster, 1992: 182). De derde orde wordt gevormd door de ontvangers van het te verdelen goed. Zij kunnen soms beslissingen beïnvloeden met bepaald gedrag, bijvoorbeeld door hard te werken en goede resultaten te behalen. Als beïnvloeding mogelijk is, is dat gelijk een incentive om het gedrag te gaan vertonen dat nodig is om voor een goed in aanmerking te komen (Elster, 1992: 142). Van de derde orde, de ontvangers, kan worden verwacht dat zij in hoge mate gedreven worden door eigenbelang, openlijk of achter de schermen (Elster, 1992: 182).

Zo zijn er volgens Elster drie niveaus waarop beslissingen over schaarste en de verdeling ervan kunnen plaatsvinden. Voor de verdeling zelf kan het principe gehanteerd worden waarbij er geen sprake is van beïnvloeding door het derde orde niveau, dus door de ontvangers. Hierbij kan het dan gaan om gelijkheidsprincipes in de vorm van bijvoorbeeld lotingen of om tijd-gerelateerde principes (senioriteit, wachttijd). Maar veelal hebben verdeelprincipes wel een relatie met een kenmerk van de ontvanger, veelal gebaseerd op 'behoefte', 'verdienste' of 'status'. Die individuele kenmerken kunnen objectief vaststelbaar zijn of moeten worden gemeten met een vorm van een meting/assessment. Daarbij zijn sommige kenmerken door het individu te beïnvloeden of juist onveranderbaar, waarbij de relatie relatief of absoluut kan zijn. Met relatief wordt bedoeld wie het meest van een relevant kenmerk heeft en in het geval van absoluut gaat het erom dat iemand wat betreft het kenmerk aan een minimumeis voldoet.

Elster wijst erop dat de kwetsbaarheid van verdeelsystemen die gebaseerd zijn op persoonsgebonden kenmerken groter is dan bij niet-persoonsgebonden kenmerken: een wachtlijst wordt als minder kwetsbaar gezien dan een beslissing van een arts wie het eerst geholpen moet worden. Hoe beter objectiveerbaar en niet manipuleerbaar een verdeelsysteem is, hoe minder kwetsbaar (Elster, 1992: 169). De meeste procedures zijn onder te brengen in een van de volgende twee soorten: een procedure met expliciete criteria die zonder extra regels toegepast kunnen worden op het uitvoeringsniveau van de tweede orde. De andere soort vraagt nadere uitwerking van de toepassing of uitleg van de principes/procedures die op uitvoeringsniveau van de tweede orde kunnen/mogen gebeuren. Op die manier ordenend, komt Elster tot een ordening van kenmerken in vier kwadranten (Elster, 1992: 69, tabel 3.1)

	<b>Een onderscheidend assessment is nodig</b>	<b>Er is geen onderscheidend assessment nodig</b>
<b>Afhankelijk van door het individu uit te voeren acties</b>	Behoeften, verdiensten, sociale belangen, verworven vaardigheden	Gewicht, leefstijl gerelateerde ziektes, burgerschap, huwelijkse staat, religie, geletterdheid, rijkdom
<b>Niet afhankelijk van door het individu uit te voeren acties</b>	Pijn, aangeboren vaardigheden, intensiteit van voorkeuren	Leeftijd, geslacht, ras, op genetische afwijkingen gebaseerde ziektes

Tabel 4 Procedure voor verdelingen

Procedures voor verdelingen hebben volgens Elster de neiging om terug te vallen op expliciete criteria die zonder extra regels kunnen worden toegepast, bij voorkeur gebaseerd op mechanische formules of onpersoonlijke mechanismen. Elster stelt dat er in samenlevingen een trend waarneembaar is om van discretionaire beslissingen in de richting te gaan van mechanische formules of onpersoonlijke mechanismen. Deze kosten minder tijd en geld, zijn minder gevoelig voor corruptie en voor het beeld dat er corruptie kan zijn. En ze hoeven niet minder effectief te zijn.

Verdelingsvraagstukken doen zich volgens Elster voor als het goederen betreft die ofwel beperkt zijn in beschikbare hoeveelheid ofwel variëren in kwaliteit. Voor het delingsvraagstuk van onderwijs geldt dat onderwijs een goed is dat in principe niet beperkt hoeft te zijn in hoeveelheid

en ook niet in kwaliteit. Zoals in voorgaande hoofdstukken is besproken, heeft Nederland er nadrukkelijk voor gekozen niet alle onderwijs voor iedereen toegankelijk te maken en te kiezen voor toedeling op basis van selectie gebaseerd op meritocratische uitgangspunten: de combinatie van inzet en intelligentie is bepalend voor wie welk onderwijs mag volgen en daarmee ook hoeveel onderwijs en hoe lang. Dit is een politieke keuze die samengaat met politieke besluitvorming.

### 7.1.2 Politiek van verdelen

Bakker en Lehning stellen dat mensen tegenover de toegang tot of de verdeling van schaarse voorzieningen of goederen vaak een wat tweeslachtige houding aannemen van enerzijds ontkennen en het niet accepteren van schaarste en anderzijds er weinig problemen mee hebben als het maar rechtvaardig gebeurt (Bakker & Lehning, 1991: 9). Want noch het 'recht van de sterkste' noch 'wie het eerst komt wie het eerst maalt' wordt gezien als een bevredigende uitweg uit het selectieprobleem waarmee de legitimiteit van de rol van bijvoorbeeld de overheid met de groei van het aantal selectievraagstukken en de selectie van burgers voor bepaalde voorzieningen steeds meer onder druk komt te staan. Bakker en Lehning zien dat er problemen ontstaan wanneer moet worden bepaald welke criteria van rechtvaardigheid legitiem zijn, wie de criteria vaststelt en aan wie selectie van mensen wordt overgelaten (Bakker & Lehning, 1991: 10). Verdelingen, zeker verdelingen van levenskansen, zoals via onderwijs, zijn zaken waar samenlevingen zich mee bezighouden, keuzes in maken en toezien op de uitvoering ervan. Zodat een ordening ontstaat die in lijn is met de verdelingsprincipes die in die samenleving blijkbaar als (voldoende) billijk worden gezien. Achter die noodzaak tot regie en belangenbehartiging gaat meer schuil. Verdelingsvraagstukken bestaan ook omdat er belangen zijn die samenhangen met het bekleden van een positie van verdeler van kansen.

Rechtvaardigheid van verdelingen vormt de kern van politieke controverses. Want daarover gaat politiek primair, volgens Stone. Een van de meest bekende definities van politieke wetenschappen is volgens Stone die van Harold Laswell. Hij definieert politieke wetenschappen als volgt: 'It is the study of who gets what, when and how.' (Stone, 2012: 39). Verdelingsvraagstukken hebben volgens Stone een paradox: gelijke verdeling (equality) betekent heel vaak ongelijkheid; en gelijke behandeling betekent vaak ongelijke behandeling. Dezelfde verdeling is vanuit het ene gezichtspunt gelijkheid en vanuit het andere ongelijkheid. Stone gebruikt de term 'equality' om te verwijzen naar 'hetzelfde'; uniformiteit als uitgangspunt van de verdeling. En 'equity' om verdelingen aan te duiden die gezien kunnen worden als gelijk in termen van fair (Stone, 2012: 41). Stone stelt dat argumenten voor het classificeren en differentiëren overtuigend zijn in de mate waarin zij bevredigend rekenschap geven van de rechtmatigheid om gevallen gelijk of juist verschillend te behandelen. Politieke redeneringen gaan volgens Stone dan ook primair over het beredeneren van gelijkheid of verschil, over goed en kwaad of over juist en onjuist (Stone, 2012: 382).

Elke verdeling heeft volgens Stone drie belangrijke dimensies: de ontvangers (wie krijgt wat?), het te verdelen goed (wat wordt verdeeld?) en het proces (hoe wordt de verdeling geregeld/uitgevoerd?) met elk hun eigen uitdagingen. Op het niveau van de ontvangers is er het vraagstuk van 'lidmaatschap' waarbij het gaat om wie er voor gelijke behandeling in aanmerking komt. Het tweede vraagstuk is het vraagstuk van beloning en waar dat op te baseren. Individuele prestaties als basis voor beloningen vormen het overgrote deel van de rechtvaardiging van inkomensongelijkheid, maar individuele prestaties in termen van verdienste is een lastiger begrip dan het op

het eerste gezicht lijkt. Individuele verdienste is lastig te definiëren maar ook lastig meetbaar te maken en daadwerkelijk te meten. Daar hoort de vraag bij hoeveel credits een individu zou behoren te krijgen voor zijn prestaties want publieke en private investeringen, politiek op alle besluitvormingsniveaus, en culturele, sociale en economische kansen dragen allemaal bij aan de individuele capaciteit om goed te presteren. Wat is daadwerkelijk de verdienste van het individu en wat van zijn context? Het derde punt dat Stone naar voren brengt is verdeling op basis van 'ranking'. Veel organisaties betalen hun mensen op basis van een rangordening. Onder veel van de verdelingen van ranking zit het idee dat verdienste en ranking correleren maar de vraag moet worden gesteld of systemen mensen helemaal correct en fair ordenen en of de criteria voor het maken van verschil daadwerkelijk de goede criteria zijn? De rechtvaardiging daarvan heeft vaak te maken met handelen of presteren uit het verleden. 'Group-based' verdelingen wijzen mensen toe aan groepen op basis van kenmerken die niks te maken hebben met individuele kwaliteiten of prestaties. De kenmerken zijn meestal niet-veranderbaar, het gaat daarbij veelal om kenmerken die individuen niet kunnen beïnvloeden. Group-based verdelingen hebben de neiging tot grote, sociale verdelingen te leiden, verdelingen die een samenleving historisch gezien verdelen in geprivilegieerden en achtergestelden. Voor wat betreft het te verdelen goed is het van belang te bepalen wat de behoefte eraan is en wat de waarde ervan is voor verschillende ontvangers. Voor het proces van verdelen is er dan de vraag hoe dat proces moet verlopen: via een competitie, via loten of via stemmen (Stone, 2012: 47). Alle met deze verdelingen samenhangende dilemma's moeten worden beoordeeld op in hoeverre ze accuraat zijn, redelijk, overtuigend en uiteindelijk fair. Het is in de optiek van Stone aan samenlevingen om die vraagstukken aan te kijken en deze vragen met elkaar te beantwoorden.

De politicologe Young (1990) stelt dat er twee assumpties liggen onder discussies over gelijke kansen waarvan de rechtvaardigheid niet ter discussie wordt gesteld. De eerste assumptie die als een gegeven wordt gezien en daarmee niet ter discussie staat, is die van een hiërarchische verdeling van arbeid met schaarse posities van hoog inkomen, status, macht en prestige aan de top en minder geprivilegieerde posities aan de onderkant. De tweede assumptie is dat de verdeling ervan moet gebeuren op basis van 'merit'. Daartoe moet de individuele competentie van mensen worden gemeten en zij die het meest gekwalificeerd zijn volgens een onpartijdige meting van dergelijke competenties mogen als beloning de meest competitieve posities krijgen (Young, 2011: 193). Maar, stelt Young, onpartijdige, waarde-neutrale, wetenschappelijke maten van verdiensten bestaan niet. Young spreekt in dit verband over de 'myth of merit'. Het principe van verdienste houdt in dat posities moeten worden toebedeeld aan de meest gekwalificeerde individuen, dat wil zeggen aan hen die de grootste geschiktheid en vaardigheden hebben voor de taken die bij de toekomstige posities moeten worden uitgevoerd. Daartoe moet de onrechtvaardige hiërarchie van standen worden vervangen door een 'natural' hiërarchie van intellect en vaardigheid (Young, 2011: 200). Dat maakt dat het een 'major issue' van rechtvaardigheid moet zijn wie bepaalt wat passende kwalificaties voor een bepaalde positie zijn, hoe die zullen worden vastgesteld en of bepaalde individuen die wel of niet bezitten (Young, 2011: 193). Zij verwijst naar Rawls, die stelt dat aangeboren talenten een arbitrair uitgangspunt zijn voor de verdeling van posities en beloningen omdat een individu net zo beperkt verantwoordelijk is voor zijn talenten als voor zijn of haar ras (Young, 2011: 201). Dit feit maakt dat velen ertoe overgaan om aangeboren talenten niet te willen nemen als invulling van 'verdiensten', maar inspanningen en daadwerkelijk geleverde prestaties. Fishkin komt volgens Young in 1983 met een 'oplossing' door

het principe van verdienste te definiëren als ‘widespread procedural fairness in the evaluation of qualifications of positions’ (Young, 2011: 201). Procedurele fairness heeft betrekking op de vereiste dat het proces van evaluatie het model van een onpartijdige competitie benadert. Een fair assessment van individuele kwalificaties zoals Fishkin die voor zich ziet, moet berusten op het verleden van een persoon of op zijn actuele prestaties op relevante taken, waarbij hij stelt dat determinatie van kwalificaties niet kan berusten op statistische gevolgtrekkingen. Het gebruik van het principe van verdienste bij de verdeling van schaarse en gewilde posities is volgens Fishkin alleen rechtvaardig als de volgende, strenge uitgangspunten worden gehanteerd: ten eerste, kwalificaties moeten worden gedefinieerd in termen van technische skills en competenties, onafhankelijk van en neutraal met betrekking tot waarden en cultuur. Ten tweede moeten de technische skills en competenties ‘job-related’ zijn zodat ze kunnen fungeren als voorspellers van excellent presteren in die posities en ten derde moeten prestaties en competenties individueel worden beoordeeld. Individuen moeten worden vergeleken en geordend met standaarden die onafhankelijk van en neutraal zijn met betrekking tot waarden en cultuur (Young, 2011: 201). Young stelt echter dat het onmogelijk is aan deze eisen te voldoen: ‘The idea of merit criteria that are objective and unbiased with respect to personal attributes is a version of the ideal of impartiality, and is just as impossible.’ (Young, 2011: 202). Dit zorgt voor druk om performance-proxies te vinden die individuele competenties en prestaties substitueren, en die wel onafhankelijk en neutraal zijn met betrekking tot waarden en cultuur. Onderwijsdiploma’s en gestandaardiseerde testresultaten functioneren, aldus Young, in onze samenleving als de belangrijkste ‘proxies’ voor het direct selecteren voor en voorspellen van loopbaanprestaties. Maar Young stelt dat ook onderwijsresultaten niet neutraal zijn en dat daarvoor hetzelfde geldt als voor neutraliteit van baangerelateerde vormen van testen: ‘Despite beliefs to the contrary, however, educational attainment and test results are no more neutral than more direct evaluations of performance.’ (Young, 1990, ed. 2011: 206). Standaardtests voldoen weliswaar aan de vereisten van procedurele fairness. Ze zijn volgens Young blind voor ras, sekse of etniciteit. En ze zijn objectief in die zin dat de beoordeling objectief kan verlopen en dat verschillende beoordelaars tot eenzelfde oordeel zullen komen. Met behulp van statistische technieken zullen leerlingen vergeleken en geordend kunnen worden zonder subjectiviteit. Maar daar stelt Young een aantal problematische aspecten van gestandaardiseerd meten tegenover. Ten eerste stelt Young dat tests zelf een cultuur hebben. Tests belonen bepaalde persoonlijke of culturele kenmerken zoals competitiviteit, de kunst om goed alleen te werken, de vaardigheid om snel te werken of een voorliefde voor abstractie. Wat een test ook beoogt te meten, individuen met deze kenmerken en vaardigheden hebben een voordeel. Daar komt bij dat inhoud die zich eenvoudig in testopgaven laat vertalen de facto grotere waarde krijgt dan inhoud die moeilijk te operationaliseren is. Er wordt veel waarde gehecht aan tests, aldus Young, omdat ze een manier lijken te zijn die voorziet in de behoefte van twee van de eisen voor de evaluatie van verdienste: een precieze maat voor het vaststellen van de competentie van ieder individu en een vergelijking en daarop gebaseerde rangordening van alle individuen. Tests lijken de behoefte van het meritocratisch uitgangspunt te bevredigen; personen kunnen worden beloond voor hun eigen, individuele prestatie. Maar omdat de tests geüniversaliseerd zijn en gestandaardiseerd, is de individualiteit van testresultaten een illusie. Volgens het proces dat Foucault het ‘normalisatieproces’ noemt, produceren tests een ‘reconstituted individuality’ vanuit een casus of score. Het normalisatieproces van de reductie van alle individuen tot een gemeenschappelijke maat is nodig om verschillen te kunnen reconstrueren als afwijking of devaluatie. Daar komt bij dat testonderzoekers, waaronder Shepard, stellen dat er

geen technische methode is die bias kan voorkomen en die kan verzekeren dat een test (even) fair is voor alle deelnemers (Shepard, 1993: 22) (zie ook hoofdstuk 8). Gestandaardiseerde tests kunnen volgens Young niet worden voorzien van precieze kwantitatieve individuele maten van technische of cognitieve competenties onafhankelijk van en neutraal met betrekking tot waarden en cultuur (Young, 2011: 210). Young komt daarmee tot de conclusie dat het gebruik van het criterium van 'verdienste' en daarop gebaseerde tests een plaats verdienen in het politieke debat. 'The ideology of merit seeks to depoliticize the establishment of criteria and standards for allocating positions and awarding benefits. Controversy about schooling, credentials, tests, and admissions and employment policies should be sufficient to show that this depoliticizing effort fails. Once we understand merit evaluation as political, then important questions of justice arise beyond distribution, questions about who should decide on qualifications and by what norms and principals.' (Young, 2011: 211).

### 7.1.3 Arbeidsmarkt en schaarste

Socioloog Mijs stelt dat de toepassing van het meritocratisch uitgangspunt voor de context van onderwijs niet alleen niet vervuld is maar ook een niet te vervullen belofte is. Verdeling van talenten over kinderen is niet gebaseerd op meritocratische principes. De meritocratische race start vanuit ongelijke en niet-meritocratische startposities (Mijs, 2016: 10). Dat meritocratie toch als uitgangspunt wordt genomen heeft volgens hem twee belangrijke redenen: de eerste is dat het een rechtvaardiging biedt voor de manier waarop beloningen, in de breedste zin van het woord, worden verdeeld en de tweede ligt in het verlengde daarvan, namelijk dat het leidt tot het tonen van inzet. Het is volgens Mijs de combinatie van deze twee elementen die de aantrekkelijkheid van meritocratie voor beleidsmakers verklaart: 'It legitimates difference, stimulates effort, and in doing so optimizes the allocation of reward.' (Mijs, 2016: 7). Uitgaan van meritocratische principes heeft volgens Mijs twee nadelige effecten: de verliezer krijgt de schuld van zijn eigen verlies en ten tweede dreigt de nadruk op 'verdienste' de invullingen van de begrippen 'behoefte' en 'gelijkheid' naar de achtergrond te dringen, wat grote gevolgen heeft voor het denken in termen van 'gelijke kansen' (Mijs, 2016: 13). Daarbij komt dat de grote aandacht in onderwijspolitiek voor het stimuleren van excellentie als uitvloeisel van meritocratisch denken leidt tot een negeren of zelfs minachten van de morele principes van 'gelijkheid' en 'behoefte' (Mijs, 2016: 17). Dit levert een beperkte, op economische rationaliteit gefundeerde blik op ten aanzien van fairheid in relatie tot onderwijs. Deze focus reduceert onderwijs tot een trainingsinstituut voor de arbeidsmarkt terwijl onderwijs in de optiek van Mijs meer zou moeten willen zijn. Van Hoof en Dronkers hebben dit al eerder, in de jaren '80, gesteld: onderwijs heeft een 'selectie- en een allocatiefunctie' voor de arbeidsmarkt. Het gevolg is dat op de arbeidsmarkt een gedifferentieerd aanbod van schoolverlaters verschijnt, die ten gevolge van hun verschillende gevolgde opleidingen en diploma's reeds voor het betreden van de arbeidsmarkt ongelijke kansen bezitten om aantrekkelijkere en beter beloonde arbeidsplaatsen te bereiken. Zo draagt het onderwijssysteem bij aan rechtvaardiging van bestaande sociale verhoudingen. Volgens Van Hoof en Dronkers is aan onderwijs een 'legitimeringsfunctie' verbonden. Verschillen in maatschappelijke positie van personen of groeperingen kunnen als acceptabel worden voorgesteld door een beroep te doen op de verschillen in het behaalde onderwijs die daaraan ten grondslag zouden liggen. Wanneer binnen het onderwijs iedereen 'gelijke kansen' heeft, kunnen de resulterende verschillen in het bereikte onderwijsniveau immers worden beschouwd als de weerspiegeling van 'onvermijdelijke' verschillen in individuele capaciteiten. Dit vormt de kern van een 'meritocratische ideologie' die de ongelijkheid in onze

samenleving moet rechtvaardigen en dit ten dele ook doet. De toegang tot vooraanstaande posities in de huidige samenleving wordt zonder voldoende opleiding steeds moeilijker. Buiten het onderwijs zijn er steeds minder mobiliteitsmogelijkheden (Hoof & Dronkers, 1980: 89). De groeiende expansie van het onderwijssysteem roept een nieuwe ontwikkeling op die deze expansie op haar beurt verder stimuleert, namelijk de noodzaak om langer en hoger onderwijs te volgen dan de ouders deden om de maatschappelijke status van het milieu van herkomst te kunnen vasthouden. Onderwijs is dan geen middel tot sociale stijging of grotere ongelijkheid van kansen, maar een middel om sociale daling te voorkomen (Hoof & Dronkers, 1980: 91).

Bij de adequaatheid van selectie ligt in Nederland historisch gezien al decennialang grote zorg van de overheid (zie ook hoofdstuk 4). Een tweede prioriteit ontstaat in Nederland als blijkt dat er sprake is van én overkwalificatie in relatie tot de arbeidsmarkt en daarmee samenhangende werkloosheid onder hoger opgeleiden in de jaren '80, én een tekort aan werknemers met een beroepsopleiding om de lagere functies op de arbeidsmarkt te vervullen. De Rooy stelt dat in het begin van de jaren zeventig 43% van de jongeren naar het beroepsonderwijs ging en dat dit percentage daalde tot 25% in 1995 (De Rooy, 2018: 183). Het aantrekkelijker maken van beroepsonderwijs met een aanpassing van het onderwijsstelsel in de vorm van het vmbo heeft een deel van het algemeen vormend onderwijs, in de vorm van de mavo, omgevormd naar meer beroepsgericht onderwijs in de vorm van vmbo-theoretische leerweg. Voor een verbetering van het imago is daar in de jaren negentig aan toegevoegd dat de route vmbo-mbo-hbo als 'koninklijk' moest worden gezien. In de praktijk betekent dit vooral een belemmering om over te stappen van vmbo naar havo. De eindtoets vervult een belangrijk rol als 'behoeder' van voldoende instroom in het vmbo. De toets zorgt er nu voor dat de helft van de leerlingen aangewezen is op de route van vmbo-mbo-hbo. Niet door met de eindtoets een vorm van geschiktheid voor beroepsgericht onderwijs te bepalen maar door vast te stellen dat de score niet toereikend is om te behoren tot die helft die voor algemeen vormend onderwijs in de vorm van havo/vwo in aanmerking komt (zie hoofdstuk 3).

Voor onderwijs als concept geldt geen schaarste. Dat er wel sprake is van schaarste is het gevolg van een bewust gecreëerde en in stand gehouden schaarste. Het belang dat gehecht wordt aan de bezetting van de onderste stromen in het onderwijs hangt samen met de economische belangen die horen bij de vervulling van de lagere functies op de arbeidsmarkt. Die functies moeten (ook) vervuld worden. In dit verband kan gesproken worden van de toedeling van 'lasten' in de terminologie zoals Elster die hanteert onder het motto 'iemand moet het doen'. Voor de beperking van de instroom aan de bovenkant van het onderwijsgebouw lijken met name financiële belangen een rol te hebben gespeeld: niet meer dure trajecten dan nodig zijn voor de economie. Dit sluit aan bij sociologisch onderzoek dat stelt dat iets vooral waarde heeft als er niet te veel mensen zijn die dat 'iets' hebben; positionele goederen krijgen hun waarde door schaarste (Hirsch, 1977). Hoe meer mensen aan het onderwijs deelnemen, hoe minder de waarde ervan zal zijn. Een onderwijsdiploma vermindert in waarde als te veel mensen dat diploma weten te bemachtigen. De grens van 20% voor toelating tot de meest begeerde vorm van onderwijs zorgt voor behoud van de waarde van dit diploma. Om waardeverlies te voorkomen kan het onderwijs het aanbod meer differentiëren en de toegankelijkheid ervan beperken. Zo wordt er nieuwe schaarste gecreëerd waar vooral de sterkeren, goed op de hoogte van wat er gaande is, van weten te profiteren.

Volgens De Rooy is dit wat er in het Nederlandse onderwijs is gebeurd in de afgelopen vijftig jaar: 'Naarmate het gewoner werd om kinderen naar het voortgezet onderwijs te sturen, groeide de behoefte aan meer en hoger onderwijs om de relatieve sociale afstand in stand te houden. Het categoriaal gymnasium groeide: van ruim twaalfduizend in 1974 naar ruwweg dertigduizend in 2016. De twintigste eeuw kenmerkt zich door een imposante opwaartse sociale mobiliteit, mede dankzij het onderwijs. Tussen 1990 en 2016 bijvoorbeeld is de bevolking in de leeftijd van 25 tot en met 64 jaar met een laag opleidingsniveau – dat wil zeggen zonder startkwalificatie – gedaald van 45 tot 23 procent. Meer in het algemeen ligt het Nederlandse opleidingsniveau boven het Europees gemiddelde. Maar de onderlinge verhoudingen tussen de verschillende sociale lagen zijn in Nederland opmerkelijk stabiel gebleven, mede dankzij de hierboven aangeduide mechanismen. Met vloed zijn alle boten gestegen, maar de grote boten steken nog steeds boven de kleinere uit.' (De Rooy, 2018: 242-3).

In de volgende paragraaf wordt onderzocht welke rol onderwijs heeft gehad in het verdelen van maatschappelijke posities, welke rol beleid hierin heeft gespeeld en hoe het beleid te karakteriseren is.

## 7.2 Verdelingen en selectie in onderwijs

Omdat onderwijs een plek is waar belangrijke sociale middelen worden verdeeld, heeft politiek filosoof Walzer in zijn boek 'Spheres of Justice' (1983) het belang onderstreept van het onderzoeken van onderwijs als een te onderscheiden 'sfeer' van rechtvaardigheid. Resh en Sabbagh werken rechtvaardigheid van verdelingen voor de onderwijscontext nader uit in hun bijdrage aan het 'Handbook of Social Justice Theory and Research' (2016). Zij onderscheiden vijf sub-sferen, van macro naar micro, die in de context van onderwijs op rechtvaardigheid van verdelingen moeten worden onderzocht: 1) het recht op onderwijs en hoe dit wordt gerealiseerd, 2) de toewijzing van onderwijsplekken, 3) de pedagogiek die in onderwijs gehanteerd wordt, 4) het overgaan of zittenblijven en 5) de relatie tussen docent en student. Voor elk van de vijf sub-sferen is het van belang de vraag te beantwoorden: wie krijgt wat van de onderwijsmiddelen, gebaseerd op welke principes? Voor rechtvaardiging van verdelingen meer in het algemeen stellen Resh en Sabbagh dat moet worden gezocht naar een balans van gelijkheid, behoefte en billijkheid. Voor onderwijs stellen zij dat het bij de toelating tot schaarse onderwijsplekken gaat om een combinatie van op billijkheid gebaseerde principes van gelijke kansen. Meritocratie geldt als uitgangspunt voor de verdeling van schaarse onderwijsplekken. Dit impliceert een rechtvaardiging van ongelijkheid mits die gebaseerd is op inspanning, prestaties en vermogen (Resh & Sabbagh, 2016: 354). Resh en Sabbagh wijzen erop dat een interpretatie van 'verdienste' als uitgangspunt voor rechtvaardiging in de uitvoering tot verschil kan leiden afhankelijk van culturele verschillen tussen landen. In de VS leidt het uitgangspunt tot een verdeling gebaseerd op scores op gestandaardiseerde prestatietests in combinatie met een IQ-score en in een land als Japan leidt het tot een groter belang dat wordt gehecht aan de inspanningen en het harde werken van een student (Resh & Sabbagh, 2016: 354). De verschillende manieren waarop het meritocratisch uitgangspunt van verdienste wordt toegepast en uitgewerkt, heeft gevolgen voor hoe verdelingen uitwerken. Dit maakt de keuzes die gemaakt worden bij de uitwerking van het concept van verdienste bijzonder relevant. Resh en Sabbagh vinden het verbazingwekkend dat er relatief beperkt aandacht is voor de uitwerking van rechtvaardigheidsprincipes voor de praktijk van onderwijs. Zij pleiten ervoor om onderwijskundige vraagstukken meer te framen als issues in het licht van rechtvaardige ver-



delingen met als belangrijke overkoepelende vraag: wat is een rechtvaardige school? (Resh & Sabbagh, 2016: 362).

In een advies aan de Onderwijsraad over de toegankelijkheid van het onderwijs, beschouwt Wilbrink in het afsluitende hoofdstuk de toegankelijkheid van het Nederlandse onderwijs vanuit meritocratisch perspectief. Voor onderwijs is, in tegenstelling tot bijvoorbeeld gezondheidszorg, niet gekozen om 'behoefte' als uitgangspunt te nemen voor selectie maar 'verdienste' waar volgens Wilbrink de redenering mee gepaard lijkt te gaan dat 'wie eerder heeft bewezen goed gebruik te hebben gemaakt van onderwijs, daarmee het recht verwerft nog meer onderwijs te volgen'. Dit laat zich volgens Wilbrink, verwijzend naar de Engelsman Winch, moeilijk verdedigen volgens in de samenleving gehanteerde juridische en morele maatstaven. Winch stelt dat het 'heel gebruikelijk is maar onverdedigbaar dat 'beter presterende' leerlingen binnen school op allerlei manieren worden voorgetrokken of privileges krijgen' (Wilbrink, 1997: 348). Wilbrink stelt vervolgens ook de vraag of meritocratisch denken wel passend is, gegeven doel en functie van onderwijs: 'Is de meritocratische vooronderstelling dat eenieder in staat is het lot in eigen hand te nemen, en daar dus vervolgens ook bij selectieve momenten op mag worden afgerekend, in empirische zin juist? Biedt het onderwijs in deze zin wel gelijke kansen?' (Wilbrink, 1997: 349). Hij stelt de principiële vraag of het mogelijk is om kinderen met gelijke capaciteiten, maar overigens in heel verschillende omstandigheden verkerend, gelijke kansen op onderwijs te bieden. Wilbrink wijst erop dat het sociale milieu ('nurture') grenzen stelt aan de meritocratisering (Wilbrink, 1997: 354). Als ongelijkheid zo groot is dat van achterstand moet worden gesproken, dan falen volgens Wilbrink op meritocratische leest geschoeide interventies (Wilbrink, 1997: 355). Hij stelt dat het een 'fictie is dat de leerling volledig verantwoordelijk is voor zijn of haar resultaten, en daar dus op mag worden afgerekend. Het hoeft geen nadere toelichting dat juist dat een bedreiging is voor de gelijke kansen van degenen die van huis uit minder bagage voor deze overlevingstocht meekrijgen dan anderen.' (Wilbrink, 1997: 368).

### 7.2.1 Vroege selectie

Nederland heeft een traditie van selecteren aan het eind van het primair onderwijs om leerlingen vervolgens door te laten stromen naar een breed aanbod van vervolgonderwijs, de delta, zoals Idenburg het formuleert in 1960: 'Gelijk het water van een rivier stroomt een jaargang kinderen naar de delta, waar hij zich vertakt in een samenstel van grotere en kleinere waterlopen met of zonder communicatie om vroeg of laat de zee te bereiken, dat wil zeggen de maatschappij binnen te gaan' (Idenburg, 1960: 476). 'Vanwaar deze scherpe schifting? Geschiedt zij als prikkel voor de leerlingen om zich tot het bereiken van vastgestelde normen naar vermogen in te spannen? Dit is stellig een van de overwegingen. Daarnaast streeft de school ernaar haar plaats in een groot en samenhangend geheel van opvolgende onderwijsinrichtingen waardig te vervullen. Maar niet in het minst selecteert de school voor de maatschappij. Haar docenten zijn belast met een taak, welke hun als opvoeders in de grond vreemd moet zijn: mee te spelen in het veelomvattend maatschappelijk schiftingsproces, waardoor jonge mensen worden gewaardeerd met het oog op de posities, welke zij straks in de samenleving zullen kunnen gaan vervullen.' (Idenburg, 1960: 491). Idenburg vat hiermee de kern van onderwijs samen: Onderwijs als prikkel voor leerlingen om zich in te spannen en onderwijs om arbeidsmarkt en samenleving van voldoende, diverse, werknemers te voorzien aansluitend bij de behoeften.

Het onderwerp van 'vroege selectie' komt steeds terug op de beleidsagenda in Nederland, de laatste jaren vooral onder druk van de OECD die er in de rapportages 'Education at Glance' regelmatig op wijst dat vroege selectie nadelige effecten heeft. Alle bewegingen die gemaakt zijn om iets van uitstel van selectie te bewerkstelligen zijn steeds weer uitgedraaid op een terugkeer naar het bestaande systeem van vroege streaming en de daaraan gerelateerde, zich steeds verder versterkende en vervroegende selectie. In Nederland wordt gezocht naar 'oplossingen' voor de vroege selectie binnen het bestaande stelsel zoals ook weer blijkt uit het rapport 'Vroeg of laat? Advies over de vroege selectie in het Nederlandse onderwijs' van de Onderwijsraad (Onderwijsraad, 2010). Daarin zegt de Raad geen voorstander te zijn van uitstel van selectie met de volgende redenering: 'Op dit moment zijn er onvoldoende aanwijzingen dat een breed en voor alle scholen verplicht uitstellen van het selectiemoment leidt tot een verbetering van schoolprestaties voor alle leerlingen. De vele wetenschappelijke studies geven onvoldoende steun voor een dergelijke algemene maatregel.' (Onderwijsraad, 2010: 9). De Raad pleit daarentegen voor aanpassingen binnen het stelsel zoals vroegtijdige opsporing van leerachterstanden, kop- of voetklassen waarin leerlingen met potentie uitstel van selectie wordt geboden, meer combinatieklassen zoals vmbo-t/havo, stapelen, combinatieklassen voor bepaalde vakken zoals gymnastiek, en experimenten met juniorcolleges (10-14 jaar) (Onderwijsraad, 2010: 10). Wat in dit advies opvalt ten opzichte van eerdere discussies is de argumentatie: uitstel van selectie wordt nu in verband gebracht met de verbetering van schoolprestaties. De OECD en ook eerdere discussies in Nederland relateren uitstel van selectie vooral aan het verbeteren van de kansen voor leerlingen die met achterstanden het onderwijs instromen. Zij zijn gebaat bij een latere selectie want nog altijd blijven de kansen van leerlingen uit achterstandsposities op een hoog schooladvies achter bij die van kinderen van hoger opgeleiden.

### 7.2.2 Selectie en afkomst

Mandemakers heeft de instroom in HBS en Gymnasium van 1880 tot aan de invoering van de Mammoetwet onderzocht. Uit zijn analyses blijkt dat de onderwijsdeelname in die periode weliswaar enorm is gegroeid maar dat de relatieve instroom in het v.h.m.o. daar niet mee in de pas loopt; de groei van het aantal deelnemers komt relatief meer terecht bij de andere vormen van vervolgonderwijs. Bewust is de deelname aan het v.h.m.o. gereguleerd zoals bijvoorbeeld blijkt uit de zogenaamde 'Stop-wet' die in 1924 aangenomen wordt om de groei van het v.h.m.o. te dempen. Omstreeks 1880 gaat 4% van de jongens van 13 naar het v.h.m.o., in 1920 is dit 8% en in 1965 is dit 20% (Mandemakers, 1996: 286). In 1880 was het aandeel kinderen afkomstig uit een arbeidersmilieu op gymnasium en hbs minder dan één procent, in 1920 was dit vier procent en in 1965 kwam dit uit op negentien procent. Dit is evenveel als het aandeel kinderen van 'hogere employees' zoals Mandemakers de kinderen van hoger opgeleiden noemt. Desondanks had in 1965 een jongen uit deze laatste categorie nog steeds bijna zeven keer zoveel kans om aan het v.h.m.o. deel te nemen als een zoon van een arbeider (geschoold of ongeschoold) (Mandemakers, 1996: 261). Dit omdat de laatste categorie veel groter was dan de eerste. De aanzienlijke kansverbetering voor de lagere sociale milieus is volgens Mandemakers niet alleen een gevolg van de uitbreiding van de capaciteit van het v.h.m.o., maar zeker ook van de daling van het gemiddeld kinderaantal (Mandemakers, 1996: 261). Tegelijkertijd laten de ontwikkelingen zien dat het aandeel leerlingen dat naar het v.h.m.o. doorstroomt in een periode van 50 jaar minder explosief verloopt dan de instroom in mulo (Mandemakers, 1996: 82).

De invloed van het sociaal milieu is sinds de Mammoetwet weliswaar afgenomen maar eind jaren '90 heeft het milieu waarin het kind opgroeit nog steeds een aanzienlijke invloed op de schoolkeuze: eind jaren negentig gaat 42 procent van de kinderen met hoogopgeleide ouders en 11 procent van de kinderen met laagopgeleide ouders naar het vwo (CBS). Er is dus wel sprake van een toename van deelname maar de relatieve verschillen zijn stabiel.

### 7.2.3 Toegankelijkheid: twee modellen

In een artikel uit 1960 vergelijkt de Amerikaanse socioloog Turner het Amerikaanse en Britse systeem van voortgezet onderwijs. Hij stelt dat de beide systemen fundamenteel van elkaar verschillen. Hij stelt dat het Amerikaanse systeem te definiëren is als een systeem waarin 'contest mobility' de norm is: de toelating is open maar na binnenkomst in het systeem is er sprake van harde concurrentie. Het Britse systeem definieert hij als een systeem waarin 'sponsored mobility' de norm is: er is een zware selectie aan de poort maar eenmaal binnen wordt men meegenomen (Turner, 1960). Mandemakers heeft in zijn onderzoek het Nederlandse voortgezet onderwijssysteem vergeleken met beide modellen. Mandemakers stelt dat het onderwijssysteem in Nederland tot 1960 gekarakteriseerd kan worden als een 'sponsored' systeem. Uit de analyse van de cohorten van 1880-'81 en 1920 maakt Mandemakers op dat als men eenmaal binnen is in het v.h.m.o. milieu nog maar betrekkelijk weinig invloed heeft. De kansen op succes zijn vergelijkbaar ongeacht de achtergrond. Dit ligt anders voor het cohort van 1960. In 1965 heeft 20% van alle dertienjarige jongens toegang gekregen tot het v.h.m.o. Volgens Mandemakers sturen alle sociale lagen dan al hun enigszins v.h.m.o.-geschikte kandidaten naar het v.h.m.o. Omdat het daadwerkelijk behalen van een diploma steeds relevanter is geworden, ontstaat er vanaf de jaren zestig meer competitie gedurende het onderwijs, wat in het voordeel is van kinderen van hoger opgeleiden: 'In 1960 is dit anders: men komt relatief gemakkelijk binnen, maar eenmaal binnen worden bij wijze van spreken alle (milieu)registers weer opengezet.' (Mandemakers: 1996: 287). Uit de analyse van de gegevens blijkt dat bij de vergelijking over de periode 1880-1965 de samenhang tussen schoolprestaties en sociaal milieu in de loop van de tijd niet is afgenomen maar juist in 1965 het grootst was (Mandemakers: 1996: 288).

De Groot spreekt over het door Turner gemaakte onderscheid van onderwijsmodellen in 'sponsored mobility' versus 'contest mobility' als respectievelijk het 'expeditiemodel' ('samen uit, samen thuis') in tegenstelling tot als het 'veldloop- of hordenloopmodel' (De Groot: 1972). De voortdurende beoordeling tijdens het onderwijs op basis van daarvoor in de ogen van De Groot niet geschikte toetsinstrumenten is een manier van werken waar De Groot zich in zijn boek 'Vijven en Zessen' (1966) tegen verzet. Met behulp van tests is het volgens De Groot mogelijk een goede, eerlijke selectie aan het begin van het voortgezet onderwijs vorm te geven waarna er sprake kan zijn van 'selectievrij' onderwijs en docenten de taak hebben de leerlingen naar de eindstreep te brengen onder het motto 'Zijn onderwijs is aangenomen werk' (De Groot, 1966: 213). De twee functies van onderwijs, die van het bevorderen van het leerproces bij leerlingen enerzijds en het controleren en beoordelen van leereffecten bij individuele leerlingen anderzijds, zijn volgens De Groot 'veeleer strijdig dan organisch verbonden' (De Groot, 1966: 213). De permanente selectie waarvan sprake is in het v.h.m.o. is De Groot een doorn in het oog. Hij vat het middelbaar onderwijs in 1966 als volgt samen: 'Wij geven onderwijs in een soort niemandsland, waarin de leerlingen blijven zweven tussen aanvaard-zijn en onaanvaardbaar-blijken tot aan het einde, bitter of gelukkig, van hun schoolcarrière.' (De Groot, 1966: 215). Het hordenloopmodel

zoals De Groot het hier beschrijft is niet het onderwijsmodel dat hij voor ogen heeft. Volgens Mertens is De Groot van mening dat het onderwijs zo moet worden ingericht dat 'na een zorgvuldige toelating, de meeste leerlingen na een aantal jaren, liefst gelijk aan de cursusduur, slagen', het expeditiemodel (Mertens, 2016: 5). Bij dit model hoort een selectieprocedure die in de optiek van De Groot moet gebeuren met behulp van gestandaardiseerde achievementtests waar volgens De Groot in de Verenigde Staten dan al goede successen mee zijn geboekt en waarin hij groot vertrouwen heeft.

In een artikel in het tijdschrift *Paedagogische Studiën* (2004) analyseert socioloog Meijnen beide modellen veertig jaar later nog eens voor het voortgezet onderwijs in relatie tot kansengelijkheid vanuit meritocratisch perspectief. Het 'expeditiemodel', volgens Meijnen, 'kapitaliseert op de stabiliteit van persoonskenmerken en op de mogelijkheid de potentie ervan al op een jonge leeftijd valide te kunnen meten. Door talent op te sporen en zorgvuldig te begeleiden binnen specifieke instellingen, ontstaat een optimale afstemming tussen persoonlijke capaciteiten en maatschappelijke structuren en wordt een efficiënte organisatie van werken en leren mogelijk.' (Meijnen, 2004: 82). Bij een 'veldloop- of hordenloopmodel' wordt, volgens Meijnen, 'weinig geloof gehecht aan stabiliteit van persoonskenmerken en wordt de kans groot geacht dat foutieve selectiebeslissingen leiden tot verlies aan talent, dan wel overmatig investeren in ten onrechte toegelatenen. Het bijbehorende onderwijssysteem kenmerkt zich door open toelating en permanente selectie. Leerlingen van uiteenlopend niveau worden zo lang mogelijk bijeengehouden ter voorkoming van systeemeffecten.' (Meijnen, 2004: 83). Bij het expeditiemodel zullen, volgens Meijnen, op termijn de kosten en baten van verkeerde selectiebeslissingen nauwelijks meer zichtbaar zijn omdat het gebaseerd is op zichzelf waarmakende voorspellingen. Dit verhindert de toetsing van de validiteit van de selectie (Meijnen, 2004: 83). Het meritocratisch gehalte van dit systeem is te bevragen omdat volgens Meijnen een in de vroege jeugd gemeten ontwikkelingsprofiel (capaciteiten, inzet, belangstelling) waarin de invloed van tal van sociale, culturele en economische factoren bepalend is, gevolgen heeft voor de instroompopulatie in het voortgezet onderwijs: 'Opleidingstypen in het voortgezet onderwijs zijn reeds vanaf het begin sociaal "gekleurd". Leerlingen uit de hogere sociale lagen stromen door naar havo/vwo, leerlingen uit de lagere statusgroepen, van autochtone en van allochtone herkomst, bevolken het vmbo en dan met name de "lagere" leerwegen. Eenmaal geselecteerd voor deze routes, zijn de achtergrondkenmerken niet dominant meer; de belangrijkste intellectuele en sociale selectie heeft dan al plaatsgevonden.' (Meijnen, 2004: 87). Een verdere vervroeging van selectie door het toekennen van selectiewaarde aan een leerlingvolgsysteem zal dit fenomeen naar verwachting de komende jaren verder versterken.

Schuyt analyseert de situatie in het Nederlandse onderwijs op basis van vergelijkbare theoretische uitgangspunten ten aanzien van kansengelijkheid. Hij stelt dat er bij het verbeteren van gelijkheid in het onderwijs theoretisch gezien in twee richtingen gedacht kan worden: 'ófwel de toegangswegen tot de beste schoolsoorten drastisch voor eenieder verbeteren maar de schoolsoorten als streng classificatiesysteem handhaven, ófwel de informele maatschappelijke invloed op de toegangswegen laten voortbestaan, maar dan de strenge selectie via het classificatiesysteem verminderen. De selectie aan het begin wordt dan omgezet in selectie onderweg of aan het eind.' (Schuyt, 1991: 331). Nederland heeft in de optiek van Schuyt een combinatie gekozen van twee elkaar versterkende eigenschappen. Er is volgens hem namelijk sprake van

'informele maatschappelijke invloed op de toegangswegen' die gepaard gaat met een 'streng hiërarchisch classificatiesysteem' (Schuyt, 1991: 331). Hij wijst op het belang van formele classificaties, omdat bij afwezigheid ervan meer informele oordelen en beoordelingen de plek zullen innemen, wat volgens Schuyt kan leiden tot wat hij noemt een 'keiharde triade: maatschappelijk bevoordeeld – beter beoordeeld – opnieuw maatschappelijk bevoordeeld' (Schuyt, 1991: 332). Tegelijkertijd is het belang van formele classificaties in het licht van die triade alleen maar van waarde in het kader van rechtvaardigheid als de toegang tot die formele classificaties voor alle betrokkenen gelijkkelijk aanwezig is. Hier ziet Schuyt een belangrijke rol voor de overheid: 'Een overheid die wel werkt met formele classificaties maar aan deze toegangsvoorwaarden weinig of geen aandacht besteedt, roept met een dergelijke wancombinatie meer problemen op met betrekking tot rechtvaardigheid dan zij probeert op te lossen.' (Schuyt, 1991: 337).

Nederland heeft een strikte toelating in combinatie met een doorgaande, permanente vorm van selectie in de vorm van overgaan, zittenblijven of afstromen gedurende het voortgezet onderwijs waarmee er alsnog sprake is van een 'triade' waarvan de rechtvaardigheid bevestigd kan worden.

#### 7.2.4 Toedelende en remediërende selectie

Bakker en Lehning stellen dat onderwijs in Nederland gezien kan worden als 'gate-keeper'. Het onderwijs is volgens hen van oudsher de institutie bij uitstek waar selectie van mensen plaatsvindt en die een belangrijke rol speelt in het verdelen van maatschappelijke posities en kansen. Bakker en Lehning maken daarbij onderscheid in twee vormen van selectie die in het onderwijs terecht zijn gekomen. Enerzijds moet in onderwijs 'toedelende selectie' plaatsvinden waarmee bedoeld wordt dat wordt verwacht, vooral vanuit het arbeidssysteem, dat in onderwijs een zodanige voorbereiding en selectie van mensen plaatsvindt dat voldaan wordt aan de 'manpower' behoefte van het arbeidssysteem. Onderwijs wordt daarmee ook object van overheidsinterventie. In het overheidsbeleid wordt onder meer getracht het onderwijs zodanig in te richten dat (meer) wordt bijgedragen aan het voldoen aan de vraag van het arbeidssysteem, het stimuleren van economische groei, bestrijding van de werkloosheid en dergelijke (Bakker & Lehning, 1991: 28). 'Achievement' heeft in de loop van de tijd de plaats ingenomen van 'ascription' als criterium voor verdeling. Waar het onderwijs in een eerdere fase werd ingezet om een bestaande standbepalende verdeling te legitimeren krijgt het nu zelf een verdelende functie. Toedelende selectie wordt daarmee een taak van het onderwijs. Daarmee moet onderwijs zich zowel inspannen om de cognitieve begaafdheid van leerlingen te ontwikkelen alsook op basis daarvan te selecteren voor vervolgopleidingen en toekomstige posities. Het vertrouwen in de mogelijkheden voor het vaststellen van cognitieve begaafdheid groeit in de eerste helft van de vorige eeuw mede door de ontwikkeling van de psychologie, in het bijzonder de testpsychologie (Bakker & Lehning, 1991: 32). Die rol verhoudt zich lastig tot de gedachte dat onderwijs ingezet moet worden als instrument ter bestrijding van ongewenste maatschappelijke selectieprocessen. Op basis van de onaanvaardbaar geachte maatschappelijke ongelijkheid heeft het onderwijs de taak groepen van leerlingen die een bij voorbaat geringe kans op succes hebben in het meritocratische selectieproces differentieel te behandelen. Met deze remediërende taak moet het onderwijs ook dienen als instrument ter vermindering van de bestaande maatschappelijke ongelijkheid bij (Bakker & Lehning, 1991: 34). Bakker en Lehning spreken in dit verband over 'remediërende selectie': selectie bedoeld om bepaalde ongelijkheid op te heffen. Toedelende en remediërende selectie worden zo op elkaar betrokken volgens Bakker en Lehning. De maatschappelijke legitimatie van selectie

binnen het onderwijs op basis van alleen cognitieve capaciteiten en prestaties gaat volgens hen samen met de gedachte dat eenieder gelijke kansen dient te hebben. Iedereen moet vanuit een zo gelijk mogelijke uitgangspositie aan het selectieproces binnen het onderwijs kunnen deelnemen.

Met de waarde die gehecht wordt aan een diploma, is de rol van onderwijs waar diploma's wel of niet worden verkregen, versterkt. Van Hoof en Dronkers (1980) wijzen op een drietal inzichten die hieraan ten grondslag hebben gelegen: ten eerste dat onderwijs een belangrijk allocatiecriterium blijkt te zijn op de arbeidsmarkt omdat onderwijs als het meest betrouwbare etiket voor de gebruikswaarde van een arbeidskracht geldt, ten tweede dat een onderwijsetiket wordt beschouwd als een indicatie voor uiteenlopende kenmerken van het aanbod (cognitief, persoonlijkheid, gedrag) en ten derde dat de arbeidsmarkt zich kan richten op de selectie op basis van andere criteria dan die aan het onderwijsdiploma kunnen worden toegeschreven (Van Hoof & Dronkers, 1980: 55-56).

Het streven naar gelijke kansen als doel van overheidsbeleid, wordt in de jaren negentig gepositioneerd binnen het economisch belang, zoals blijkt uit de nota van het ministerie van OCW 'Niet zomaar gekozen: over kerntaken, leerwegen en overheidsbudget' (1993). Aan de ene kant zijn er de economische belangen die met onderwijs moeten worden gediend, waarbij een afweging nodig is van de belangen van het individu versus de belangen van de samenleving: 'Enerzijds bereidt het individuen voor op hun maatschappelijk functioneren, en anderzijds draagt het als onderdeel van de maatschappelijk-economische infrastructuur bij aan economische groei en welvaart. De functie van voorbereiding van individuen op hun maatschappelijk functioneren is te onderscheiden naar kennisoverdracht, individuele ontplooiing, sociale en staatkundige vorming, socialisatie (overdracht van waarden en normen), cultuuroverdracht en beroepsvoorbereiding. De bijdrage van economische groei en welvaart komt deels tot stand via de uitwerking op individuen: de groei van 'menselijk kapitaal' is zonder onderwijs moeilijker te realiseren.' (OCW, 1993a: 7). Aan de andere kant stelt het ministerie in dezelfde nota dat het met zijn beleid rechtvaardig wil handelen hieraan de volgende invulling gevend: 'De belangrijkste reden voor de overheidsbemoediging met het onderwijs is evenwel het rechtvaardigheidsmotief. Vanwege het grote belang van onderwijs voor de toekomst van elk individu rekent de overheid het tot haar centrale taak te voorkomen dat sociale ongelijkheden zich via het onderwijs reproduceren. Het bevorderen van de toegankelijkheid van het onderwijs blijft bij voortduring een belangrijk, zo niet het belangrijkste doel.' (OCW, 1993a: 8).

Rechtvaardigheid en meer toegespitst het concept kansengelijkheid wordt vanaf de jaren negentig vooral uitgewerkt in het licht van 'procedurele fairness', zodat er in onderwijs sprake kan zijn van een 'eerlijke' competitie. Daarbij wordt vooral waarde gehecht aan het creëren van een 'level playing field' als voorwaarde voor de waarde die kan worden toegekend aan de meritocratische wedstrijd die het onderwijs moet zijn. Voor de winnaars van die competitie is de uitkomst alleen maar waardevol als het idee bestaat dat de wedstrijd eerlijk is verlopen en dat zij de terechte winnaars zijn. En de verliezers van deze wedstrijd worden op die manier niet zomaar verliezers maar verliezers die de kans hebben gehad om te winnen, maar helaas niet de besten waren. Schaarste is een voorwaarde voor het legitimeren van een wedstrijdmodel en het streven naar een toegankelijk onderwijsstelsel is een voorwaarde om de wedstrijd te kunnen spelen, zodanig dat de winnaars als winnaars worden gezien en de verliezers als verliezers, zodat de uitkomsten van de wedstrijd als legitiem en eerlijk kunnen worden beschouwd.

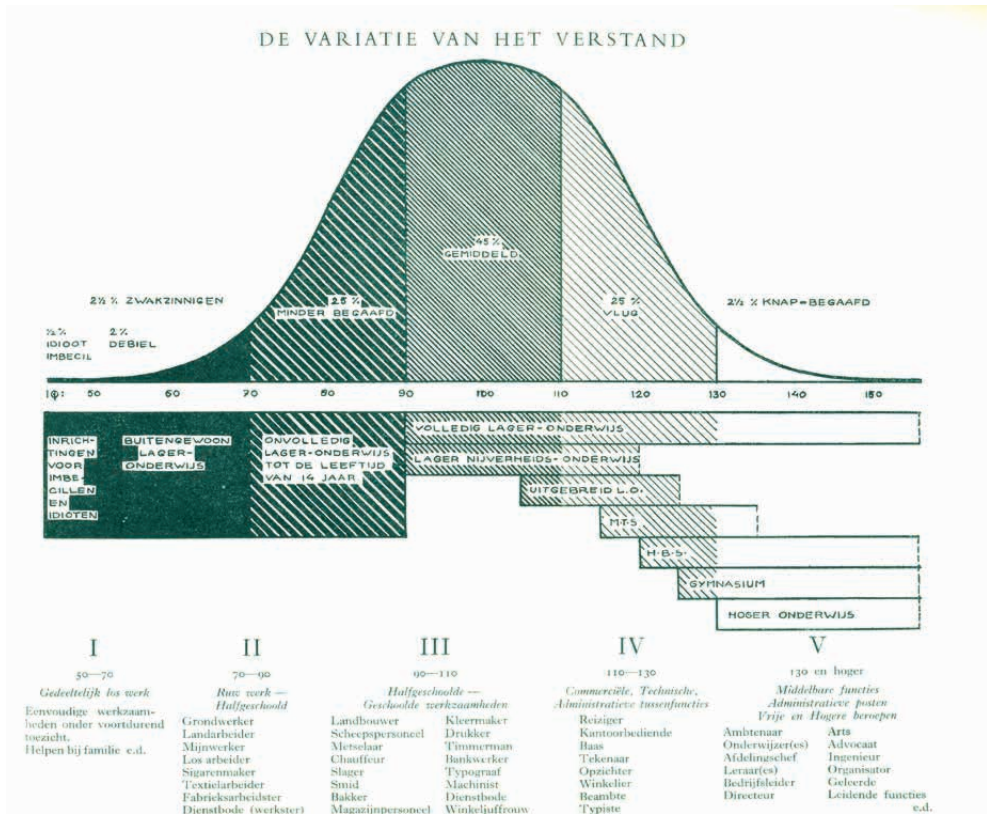
### 7.3 Verdelingen en testen in de praktijk

In lijn met de teneur die Elster in verdelingsvraagstukken in de afgelopen decennia waarneemt, geldt voor de verdeling van de plekken in het Nederlandse onderwijs ook dat wordt gestreefd naar oordelen die gebaseerd zijn op objectief toe te passen maten: 'Hoe beter objectieverbaar en niet manipuleerbaar een verdeelsysteem is, hoe minder kwetsbaar.' (Elster, 1992: 169). De waarde die in Nederland wordt toegekend aan de eindtoets in combinatie met de uitkomsten van het LVS passen binnen de door Elster gesignaleerde trend in samenlevingen om van discretionaire beslissingen in de richting te gaan van mechanische formules of onpersoonlijke mechanismes. Jarenlang is de uitvoering van de verdeling de verantwoordelijkheid geweest van het uitvoeringsniveau van de 'tweede orde': de scholen. Zij bepaalden zelf de plaatsing, al dan niet met behulp van tests en in overleg met ouders en kinderen. Met de wetswijziging van 2011 is het verdeelsysteem deels verplaatst naar het 'eerste orde niveau': de overheid heeft zelf regie over de verdeling. Volgens strikte richtlijnen moet het verdeelmechanisme worden toegepast en de overheid controleert de uitvoering ervan. Omdat de verdeling mechanisch is georganiseerd is, is inzicht en controle eenvoudig mogelijk gemaakt. Voor de toedeling van onderwijsplekken is gekozen voor een 'onderscheidend assessment' in de terminologie van Elster. In de beginjaren van de eindtoets is steeds gewezen op het feit dat 'trainen' voor de eindtoets niet zinvol zou zijn; het zou de meting verstoren. Met de belangen die op dit moment met de plaatsing gepaard gaan, voor school én leerling, is een trend waarneembaar om het individu te stimuleren acties te ondernemen die de score en daarmee de toewijzing beïnvloeden. Soms gebeurt dit actief door de school, al dan niet onder druk van ouders, maar ook ouders zelf kiezen er steeds vaker voor dat te doen wat nodig is om hun kinderen zo goed mogelijk voor te sorteren voor hun plaatsing in het voortgezet onderwijs (Elffers, 2018: 9). Zo heeft de selectie in Nederland in de laatste decennia wel degelijk een ander karakter gekregen. Een objectief te beoordelen, onderscheidend assessment is niet langer vrijblijvend maar wordt nu voorgeschreven. En er wordt op toegezien. Dit leidt tot extra inspanningen van scholen en leerlingen om te proberen de scores te verhogen waarvan bekend is dat scores met training te beïnvloeden zijn. Gezien het relatieve karakter van de eindtoets, gaat stijging van de scores van de een per definitie ten koste van de positie van een ander. Zo dient selectie een tweede doel: het vergroten van de inspanningen van scholen en leerlingen om te behoren tot de besten.

#### 7.3.1 Toedeling en intelligentie

Intelligentietests hebben zich in de loop van de 20e eeuw ontwikkeld tot instrumenten waarmee niet alleen begaafdheid wordt vastgesteld maar waarmee tevens de toelating tot de arbeidsmarkt wordt gereguleerd. In zijn boek 'Tests op school' (1952) laat Prak zien hoe IQ's, die per definitie normaal verdeeld zijn, gerelateerd worden aan categorieën functies op de arbeidsmarkt en aan onderwijstypen (Prak, 1952: 232).

Ook in Engeland is aan het begin van de 20e eeuw vertrouwen in intelligentietests en de potentie daarvan. Voor de toelating tot de 'Grammar School' wordt na de Tweede Wereldoorlog besloten dat de beste twintig procent van de leerlingen, gebaseerd op een intelligentietest, toegang krijgen tot deze exclusieve vorm van onderwijs. Stellwag stelt in 1955 dat de schaarste van onderwijsplekken in Engeland de selectie beïnvloedt: 'Hierdoor moet de toelating wel het karakter krijgen van een vergelijkend examen, waarbij de knapste leerlingen aan de Grammar School worden toegewezen.' (Stellwag, 1955: 147). Standaert stelt dat de druk van dit zogenaamde



Figuur 6 Variatie van het verstand

elf-plus examen voor vele ouders (en kinderen) een nachtmerrie was. Het leidde tot een terugslageffect op het voorgaande onderwijs wat er volgens Standaert vervolgens weer toe leidde dat de meeste lagere scholen al vanaf zeven jaar aparte groepen vormden voor 'slimme' leerlingen om hen voor te bereiden op het elf-plus-examen (Standaert, 2016: 4).<sup>25</sup> Ook in de Verenigde Staten is het meten van intelligentie, met behulp van de SAT-test, in die periode ingevoerd als vorm van toelating tot universiteiten. De SAT-score speelt nog altijd een belangrijke rol bij de toelating tot de prestigieuze universiteiten (Sandel, 2020).

Het denken vanuit intelligentie in combinatie met inzet als criterium voor de verdeling van plekken in het voortgezet onderwijs ligt ook in Nederland aan de basis van de verdeling van onderwijsplekken (zie ook hoofdstuk 4). Het percentage leerlingen dat in 1965 wordt geplaatst in het v.h.m.o. in Amsterdam ligt rond de twintig procent. Dit is voor de voorloper van de eindtoets richtpunt voor de constructie. De keuze voor het werken met percentielgroepen van twintig procent sluit hierbij aan. Het percentage van twintig procent is een percentage dat in Engeland dan al ruim twintig jaar gangbaar is voor de selectie van 'de besten'. In Nederland is dit percentage vanaf de jaren zestig nog altijd het uitgangspunt voor het krijgen van een vwo-advies. Het werken met een test die leerlingen vergelijkt en ordent in percentielgroepen maakt het

25] Vanaf 1964 zijn in Engeland meer comprensieve vormen van onderwijs ingevoerd en is de 'grammar school' gemarginaliseerd alhoewel deze 'public schools' voor de Engelse elite nog altijd een aantrekkelijke en belangrijke route zijn.



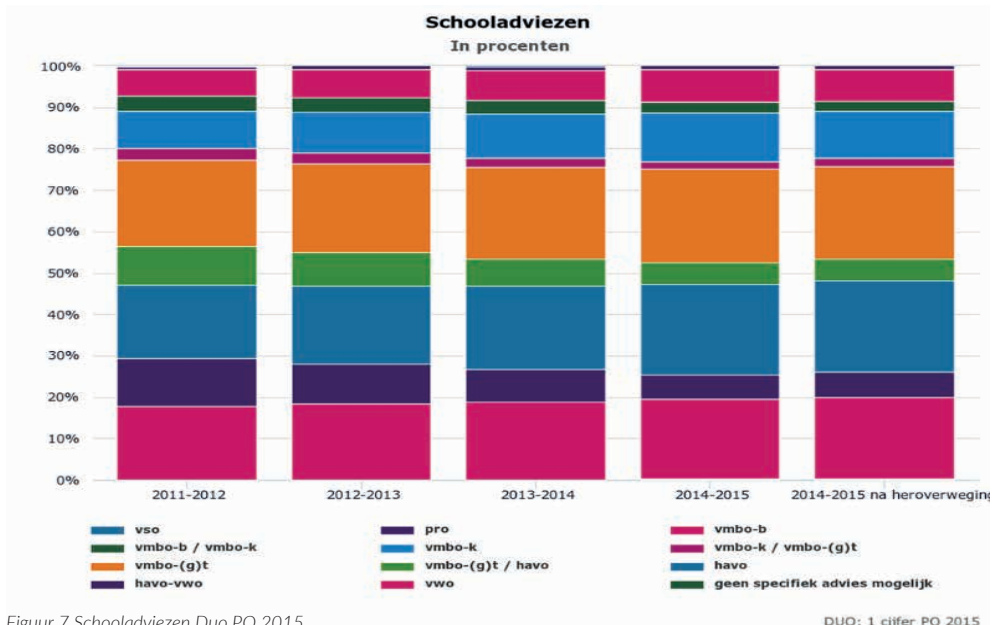
eenvoudig de hoogst scorende twintig procent aan te wijzen. Dit geldt ook voor de laagste vijftig procent. De min of meer jaarlijks vaste toewijzing van de leerlingpopulatie aan vormen van beroepsonderwijs in combinatie met beperkte overstapmogelijkheden, draagt zorg voor de realisatie van een stabiele instroom van arbeidskrachten richting arbeidsmarkt. Sinds de invoering van het vmbo is deze instroom stabiel: ongeveer de helft van de leerlingen wordt geadviseerd een traject te volgen in een van de vormen van het vmbo. Voor de havo is het beeld niet constant geweest. In eerste instantie was het havo-advies ondergebracht in het percentage van havo-vwo. Dit is wellicht toe te schrijven aan het feit dat de havo bij de invoering van de Mammoetwet een wat onduidelijke status had: was het wel of geen hbs-niveau? Het percentage leerlingen dat tot aan de invoering van de Mammoetwet een v.h.m.o.-advies kreeg, was ongeveer twintig procent. Bij het uitkristalliseren van de structuur van het hoger onderwijs, hebben de toeleidende rol die daarbij gezien wordt voor havo en de afspraken op Europees niveau voor de groei van het hoger onderwijs, geleid tot een aanpassing van de 'verdeelsleutel' voor instroom in het voortgezet onderwijs die sinds het begin van deze eeuw, na invoering van het vmbo, nu tamelijk stabiel is: vijftig procent van de leerlingen krijgt een beroepsgerichte leerweg aangeraden, dertig procent een advies gericht op een leerweg richting havo/hbo (adviezen vmbo-t/havo) en twintig procent een vwo-advies met daaraan gekoppeld de mogelijkheden voor directe doorstroming naar de universiteit. Deze percentages worden ook gehanteerd voor de toekomstvoorspellingen die het ministerie van OCW uitvoert ter onderbouwing van de meerjarige onderwijsbegroting, de zogenaamde referentieramingen. In de referentieraming van 2019 is te zien dat tot en met 2025 het ministerie van OCW ervan uit gaat dat aan de instroomverdelingen zoals die nu gelden niet veel zal veranderen, behalve veroorzaakt door een verwachte krimp van de algehele leerlingpopulatie (zie referentieraming 2019, OCW, voortgezet onderwijs).<sup>26</sup>

### 7.3.2 Eindtoets, LVS en verdelingen

De eindtoets en de toetsen van het LVS ordenen leerlingen op basis van percentielscores. Prestaties van leerlingen krijgen waarde door de vergelijking van scores met andere leerlingen. De score van alle leerlingen wordt via een lineaire transformatie omgezet in een standardscore oplopend van 501 tot 550 waarbij 535 het gemiddelde is (CvTE, 2015). In combinatie met het streven naar de realisatie van een normaalverdeling, zorgt dit voor een stabiele spreiding van leerlingen over het gehele scoregebied. De daaraan gekoppelde categorieën voor schooladvisering zorgen vervolgens voor stabiele leerlingenaantallen per adviescategorie. De stabiliteit van de verwijzingen is te zien in figuur 27 van Dienst Uitvoering Onderwijs (DUO)(2015). De belangrijkste factor die zorgt voor 'afwijkingen' van dit van tevoren doordachte patroon, zijn de oordelen van scholen ten aanzien van de advisering. Als daar grote afwijkingen zijn die door ouders en leerlingen gevolgd worden, bestaat de kans dat de uiteindelijke selectie en plaatsing er net iets anders komt uit te zien dan met het model van de eindtoets wordt nagestreefd.

Bij het wetsvoorstel voor de invoering van de verplichting van de afname van een eindtoets is besloten de afname van de eindtoets te laten plaatsvinden nadat het schooladvies al is gegeven en de aanmeldingsperiode voor het voortgezet onderwijs is verstreken. Dit onder druk van een deel van de Tweede Kamer, die daarmee het gegroeide belang van de waarde van de eindtoets als momentopname, wil terugbrengen. Dit is een besluit waarbij het schooladvies aan belang lijkt te winnen. Met de verplichting tot het gebruik van het LVS voor het volgen van de ontwikkeling

26| <https://www.rijksoverheid.nl/documenten/rapporten/2019/09/17/referentieraming-ocw-2019>



Figuur 7 Schooladviezen Duo PO 2015

van leerlingen maar ook voor het opstellen van het schooladvies, wordt dit effect in de praktijk weer tegengegaan. Scholen worden nadrukkelijk gewezen op het belang hun schooladvies te onderbouwen op basis van de resultaten van de LVS-toetsen rekenen en begrijpend lezen groep 6, 7 en 8 die aanvankelijk vooral bedoeld waren voor het volgen van de ontwikkelingen van leerlingen, maar nu een belangrijke rol hebben gekregen in de onderbouwing van de schoolselectie. De Inspectie van het Onderwijs stuurt aan op schooladvisering ondersteund met gegevens van het LVS (zie ook hoofdstuk 4). Voor leerlingen die met achterstanden het basisonderwijs in zijn gekomen is de wetsaanpassing van 2011 dus geen selectie op een later moment geworden, maar heeft het door de toename van belang dat aan de uitkomsten van toetsen van het LVS wordt toegekend, juist geleid tot een feitelijke vervroeging van de selectie. Door de uitkomsten van het LVS mee te nemen in de onderbouwing van het schooladvies en daarbij de focus te leggen op lezen en rekenen/wiskunde ontstaat een zorgvuldig rankingproces van jaren waarbij ruimte voor groei en ontwikkeling steeds verder teruggedrongen wordt. De resultaten van groep 6 en 7 zijn niet langer vrijblijvend. Daarmee is de periode waarin relatieve rust en ruimte is voor het inhalen van achterstanden met meer dan een jaar ingekort. Omdat scholen minder ruimte hebben af te kijken van ofwel het schooladvies ofwel de gecombineerde uitkomsten van de LVS-toetsen voor groep zeven en acht, is het belang van de vaardigheden taal en rekenen, zoals gemeten met de eindtoets, toegenomen. En is er minder ruimte voor het meewegen van andere elementen. Daarmee is de balans verstoord die er aanvankelijk bij de selectie was door te werken met een combinatie van het oordeel van de school, die daarbij de ruimte had het kind en zijn context als geheel mee te wegen in het schooladvies. Er is nu een eenzijdige fixatie op scores gebaseerd op een opeenstapeling van vergelijkbare cognitieve, door objectief te scoren meting vastgelegde oordelen op basis van dezelfde smalle interpretatie van onderwijsdoelen, namelijk begrijpend lezen en rekenen. De meest begeerde plekken in het onderwijs zijn nog even schaars als in de jaren zestig. En de kans op het verwerven van die plek is het grootst voor kinderen van ouders die zelf hoger opgeleid zijn en vaker met voorsprong het onderwijs binnenkomen.

### 7.3.3 Eindtoets, LVS en effecten voor sociale groepen

De gelijkheid van kansen wordt in Nederland nog altijd beschouwd vanuit maatschappelijke groepen en hun onderwijsuitkomsten. Voor het vaststellen van hun succes in het basisonderwijs worden achtergrondvariabelen gecorreleerd met prestaties op de eindtoets en daarvan wordt onderzocht of bij gelijke prestaties in termen van toetsscores de doorstroomadviezen, gegeven door de scholen, wel leiden tot gelijke plaatsingsadviezen voor groepen geordend op basis van achtergrondvariabelen. De score op de toets is hierbij de norm geworden waartegen het plaatsingsadvies van de scholen wordt afgezet en waarover scholen zich moeten verantwoorden (zie ook hoofdstuk 3).

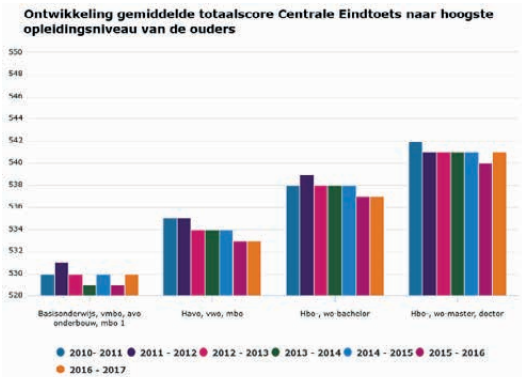
Veel scholen realiseren zich dat de periode in het basisonderwijs voor leerlingen met achterstanden te kort is om achterstanden in te kunnen lopen. Bij de advisering aan het eind van het basisonderwijs is in de jaren '80 sprake geweest van advisering waarbij het uitgangspunt voor kinderen uit achterstandsposities is dat zij het voordeel van de twijfel kunnen krijgen. Bijvoorbeeld op basis van de groei die ze hebben laten zien in de periode dat ze onderwijs volgen. Deze manier van denken en adviseren komt in de loop van de jaren tachtig onder druk te staan. Roeleveld et al. verwijzen naar een onderzoek van De Jong uit 1987 naar de schooladviezen in Rotterdam. Dit onderzoek laat zien dat allochtone leerlingen bij gelijke prestaties vaker een hoger advies krijgen dan hun autochtone jaargenoten (Roeleveld et al., 2011). Analyses van Driessen (1991), die gebruik maakt van landelijke gegevens, bevestigen dit beeld volgens Roeleveld (Roeleveld et al., 2011). Uit enkele studies uitgevoerd vanaf de tweede helft van de jaren '90, blijkt volgens Roeleveld dat 'de overadvisering van allochtone leerlingen substantieel is verminderd, dan wel geheel is verdwenen' (Roeleveld et al., 2011). Vanaf begin van deze eeuw is er eerder sprake van het tegendeel, namelijk 'onderadvisering'. Claassen en Mulder concluderen dat de in 1988 en 1992 nog geconstateerde overadvisering in 2000 is omgeslagen in onderadvisering (Claassen&Mulder, 2003). Driessen en Smeets (2007) hebben grootschalig onderzoek gedaan of er bij de overgang van het basis- naar het voortgezet onderwijs sprake is van een correcte advisering, en geen onder- of overadvisering, waarbij de capaciteiten vastgesteld met behulp van de uitkomsten van toetsen het uitgangspunt zijn. Voor de analyses is gebruik gemaakt van grootschalige, landelijke gegevensbestanden. De resultaten van die studie laten zien dat er geen sprake is van systematische onderadvisering van onderscheiden etnische groepen in verschillende gemeentetypen (Roeleveld et al., 2011).

Uit het onderzoek van Roeleveld et al. blijkt ook dat er over de periode 1988/89-2007/08 een positieve ontwikkeling te zien is wat betreft zowel de taal- als rekenprestaties van allochtone doelgroepleerlingen, hoewel die in groep 8 voor taal nog altijd de laagste scores behalen (ook lager dan de autochtone achterstandsléerlingen). Voor autochtone doelgroepleerlingen zijn er volgens het onderzoek in het basisonderwijs geen positieve ontwikkelingen in prestaties. Zij halen hun achterstand niet in, althans niet de groep die na de striktere definitie nog tot deze doelgroep behoort. De doorstroom naar hogere vormen van voortgezet onderwijs verbetert wel voor alle doelgroepleerlingen, zowel de autochtone als allochtone, maar hun achterstand op dit gebied blijft bestaan, omdat ook niet-doelgroepleerlingen vaker naar hogere vormen van voortgezet onderwijs gaan (Roeleveld et al., 2011: 31).

De discussie over en onderzoek naar het al dan niet voorkomen van over- of onderadvisering,

heeft het belang van de toets versterkt. De uitkomst van de toets, in termen van standaardcores, is het uitgangspunt waartegen oordelen van scholen worden gepositioneerd als over- of onderadvies, wat allebei leidt tot vermeende 'subjectiviteit' die voorkomen moet worden. De versterking van de rol van de eindtoets en het LVS komen voort uit de wens beïnvloeding zoveel mogelijk tegen te gaan. De keuze voor een eindtoets waarbij als 'proxy' voor het vaststellen van een schoolgeschiktheid voor de diverse vormen van beroeps- en algemeen vormend onderwijs gekozen is voor een bepaalde invulling van taal- en rekenvaardigheid, is echter eveneens een keuze die gevolgen heeft voor diverse groepen. Zeker omdat het 'talige' element ook in rekenopgaven aanwezig is, is de eindtoets met name te typeren als een selectie op talig, cognitief functioneren. Dit is in het voordeel van bepaalde groepen en in het nadeel van andere groepen. Het vermijden van 'subjectiviteit' door de inzet van een meerkeuze-instrument, is in het voordeel van de groepen die met een relatieve voorsprong het onderwijs in stromen en ouders hebben die hoger opgeleid zijn.

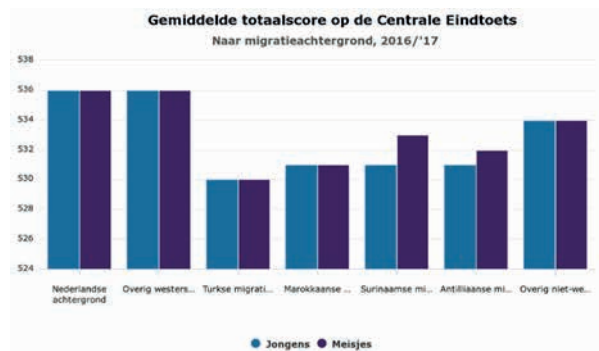
Op basis van de scores op de eindtoets wordt door Cito/CBS jaarlijks onderzocht hoe de scores op de eindtoets zich verhouden tot verschillende achtergrondvariabelen. Het opleidingsniveau van ouders hangt samen met de gemiddelde totaalscore die leerlingen behalen op de Centrale Eindtoets. Hoe hoger het opleidingsniveau van de ouders, hoe groter de kans voor een kind op een hoge score op de eindtoets en daarmee op een hoger schooladvies.



Figuur 8 Ontwikkeling gemiddelde totaalscore Centrale Eindtoets naar hoogste opleidingsniveau van de ouders

Dit geldt ook voor de migratieachtergrond van de leerlingen. De gemiddelde scores van leerlingen met een Nederlandse of westerse achtergrond bevinden zich op een hoger niveau dan die van diverse groepen leerlingen met een migratieachtergrond. Hun standaardcores liggen gemiddeld genomen beduidend lager dan die van hun Nederlandse leeftijdsgenoten.

De keuze voor het testen van taal- en rekenvaardigheid kan voor een deel de verschillen die zich voordoen op de standaardcores op de eindtoetsen verklaren omdat taal



Figuur 9 Gemiddelde totaalscore op de Centrale Eindtoets naar migratieachtergrond

zowel een te meten construct is als het middel waarmee de inhoud wordt vormgegeven. Het feit dat bij het meten van rekenvaardigheid een substantieel deel van de opgaven jaarlijks bestaat uit sommen die zijn ingebed in een context (de zogenaamde 'verhaaltjessommen') maakt het relatieve gewicht van 'taalvaardigheid' groter.

### 7.3.4 Kansengelijkheid en toetsing

Gelijke kansen in het onderwijs is in Nederland een belangrijk streven in het onderwijsbeleid dat onder kabinetten met uiteenlopende signatuur al decennia lang terugkeert op de politieke agenda. Als de Inspectie van het Onderwijs zich zorgelijk uitlaat over de kansengelijkheid dan leidt dat tot politieke commotie. In de afgelopen jaren spreekt de jaarlijkse rapportage over de staat van het onderwijs over het onder druk staan van kansengelijkheid. Recent, in maart 2019, schrijven de beide ministers van onderwijs in een brief aan de kamer dat dit hun zorgen baart en zetten ze uiteen hoe ze de bevordering van kansengelijkheid in het onderwijs de komende jaren voor zich zien: 'Bij de bevordering van kansengelijkheid gaat het erom dat jongeren in staat worden gesteld – en worden gemotiveerd en waar nodig ondersteund – om hun talenten te ontwikkelen. Het schoolsucces van kinderen zou enkel afhankelijk moeten zijn van hun capaciteiten en inzet en niet van het inkomen of het opleidingsniveau van hun ouders.' (OCW, 19 maart 2019 brief aan TK). Ook uit deze brief, in lijn met voorgaand beleid, blijkt dat de overheid vooral een rol ziet voor zichzelf om te zorgen dat de competitie, die inherent verbonden blijft aan het normatieve ideaal van kansengelijkheid, eerlijk verloopt. Kansengelijkheid in Nederland is fundamenteel verbonden geraakt met het reguleren van het competitieve proces dat inherent is aan onderwijs zodat er sprake kan zijn van een 'faire' selectie. De Canadese politicoloog Jacobs stelt dat het waarborgen van gelijke kansen gaat over de normatieve regulatie van competities over waardevolle kansen in de samenleving. Voor een normatieve regulatie van kansen in de samenleving onderscheidt Jacobs naast procedurele fairheid en achtergrond fairheid een derde vorm, die van 'stakes' fairheid (Jacobs, 2016: 325). De eerste twee aspecten kunnen in het onderwijsbeleid rekenen op aandacht. Dat geldt niet voor het derde aspect, de 'stakes' fairheid. Daarmee bedoelt Jacobs dat het van belang is te kijken naar wat er op het spel staat, naar de risico's die met de competitie zijn verbonden voor alle deelnemers; winnaars en verliezers. De traditionele blik van kansengelijkheid is volgens Jacobs vooral ééndimensionaal en gericht op de fairheid van procedures. Soms is sprake van een tweedimensionale manier van kijken bij verdelingen en wordt rekening gehouden met de achtergrond van deelnemers en de verdeling van kansen over verschillende groepen. Het toevoegen van een derde dimensie, de 'stakes fairness', weerspiegelt de zorg die met verdelvingsvraagstukken samenhangt, namelijk de verdeling van voordelen (benefits) en de lasten (burdens) die aan winnen of verliezen verbonden zijn. Jacobs onderscheidt binnen stakes fairness drie aspecten waar rekening mee moet worden gehouden: 1) zorg voor het beperken van risico's waaraan deelnemers in de competitie worden bloot gesteld, 2) zorg over hoeveel er feitelijk op het spel staat, en 3) de zorg voor het limiteren van de impact die de ene competitie heeft op een volgende. Een formule als 'the winner takes all' wordt zelden als fair ervaren. Jacobs stelt dat het vraagstuk hoe om te gaan met winnaars en verliezers in onderwijsbeleid niet heel gebruikelijk is en spreekt in dit verband van een 'normative blind spot' (Jacobs, 2016: 328). Ook voor het Nederlandse debat over kansengelijkheid geldt dat dit debat vooral gaat over de procedurele kant van het selectievraagstuk. Deze focus heeft de aandacht afgeleid van de steeds sterker geworden gevolgen voor groepen met een bepaalde achtergrond en voor de gevolgen die met het winnen of verliezen van de Nederlandse onderwijscompetitie samenhangen. Het beperken van de beantwoording van de validiteitsvraag tot vraagstukken die samenhangen met het instrument op zich en niet met de effecten van het instrument in onderwijs en samenleving, hangt hier mee samen (zie ook hoofdstuk 3). De aandacht voor validiteit beperkt zich tot aspecten van een toetsinstrument zelf. Maar juist de test in relatie tot de toepassing in de praktijk van het onderwijs is relevant voor beleid, zeker voor beleid dat zegt waarde toe te kennen aan

onderwijs als plaats voor het bieden van gelijke kansen. Normatieve vragen over de fairheid van verdelingen als onderdeel van het vraagstuk van validiteit zijn daarbij wel degelijk van belang; kan op basis van het gebruikte instrument de competitie met een dergelijk gewicht voor winnaars en verliezers als fair en daarmee als rechtvaardig worden beschouwd? Dit als vervolg op meer algemene vragen naar de fairheid van verdelingen in de context van onderwijs zoals: is schaarste fair? Is de competitie voor het winnen van de meest begeerde plekken fair? Is het instrument dat daarbij ingezet wordt fair? Voor het stellen en zeker ook voor de beantwoording van die vragen is het relevant om te onderzoeken of er in het Nederlandse onderwijs nu sprake is van selectie, toelating of plaatsing.

#### **7.4 Selectie, toelating of plaatsing?**

Nederland heeft een leerplicht tot de leeftijd van 16 jaar. Daarmee wordt er voor iedereen een plek in het Nederlandse onderwijs gerealiseerd, wat zou pleiten voor het spreken over een proces van allocatie en daarmee van een plaatsingsvraagstuk: er is voor iedereen immers een deel van het te verdelen goed onderwijs beschikbaar. Het feit dat leerlingen afhankelijk van hun resultaten op de eindtoets verschillende programma's volgen die tot verschillende diploma's of certificaten leiden, maakt dat er volgens Cito sprake is van een 'toets met een classificerend karakter' (Sanders, 2017b: 12). In Nederland bestaat het beeld dat de toewijzing van plekken samenhangt met de capaciteiten van leerlingen en dat er sprake is van een 'geschiktheidsbepaling' met behulp van vorm van 'ability-testing'. In het maatschappelijk discours wordt gesteld dat de verdeling gebaseerd is op 'behoefte' van de individuele leerling ('need') en wordt gesteld dat iedereen dat krijgt wat hij nodig heeft of wat bij hem past. De behoeften worden vastgesteld door een extern gevalideerd instrument waarvan de suggestie uitgaat dat de behoeften-inventarisatie objectief verloopt. De vraag is of de manier waarop nu behoeften worden geïnventariseerd de best passende vorm is om behoeften in kaart te brengen en om welke behoeften-inventarisatie het hierbij gaat. Staat hier de behoefte van een individu aan passend onderwijs centraal of eerder de behoeften van een arbeidsmarkt aan voldoende, passend personeel? Vastgestelde prestaties gebaseerd op cognitieve onderwijshoudens, samengeperst in één maat, komen voort uit en wortelen nog altijd in een traditie van 'ability-testing' bedoeld voor selectie, namelijk de selectie van de besten die als beloning de langste onderwijstrajecten aangeboden krijgen met de beste kansen op vervolgonderwijs en de daarmee samenhangende maatschappelijke posities. Het fenomeen schaarste is inherent aan selectie en de vraag is relevant hoe passend het fenomeen schaarste is in relatie tot onderwijs. In termen van Elster gaat het hier niet alleen om een schaars goed dat verdeeld kan worden, maar ook nog eens om een goed dat in principe zo kan worden verdeeld dat er van alles genoeg is voor iedereen. In de perceptie van ouders en leerlingen zijn de plekken die in het onderwijs nu beschikbaar zijn kwalitatief niet gelijkwaardig. Het aantal van de meest begeerde plekken is beperkt, de plekken in het vwo zijn toegankelijk voor niet meer dan een vijfde deel van de populatie. Daarmee is er sprake van een gecreëerde schaarste van een goed dat bewust beperkt wordt in hoeveelheid. In een dergelijk geval is volgens Elster een procedure van belang waarbij zowel het goed als de afnemer in ogenschouw worden genomen. In een dergelijke situatie zijn er volgens Elster een drietal opties. De eerste betreft een loting waarbij iedereen evenveel kans heeft op een toelating tot het schaarse goed. De tweede is een toelating waarbij individuen vergeleken worden met een norm en alleen zij die aan de norm voldoen, worden toegelaten. De derde optie is dat er sprake is van een selectieprocedure waarbij individuen worden vergeleken met elkaar, meestal door een rankinglist te produceren waarbij

de acceptatie top down gaat net zolang totdat het goed op is. De eindtoets in combinatie met de inzet ervan voor de plaatsing van leerlingen in het voortgezet onderwijs laat zien dat er geen sprake is van loting noch van criteriumgerelateerd toelaten: inhoudelijke criteria gekoppeld aan de verschillende onderwijstypen ontbreken in Nederland. Er is, in termen van Elster (en Hofstee), sprake van een selectieprocedure, uitgaande van een gecreëerde schaarste aan plekken in het vwo die top down wordt uitgevoerd. De toelating tot vwo op basis van prestaties op een en dezelfde schaal van jaarlijks stabiele standaardscores heeft niet alleen een vergelijking mogelijk gemaakt van leerlingen met elkaar maar ook een rangordering waarmee toegang tot de meest begeerde onderwijsplekken snel en efficiënt kan worden geregeld. Er is gekozen voor het ordeningsprincipe van de tot één standaardscore gecombineerde prestatiescores op taal- en rekentoetsen. Middels langere en duurdere onderwijstrajecten die toegang geven tot de hoogste vormen van vervolgonderwijs investeert de overheid in degenen die de beste kansen hebben die investering waar(d) te maken en daar ook het meeste recht op hebben gegeven hun geleverde prestaties in het verleden.

Het competitie­model waarvan altijd al sprake was in het voortgezet onderwijs, heeft zich uitgebreid naar het basisonderwijs. Het toegenomen belang van prestaties op de toetsen van het LVS in combinatie met de toegenomen waarde van de score op de eindtoets heeft ervoor gezorgd dat de rust en de ruimte die er aanvankelijk was voor de ontwikkeling in het primair onderwijs plaats maakt voor een steeds belangrijkere en vroegere competitie. Want het gevecht om de schaarse plekken in het voortgezet onderwijs vindt steeds vroeger plaats nu de scores op de LVS-toetsen van groep 6,7 en 8 meegenomen moeten worden in het schooladvies. Hoe jonger de selectie plaatsvindt, hoe sterker de milieu-­invloed is. En hoe minder meritocratisch de selectie zal kunnen zijn, want hoe jonger de prestaties worden gemeten hoe meer ze zijn toe te schrijven aan aanleg, opvoeding en achtergrond en met verdiensten in termen van schoolse prestaties nog niet veel te maken kunnen hebben.

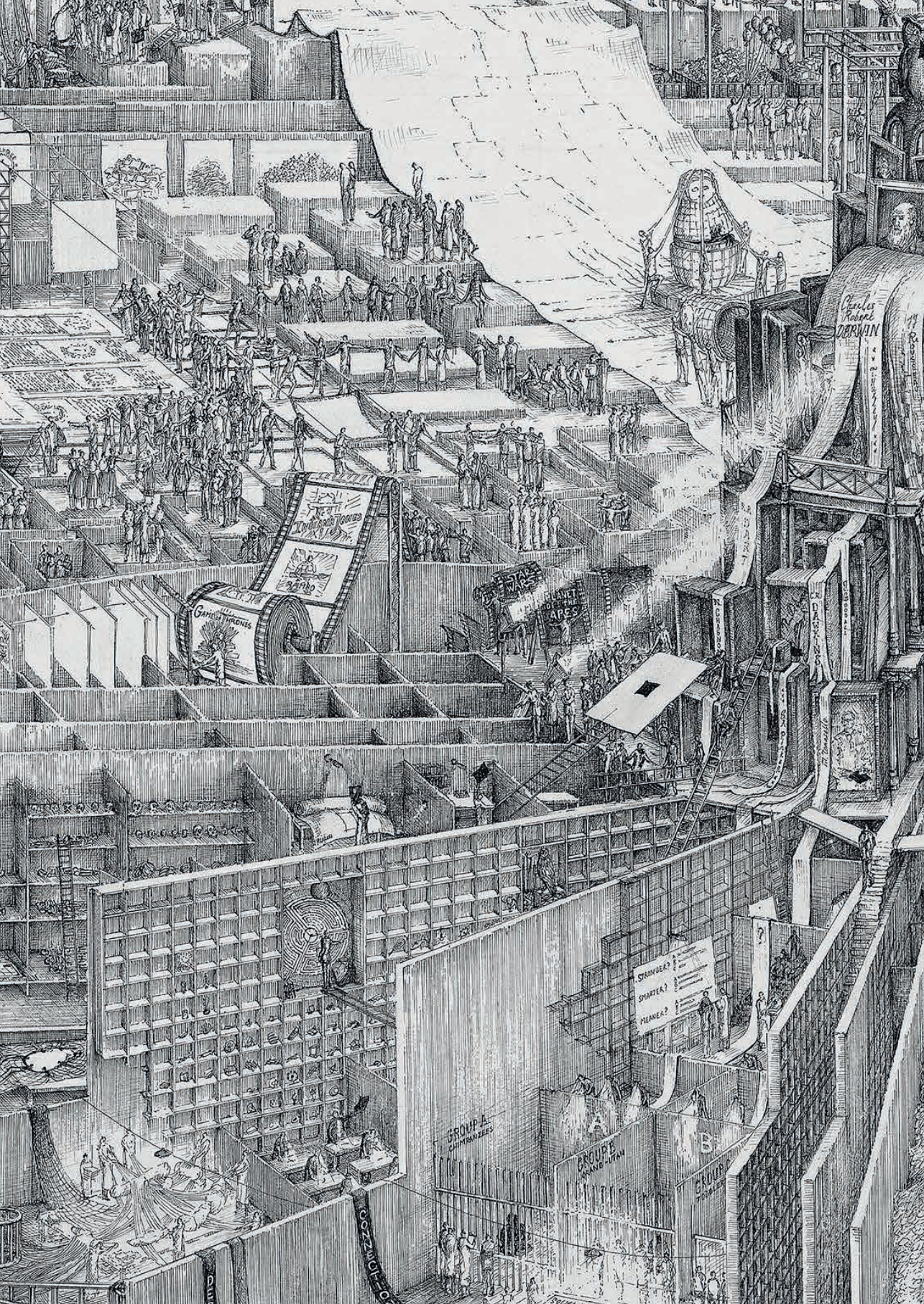
## 7.5 Samengevat

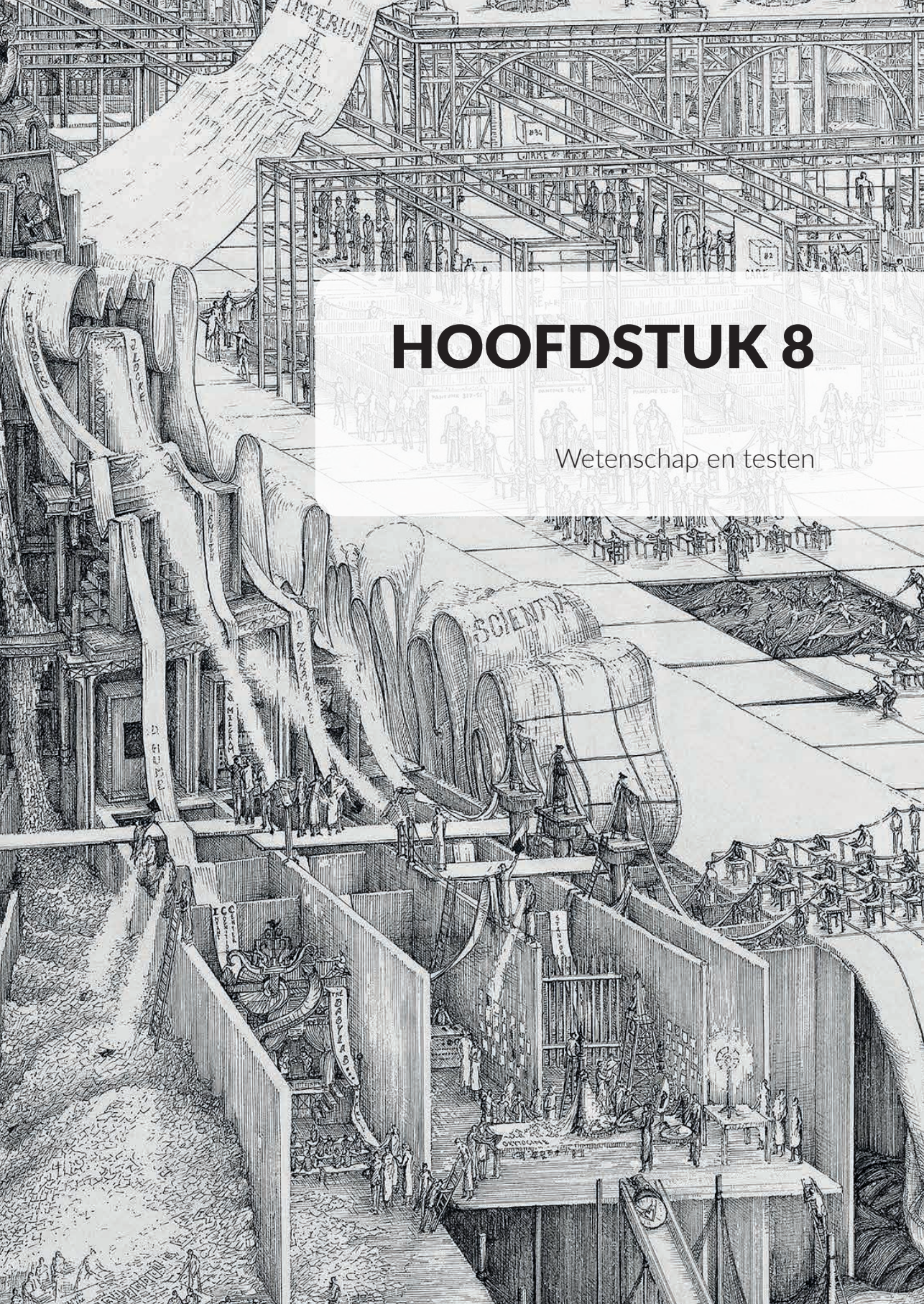
Onderwijs­onderzoek laat zien dat het algemene opleidingsniveau in Nederland de afgelopen eeuw weliswaar is gestegen, maar de verdeling over de verschillende niveaus van onderwijs laat grote stabiliteit zien. Het volgen van onderwijs is enerzijds een verplichting waarbij de overheid zorgt dat er voor iedereen voldoende van het goed onderwijs beschikbaar is. Tegelijkertijd is er van het meest begeerde stukje van het voortgezet onderwijs maar een beperkt stukje van de taart beschikbaar, ongeveer twintig procent van de leerlingen kan een plek krijgen in het vwo. Daarmee is onderwijs, en de verdeling van plekken, wel degelijk te zien als een schaars goed met een daarmee samenhangend verdelings­vraagstuk dat in termen van rechtvaardigheid beschouwd kan worden. Het Nederlandse onderwijs is te zien als een model met een combinatie van een systeem van vroege selectie (expeditie­model) en een competitie (veldloop- of hordenloop­model). Al vroeg moeten de leerlingen met elkaar de competitie aan in de strijd om de beste plekken. De selectie vindt formeel plaats op 12-jarige leeftijd maar de verplichting van het gebruik van een LVS heeft een feitelijke vervroeging van de selectie gestimuleerd. De criteria die bij de verdeling worden gehanteerd laten twee gezichten zien. Enerzijds wordt gezegd dat meritocratische uitgangspunten de basis van de selectie vormen zodat de besten kunnen winnen. Tegelijkertijd is de huidige selectie gericht op het bepalen van geschiktheid. Door de verdeling te baseren op steeds vroeger verzamelde onderwijs­resultaten is de beste vooral degene die met

de beste papieren aan de start van de wedstrijd verschijnt. De vraag naar de achtergrond van de gecreëerde schaarste en in combinatie daarmee de fairheid, de rechtvaardigheid, van de huidige manier van verdelen en de gevolgen ervan voor winnaars en verliezers, dringt zich op. De eindtoets in combinatie met het LVS fungeert als een steeds harder wordend instrument voor de verdeling van onderwijsposities in termen van kansen voor de winnaars en van de gevolgen voor verliezers. Het roept de vraag op of het huidige onderwijs, in combinatie met de gehanteerde selectie, te typeren is als een model dat bijdraagt aan de ontwikkeling van capaciteiten van alle leerlingen en daarmee dienstig is aan het bieden van kansen aan allen. Het selectieve onderwijs, in combinatie met harder wordende selectie, is niet gericht op inclusie, maar werkt als een systeem om successievelijk leerlingen uit te sluiten van kansen om hun mogelijkheden te verkennen en te ontwikkelen. In het licht van de mate van rechtvaardigheid van deze verdelingen en de gevolgen ervan is het van belang nader te onderzoeken of de test als instrument voldoet aan de kwaliteitseisen zoals die zijn vastgelegd in (internationale) richtlijnen en standaarden voor ontwikkeling en gebruik van psychologische tests en studietoetsen. Daarover gaat het volgende hoofdstuk.



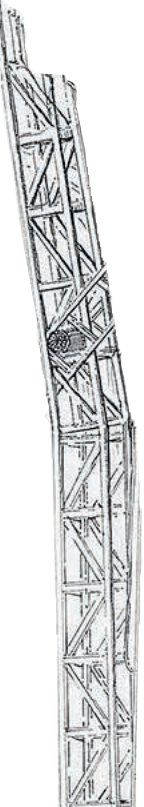






# HOOFDSTUK 8

Wetenschap en testen



## Hoofdstuk 8 Wetenschap en testen

*'De ene slaagt beter dan de andere, dat is een feit. Als hij beter slaagt, zegt u, dan is dat omdat hij intelligenter is. Hier begint de uitleg duister te worden. Heeft u een ander feit aangewezen dat de oorzaak zou zijn van het eerste? [...] Wanneer u zegt 'hij is intelligenter' doet u niets anders dan de ideeën samenvatten die van het feit verslag doen. U heeft het feit een naam gegeven. Maar de naam van een feit is niet de oorzaak van dat feit, ten hoogste is het er de metafoor van.'*

*Uit: Jacques Rancière, 'De onwetende meester' (2007)*

In het in 1969 verschenen werk van Gunnar Myrdal met de naam 'Objectivity in Social Research' wijst Myrdal op het onvermijdelijke 'a priori element' waar al het wetenschappelijk onderzoek mee te kampen heeft: vragen worden geformuleerd voordat antwoorden kunnen worden gegeven. En de vragen die wetenschappers formuleren weerspiegelen hun eigen, al dan niet verborgen, belangstelling voor en geloof in de wereld, en zijn dus te zien als waardebevestigingen. Daarmee ligt, volgens Myrdal, waardebevestiging aan de basis van elk wetenschappelijk onderzoek (Myrdal, 1969: 9). In zijn ogen kun je niet eenvoudigweg 'objectief' zijn, maar je kunt wel een poging wagen te 'objectiveren' waar je mee bezig bent. We kunnen volgens Myrdal niet veel meer doen dan daar open over zijn: de onderliggende waarderingen naar de oppervlakte brengen, daar bewustheid over creëren, specifiek en expliciet, zodat duidelijk is wat de theoretische uitgangspunten van onderzoek zijn alvorens praktisch aan de slag te gaan (Myrdal, 1969: 56). Objectiviteit is meer een streven dan een te realiseren ambitie. Porter definieert 'objectiviteit' als 'strijd tegen subjectiviteit'. Hij zegt daarvan zelf dat het een zwakke definitie is, maar hecht vooral waarde aan strategieën om die strijd te winnen. Het gaat volgens hem om een 'set of strategies for dealing with distance and distrust' (Porter, 1995: IX). Een van die strategieën is het volgen van regels om subjectiviteit tegen te gaan: regels zouden persoonlijke bias of voorkeuren moeten voorkomen, zodat uitkomsten van onderzoek daardoor niet kunnen worden beïnvloed. Maar Porter wijst erop dat het disciplineren door regels te volgen ook waarheidsvinding in de weg kan staan. Mechanische objectiviteit, kennis volledig gebaseerd op expliciete regels, kan niet worden bereikt. Daarom acht hij 'informed intuition' minstens zo belangrijk (Porter, 1995: 7). Huisman plaatst conceptuele meerduidigheid van wat verstaan moet worden onder 'objectiviteit' in de context van bias. Om bias in de wetenschappelijke wereld te kunnen benoemen alvorens deze aan te kunnen pakken, acht zij daarom allereerst een definitie van objectiviteit van belang. Het begrip 'objectieve kennis' is meerduidig: gaat het bij 'objectieve kennis' om kennis die door een subject op objectieve wijze is verkregen of wordt daarmee kennis bedoeld die overeenkomt met de werkelijke kenmerken van het object? (Huisman, 1997: 33). Toetsing wordt in de dagelijkse praktijk snel verbonden met de wetenschap waarmee objecten op een objectieve manier gekend kunnen worden. Of en in hoeverre dit kenproces als geslaagd kan worden beschouwd heeft te maken met het vraagstuk van validiteit en bias. Dit geldt voor onderzoek naar het vaststellen van objectieve waarde en waardering binnen de toetswetenschap zelf, maar is ook relevant daar waar tests een plek krijgen in andere vormen van wetenschappelijk onderzoek. Want toetsing heeft zich enerzijds ontwikkeld tot een zelfstandige wetenschap én is tegelijkertijd verweven mét wetenschap, in die zin dat andere wetenschappen tests, of de resultaten daarvan, zijn gaan inbedden in eigen onderzoek. Dit is in eerste instantie gebeurd in de psychologie, later ook in de (onderwijs)sociologie en is nu ook relevant voor tal van onderwijskundig onderzoek. Hoe toet-

sing en onderzoek zich los van elkaar en gezamenlijk hebben ontwikkeld staat centraal in dit hoofdstuk. Het hoofdstuk start met een verkenning van het begrip 'objectiviteit' want dit begrip verbindt toetsing aan veel wetenschappelijk onderzoek (8.1). Vervolgens worden de lijnen verkend waarlangs onderzoek in de Verenigde Staten zich heeft ontwikkeld (8.2) omdat met name deze manier van denken over testen en onderzoek de positie van tests in onderzoek in Nederland heeft beïnvloed, zoals blijkt uit de paragraaf waarin de onderwijswetenschapsontwikkeling in Nederland centraal staat (8.3). Vervolgens worden kernbegrippen uit de toetswetenschappen zelf nader onderzocht en op waarde geschat (8.4).

### 8.1 Objectiviteit in onderzoek en tests

Het begrip 'objectiviteit' heeft een positieve connotatie in het dagelijks taalgebruik, net als dat de vermeende keerzijde ervan, 'subjectiviteit', een negatieve connotatie heeft. Van Dale maakt bij het geven van een omschrijving van het woord 'objectief' een onderscheid tussen de betekenis van het bijvoeglijk naamwoord 'objectief' en het zelfstandig naamwoord 'objectief'. Als bijvoeglijk naamwoord stelt Van Dale dat 'objectief' betekent: 'zonder zich door eigen voorkeur te laten beïnvloeden; onbevooroordeeld, onbevangen' en positioneert 'objectief' hierbij als de tegenstelling van 'subjectief'. Als zelfstandig naamwoord betekent 'objectief' enerzijds '(in kijkers, camera's) de naar het object gerichte lens of het samenstel van lenzen' en anderzijds 'het doelwit, oogmerk'. Daarmee is objectiviteit verbonden met zowel een letterlijke manier van waarnemen, de lens waarmee naar een probleem gekeken wordt, als met een aan te nemen houding die hoort bij waarnemen, namelijk die van een onafhankelijke waarnemer. Beide aspecten zijn verbonden met objectiviteit in onderzoek. Met het boek 'Wetenschap in Actie' heeft de Franse wetenschapsfilosoof Latour wetenschap en de vermeende objectiviteit ervan in een ander daglicht gesteld: feiten bestaan niet maar worden gecreëerd door de rol en positie die ze verwerven in netwerken. De creatie van feiten is volgens Latour een sociaal proces en moet als zodanig worden onderzocht. Een belangrijk punt van Latour is dat een wetenschappelijk feit pas 'waar' wordt wanneer het niet alleen in een goed onderbouwd wetenschappelijk artikel is gepubliceerd, maar ook door collega-onderzoekers in hun artikelen wordt aangehaald en dus erkend (Latour, 1988). Zoals Latour het stelt: 'Science is not politics. It is politics by other means.' (Latour, 1993: 229). Pearl en Mackenzie roepen wetenschappers op het 'eeuwenoude dogma – objectiviteit omdat het nu eenmaal objectiviteit is – achter zich te laten', want, zo stellen zij: 'Een snuffe wijze subjectiviteit vertelt ons meer over de echte wereld dan welke hoeveelheid objectiviteit dan ook.' (Pearl & Mackenzie, 2018: 116). Bewijs gaat volgens hen altijd gepaard met een mate van onzekerheid (Pearl & Mackenzie, 2018: 137) en zij wijzen op het belang van het niet zomaar aannemen van een uitslag, maar juist steeds de vraag te stellen: hoe groot is de kans dat het waar is wat er uit een test komt? Of: hoe groot is de kans dat het niet waar is? En daaraan gekoppeld vragen die betrekking hebben op de situatie: wat als er niet gemeten was, wat dan?

Onderzoek in de sociale wetenschappen in Nederland is sterk beïnvloed door de man die gezichtsbepalend is geworden voor toetsing in Nederland, A. D. de Groot. Zijn boek 'Methodologie' uit 1961 geldt in de jaren zestig en zeventig als hét standaardwerk voor eenieder die zich met empirisch onderzoek binnen de sociale wetenschappen gaat bezighouden (Busato et al., 2014: 261). De kern van het werk is de cyclus van het empirisch-wetenschappelijk onderzoeken: observatie, inductie, deductie, toetsing en evaluatie. De mate van objectiviteit die de wetenschapsbeoefenaar aan de dag kan leggen verschilt per fase. Voor de eerste fase, die in zijn ogen meer cre-

atief is, acht hij subjectiviteit onvermijdelijk en zelfs noodzakelijk (De Groot, 1961: 177). En dat geldt volgens hem ook voor de fase van de evaluatie. Het streven naar objectiviteit ziet hij vooral als relevant voor de opzet van het daadwerkelijke onderzoek en de uitvoering ervan (De Groot, 1961: 178). Zijn beroemd geworden adagium is: 'Als ik iets weet, kan ik iets voorspellen; kan ik niets voorspellen, dan weet ik niets.' (De Groot, 1961: 20). Alleen objectief getoetste kennis is volgens hem van waarde. De waarheidswaarde van de afspiegeling is daaraan ondergeschikt.

Bij legitimering van het gebruik van tests wordt met regelmaat gewezen op het belang van objectiviteit van het oordeel waarbij objectiviteit van een test geplaatst wordt tegenover de subjectiviteit van het oordeel van een professional, in het geval van onderwijs: de leerkracht. Want van mensen is bekend dat zij, bewust of onbewust, zich laten sturen door andere aspecten dan waar de beoordeling op is gericht zodat zij in die zin niet kunnen komen tot een 'objectief' oordeel. Chapman verklaart de aantrekkelijkheid en groei van intelligentietests vooral vanuit dat perspectief. Selectie op basis van persoonlijke oordelen is gevoelig voor subjectiviteit in de vorm van bias voor leeftijd, sekse en ras. Rist heeft met zijn onderzoek uit 1970, 'The selffulfilling prophecy in Ghetto Education', laten zien welke uitkomsten classificatie heeft als dat gebeurt op basis van subjectieve oordelen. Zijn conclusie is dat school een samenleving niet alleen weerspiegelt maar deze ook zelf construeert (Rist, 2000: 171). Onafhankelijke, objectieve metingen kunnen aan mogelijke schadelijke effecten van subjectieve oordelen een tegenwicht bieden, maar brengen hun eigen vormen van subjectiviteit in stelling. Chapman laat zien dat de opkomst van intelligentietests samengaat met systemen van 'tracking'. Het 'objectieve' imago dat intelligentietests hebben in de zin van het kunnen beoordelen van de uitkomsten, onafhankelijk van de mogelijke subjectiviteit van een beoordelaar, heeft groepsgewijze afname van intelligentietests wind in de zeilen gegeven. Maar juist praktijken van classificeren op basis van intelligentietests, zoals bij tracking, hebben volgens Chapman vaak belemmerend gewerkt voor de kansen op het bereiken van goede posities op de arbeidsmarkt van de VS van vooral zwarten en immigranten uit Zuid- en Oost-Europa. Chapman spreekt in dit verband over tests als 'road blocks' (Chapman, 1988: 175).

Tests zijn niet zomaar te beschouwen als objectieve, wetenschappelijke instrumenten. In lijn met de tweede methodologische regel die Latour onderscheidt, is het van belang instrumenten te onderzoeken in samenhang met alle veranderingen die ze later in handen van anderen ondergaan (Latour, 1988: appendix 1).

## 8.2 Internationale ontwikkelingen en testen

Met de zin 'Thorndike 'won' and Dewey 'lost' vat de Amerikaanse onderwijskundige Condliffe Lagemann de strijd samen die aan het begin van de 20ste eeuw het onderwijsonderzoek in de Verenigde Staten beheerst (Condliffe Lagemann, 2000: XI). Thorndike is bepalend geweest voor de ontwikkeling van onderwijspsychologie. Thorndike heeft onderwijspsychologisch onderzoek ontwikkeld, gebaseerd op empirische data, verzameld onder gecontroleerde condities bij voorkeur in een laboratoriumsetting, daarmee aansluitend op de traditie zoals gegroeid in die tijd van natuurwetenschappelijk onderzoek. Voor Thorndike zijn directe experimenten interessant waarbij een enkele variabele wordt onderzocht: 'He was far more interested in tests of simple variables than he was in observation within naturalistic settings.' (Condliffe Lagemann, 2000: 59-62). Belangrijk is zijn wetmatigheid: 'The Law of Effect'. Hiermee heeft hij vastgesteld dat gedrag waarop een positieve respons volgt zich in de toekomst vaker voor zal doen dan gedrag

dat negatief wordt beloond (Mayer, 2003: 131). Leren is in de optiek van Thorndike een zaak van het versterken of verzwakken van de associaties tussen stimulus (S) en respons (R) (Mayer, 2003: 132). Thorndike heeft met zijn onderzoek niet alleen een belangrijke bijdrage geleverd aan theorievorming over leren maar ook over testen. In tegenstelling tot de gangbare gedachte van eerst Galton en later Spearman, dat er zoiets zou bestaan als een algemene intelligentie factor, 'factor G', stelt Thorndike dat individuele verschillen in scores op intelligentietests vooral terug te voeren zijn op kennis en meer precies op kennis die volgens Thorndike wordt opgebouwd door veel ervaringen met S en R (Mayer, 2003: 136). Hij verschuift het denken van 'intellectual ability' naar 'ability to learn'. Nog altijd is 'ability' het basisuitgangspunt onder de moderne intelligentie-theorieën (Mayer, 2003: 136). Thorndike wil met tests de superieure studenten kunnen identificeren omdat niet iedereen volgens hem even veel baat heeft bij bepaalde vormen van onderwijs. Het doen van experimenteel onderwijspsychologisch onderzoek, zoals Thorndike dat doet, is gericht op het kunnen voorspellen van uitkomsten van onderwijs. Hoe exacter de onderwijspsychologie zou worden, hoe beter die psychologie onderwijsuitkomsten zou kunnen voorspellen, en hoe minder belangrijk de rol van de onderwijsgever kon worden (Condliffe Lagemann, 2000: 60). Het gedachtegoed van Thorndike staat daarmee haaks op dat van Dewey, die als onderwijspsycholoog veel meer een holistische manier van kijken voorstaat. Het is volgens Dewey aan de docent om als professional uitkomsten van (laboratorium)onderzoek te onderzoeken op hun waarde in de dagelijkse praktijk en ze afhankelijk van de mogelijkheden wel of niet in te passen in de onderwijspraktijk (Dewey, 1929: 76). Onderwijs is in zijn ogen complex en er zijn veel factoren die tegelijkertijd op elkaar inspelen en die nooit allemaal gelijktijdig middels experimenteel onderzoek onder controle gehouden kunnen worden (Dewey, 1929: 19). Het doel van onderwijs is volgens Dewey het realiseren van 'groei' van een kind. Vaste, uitgewerkte doelstellingen voor onderwijs zijn niet nuttig omdat omstandigheden snel kunnen veranderen. Het is vooral van belang te zorgen dat individuen zelf leren nadenken en zich aan die veranderingen kunnen aanpassen (Bredo; 2003: 103). Het denken van Dewey krijgt in de Verenigde Staten pas vanaf de jaren vijftig echt belangstelling. Voor veel hedendaagse pedagogen is Dewey nog altijd belangrijk voor hun denken over onderwijs (zie Biesta, 2012; Meirieu, 2016; Pols, 2016; Berding, 2016). De maatschappelijke en wetenschappelijke aantrekkingskracht van de belofte die schuilt in de manier van kijken en ordenen zoals Thorndike die voorstaat, is zeker aanvankelijk groter en Thorndike is bepalend geweest voor hoe onderwijspsychologie en in lijn daarmee het onderwijs in de Verenigde Staten zich heeft ontwikkeld.

### 8.2.1 Intelligentie en IQ

In zijn controversiële boek (1996) schetst Stephen Jay Gould de historie van intelligentietests. Het testen van intelligentie op grote schaal heeft vooral in de loop van de vorige eeuw plaatsgevonden in de Verenigde Staten. De basis voor het denken komt echter uit Europa. De Fransman Alfred Binet, destijds directeur van een psychologisch laboratorium in Parijs, komt begin 20e eeuw tot de slotsom dat schedelmetingen niet langer een goede indicator vormen voor het vaststellen van intelligentie (Jarvin & Sternberg, 2003: 65). Binet verdiept zich in het fenomeen van abnormaliteit bij kinderen. In 1904 krijgt hij de opdracht van de Franse minister van onderwijs om een manier te ontwikkelen om bij kinderen een mentale achterstand te kunnen vaststellen (Jarvin & Sternberg, 2003: 70). Binet ontwikkelt een serie vragen en proefjes die hij een intelligentietest noemt. Hiermee wil Binet van kinderen met leerproblemen vaststellen wie alsnog zou kunnen profiteren van onderwijs, ondergebracht in speciale klassen (Jarvin & Sternberg, 2003: 71). Bi-



net ontwikkelt 30 kleine testjes van wisselende moeilijkheidsgraad en hoopt dat het geheel van testjes een indicatie oplevert van het algemene vermogen van een kind. Binet stelt daarbij echter expliciet dat een op deze manier verkregen score 'niets zegt over de mate van intelligentie omdat intellectuele kwaliteiten niet optelbaar zijn en daarom niet gemeten kunnen worden zoals lineaire oppervlakken' (Gould, 1996: 200). Voor Binet is intelligentie niet op te vatten als een aangeboren eigenschap en een testscore is voor hem nadrukkelijk geen maat voor gefixeerde intelligentie (Jarvin & Sternberg, 2003: 73). Binet en Simon formuleren het zelf als volgt: 'It is understood that these diagnoses apply only to the present moment. One who is imbecile today, may by the progress of age become a moron, or on the contrary remain an imbecile all his life. One knows nothing of that; the prognosis is reserved.' (Binet&Simon, 1916: 270). De schaal die Binet ontwikkelt, is nadrukkelijk slechts een instrument en moet aangevuld worden met andere bronnen van informatie die meer kwalitatief van aard zijn om een beeld te kunnen krijgen van de globale persoon (Jarvin & Sternberg, 2003: 73). Binet en Simon wijzen er ook nadrukkelijk op dat de schaal niet bedoeld is voor 'normale' leerlingen maar juist voor de leerlingen die achterblijven; 'those of inferior grades of intelligence' (Binet&Simon, 1916: 263).

De bedoelingen van Binet gaan verloren bij de overtocht van zijn schaal over de oceaan naar de Verenigde Staten waar de schaal en de testjes in handen komen van de Amerikaan Lewis M. Terman die in lijn met Galton intelligentie vooral ziet als een aangeboren eigenschap (Jarvin & Sternberg, 2003: 74). Terman gebruikt de schaal om kinderen met een lage intelligentie te scheiden van de anderen, niet met de intentie om ze extra onderwijs te geven maar om onderwijs voor hen als niet zinvol te verklaren en aan hen juist geen extra aandacht te hoeven geven (Jarvin & Sternberg, 2003: 74). Terman en zijn opvolgers gaan intelligentietests gebruiken om subjecten te ordenen ten opzichte van elkaar waarbij sommigen 'voor' lopen en anderen 'achter', waarvan sommigen nog bij te spijkeren zijn en anderen niet. De normaalverdeling wordt de basis voor een verdeling van normaliteit.

### **Intelligentietests**

Het is Terman die in 1916 de intelligentietest, zoals door Binet ontwikkeld, omvormt tot de Stanford-Binettest en daarmee de 'testing movement' in Amerika start (Feldhusen, 2003: 155). Hij interpreteert de resultaten op de test met behulp van een enkelvoudige index: het intelligentie-quotiënt. De maat voor IQ wordt gedefinieerd als de mentale leeftijd (testscore) gedeeld door de chronologische leeftijd, vermenigvuldigd met het getal 100 (Feldhusen, 2003: 163). Samen met Yerkes vormt hij de Stanford-Binettest om tot een screeningstest voor nieuwe recruten voor het leger. Door gebruik te maken van meerkeuzevragen is grootschalige afname van IQ-tests mogelijk: de Army-alpha- en de Army-bèta test. Na de oorlog werken Yerkes en Terman aan de ontwikkeling van intelligentietests voor onderwijs, daarbij samenwerkend met Thorndike. Zij ontwikkelen de National Intelligence Test voor 'grade 3 to 8' (Feldhusen, 2003: 159). Een paar jaar later ontwikkelt Terman samen met Kelley en Ruch de eerste gestandaardiseerde test van 'academic achievement', de SAT, die bij 350.000 kinderen wordt afgenomen. Deze intelligentietest wordt door scholen gebruikt om het onderwijs te evalueren en leerlingen te plaatsen voor academisch vervolgonderwijs (Feldhusen, 2003: 159).

Condliffe Lagemann stelt dat er tussen 1917 en 1928 zo'n 1300 tests zijn ontwikkeld in de VS en dat er in 1949 al 2600 tests beschikbaar zijn. Zij spreekt van een 'explosieve groei' (Condliffe

Lagemann, 2000: 88). In 1947 wordt in de Verenigde Staten 'The Educational Testing Service' (ETS) opgericht, een belangrijke stap voor het verzelfstandigen van psychologische tests en de wetenschap van testen. Van het oprichten van een dergelijk instituut gaat een dubbele werking uit: het bevestigt enerzijds het geloof in het nut van gestandaardiseerd meten in de context van onderwijs en anderzijds leidt de inzet van tests tot een geloof in tests en daarmee tot de overtuiging dat onderwijs op selectie gericht moet zijn: 'The very existence of ETS helped perpetuate existing educational practices. Its establishment was important in supporting the belief that standardized tests should play a central role in college and graduate school admission [...] in addition, belief in the value of admissions testing was essential to the idea that schools and colleges should serve selective as well as educational functions.' (Condliffe Lagemann, 2000: 156). Institutionaliseren van testen maakt van tests meer dan een hulpmiddel in dienst van een professional, maar vooral een bepaler van wat er in het onderwijs gebeurt.

Ook in Engeland groeit het vertrouwen in intelligentietests en de potentie daarvan. De Engelman Burt weet het in 1944 gedaan te krijgen dat leerlingen op de leeftijd van 11 jaar op basis van intelligentietests geselecteerd worden voor de 'Grammar School' (de school voor de 20 procent best scorende leerlingen) (Standaert, 2014: 115). Burt is ervan overtuigd dat intelligentie een statische en aangeboren competentie is, die goed meetbaar is (Burt, 1943: 83). Op basis van erfelijke vermogens is het volgens Burt mogelijk een rangordening te maken op basis waarvan 20% toegelaten kan worden tot het hoogste niveau van voortgezet onderwijs in Engeland, de 11-plus examination. Burt verdedigt deze scheiding als een verstandige stap: 'Het is in het belang van zowel de kinderen zelf alsook het land in zijn geheel dat degenen met de hoogste vermogens – de slimsten van de slimmen – zo nauwkeurig mogelijk worden uitgeselecteerd. Van de tot dusver beproefde methoden is het zogenaamde '11 en ouder onderzoek' verreweg het meest betrouwbare gebleken.' (Gould, 1996: 355). Nog altijd wordt dit elf-plus examen in delen van Engeland gebruikt voor de selectie bij nog resterende Grammar-scholen (Standaert, 2014: 115).

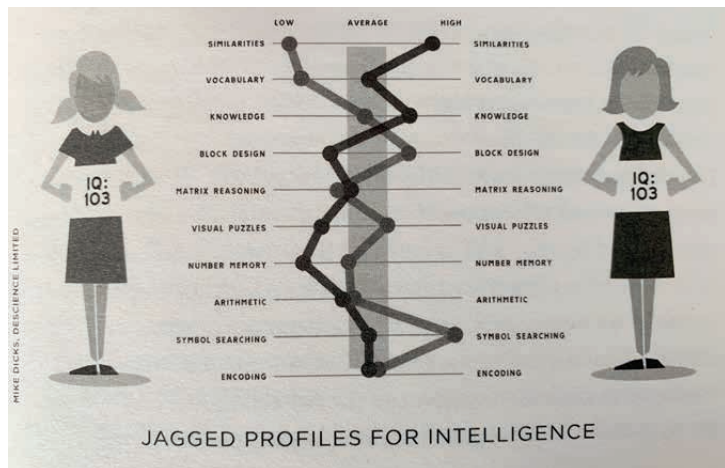
Testpsychologen hebben vanuit het idee dat intelligentie een vaststaand gegeven, een eigenschap is, zich beziggehouden met het ordenen van mensen op basis van dit gegeven en daarmee samenhangend met het voorspellen en plaatsen op basis van test scores. Met name in de Verenigde Staten is het gebruik van intelligentietests wijd verspreid en gekoppeld aan het doen van voorspellingen en selecteren voor de arbeidsmarkt zonder dat er overeenstemming is bereikt over wat intelligentie is, hoe deze te meten en of deze meetbaar is. Dit soort tests maakt onderdeel uit van de toelating tot universiteiten als Harvard, Yale en Princeton in de Verenigde Staten (Karabel, 2005). Het ontbreken van een inhoudelijk duidelijke definitie heeft de aantrekkelijkheid van het concept en het willen kunnen meten ervan niet gehinderd.

Het ongemak dat in de periode na de Tweede Wereldoorlog ontstaat over erfelijkheid en daarmee intelligentie, heeft de weg vrijgemaakt voor tests die aan dezelfde (psychometrische) kenmerken voldoen maar onder tal van andere benamingen hun weg naar het grote publiek hebben weten te vinden. In de Verenigde Staten leidt de gevoeligheid die in de loop van de vorige eeuw groeit ten aanzien van intelligentietests ertoe dat men ervoor kiest te spreken over 'geschiktheid' ('aptitude') en 'geschiktheidsmetingen' in plaats van 'intelligentie' en 'IQ-tests'. Daar waar de term 'intelligentie' uit de benamingen van tests is verdwenen, zijn de psychologische en psychometrische uitgangspunten gebleven, zoals bij 'achievement tests', 'qualifying tests', 'ability tests' en

'aptitude tests'. Hanson stelt dat het verschil beperkt is: 'The terminological distinction is of little substance.' (Hanson, 1993: 212). Intelligentie en geschiktheid zijn twee begrippen die nauw met elkaar verweven zijn geraakt vanwege het gevoelige debat over het al dan niet aangeboren zijn van intelligentie en het al dan niet veranderbaar zijn van intelligentie. De ontwikkeling van intelligentietests en de manier waarop intelligentie daarbij is geoperationaliseerd, hebben gemaakt dat intelligentie in de algemene beeldvorming is uitgegroeid tot een kenmerk dat onafhankelijk van het onderwijs vastgesteld kan worden en gezien wordt als een eigenschap waar mensen veel of weinig van hebben. Hanson stelt dat het aan de oppervlakte lijkt of intelligentie ontsproten is aan een deductief proces dat als volgt lijkt te verlopen: intelligentie is een enkel kenmerk, verschillend verdeeld over individuen; intelligentie wordt gemeten met een intelligentietest; individuen scoren verschillend op zo'n test; dus heeft iedereen een verschillende hoeveelheid intelligentie die gekoppeld is aan de score op zo'n test. Maar in feite start volgens Hanson de redenering bij de conclusie en is het denken als 'inductief' te karakteriseren: individuen behalen bepaalde scores op een intelligentietest, intelligentietests meten intelligentie, iedereen heeft dus zoveel intelligentie als met de score op een intelligentietest wordt vastgesteld, dus is intelligentie een ding dat verschillend verdeeld is over individuen (Hanson, 1993: 295).

Rose wijst in 2016 weer opnieuw op het feit dat voor intelligentie het 'jaggedness-principle' geldt: op tal van subtests van intelligentie scoren mensen verschillend waardoor feitelijk van ieder individu een eigen, ander profiel ontstaat. Dit is ook het uitgangspunt dat Binet bij de ontwikkeling van zijn schaal hanteert. Dit is echter een gegeven dat bij de doorontwikkeling van IQ-tests naar de achtergrond is geschoven. Door complexe, menselijke eigenschappen op een eendimensionale schaal te uitdrukken

in één getal, waarmee vervolgens geordend en geselecteerd kan worden, gaat alle onderliggende variëteit verloren. Daarmee wordt elke individualiteit volgens Rose ontkend en is aan de score geen individuele opbouw meer te herkennen. Terwijl de score per individu geheel anders kan zijn opgebouwd, is dat aan



Figuur 10 Jaggedness

de oppervlakte niet meer zichtbaar en lijken mensen met eenzelfde score op bijvoorbeeld een intelligentietest 'hetzelfde' terwijl ze een volledig verschillend, 'jagged', patroon van scores op de verschillende onderdelen kunnen hebben (Rose, 2016: 81). Talenten, welke dan ook, zijn altijd multidimensionaal en zijn niet tot een en dezelfde schaal terug te brengen of op te tellen tot één score waarmee vergelijkingen gemaakt worden (Rose, 2016: 85). Daarbij interacteren contexten altijd met persoonlijkheidskenmerken. Persoonlijkheidskenmerken, traits, zijn geen vaststaande gegevens maar worden beïnvloed door de omgeving.

In 1976 schrijft onderwijskundige Bloom over de rol die onderwijs kan spelen in het laten bereiken van onderwijsdoelen door alle leerlingen. In het voorwoord kijkt hij terug op de ontwikkelingen in het denken over de mogelijkheden van leerlingen. In de jaren vijftig was volgens Bloom het beeld dat er goede en zwakke leerlingen zijn. In de jaren zestig verandert dat en is de gedachte dat er snelle en langzame leerlingen zijn. In dit boek staat Bloom op het standpunt dat de meeste leerlingen op een vergelijkbaar punt van leren kunnen komen mits ze te maken krijgen met stimulerende leeromstandigheden. Variaties in leren en het niveau dat leerlingen bereiken wordt bepaald door de leerervaringen van leerlingen en de kwaliteit van de instructie die ze krijgen. Bloom pleit ervoor individuele verschillen die aanvankelijk tussen leerlingen bestaan, te negeren en onderwijs te richten op kwalitatieve instructie voor alle leerlingen, ongeacht hun capaciteiten (Bloom, 1976). Onderwijskundige Didau stelt dat onderwijs erop gericht moet zijn leerlingen slimmer te maken, niet slimmer dan een ander in vergelijkende zin, maar onderwijs moet het niveau van slimheid van ieder kind zoveel mogelijk verhogen. Intelligentie is volgens Didau een mechanisme waarmee kennis kan worden verworven. De context waarin iemand functioneert, ook de onderwijscontext, moet worden verbeterd om dit mechanisme beter zijn werk te kunnen laten doen (Didau, 2019: 58). Daarvoor is het van belang een aantal mythes over intelligentie los te laten zoals het veronderstellen dat intelligentie een natuurlijke, aangeboren capaciteit is voor denken en leren. Het gebruik van IQ-tests voor selectie in onderwijscontexten zal nergens toe leiden omdat uitkomsten meer zeggen over de sociaal-economische status dan dat ze kunnen dienen als een redelijke 'proxy' van intelligentie (Didau, 2019: 82).

Internationaal gezien blijft de discussie over intelligentie en met name het al dan niet aangeboren zijn ervan, opkomen en weer gevoerd worden. Ook in een tijd waarin vooral gekeken wordt naar de rol van sociaal-culturele factoren die voor bepaalde groepen belemmerend zijn om aan het onderwijs deel te nemen en daar succesvol te zijn, komt aandacht voor de rol van erfelijke bepaaldheid van intelligentie terug evenals voor de mate waarin onderwijs kan bijdragen aan de ontwikkeling van intelligentie. Het moet niet gaan over de vraag of en in hoeverre iets erfelijk bepaald is maar over de kneedbaarheid en flexibiliteit van een individu en daaraan gerelateerd de vraag hoe onderwijs kan bijdragen aan de ontwikkeling van een individu. Onderwijs is er om het effect van ongelijke omgevingsfactoren te verkleinen door kinderen krachtig, cultureel rijke kennis aan te bieden op een manier die kinderen vooruit helpt in hun latere leven: 'We can all get cleverer.' (Didau, 2019: 279).

De kennelijke aantrekkelijkheid van het concept van intelligentie, de overtuiging dat het met objectieve tests vast te stellen is en dat er ordeningen mee kunnen worden aangebracht, verklaren voor een belangrijk deel het aanvankelijke succes van intelligentietests en alle varianten die op dezelfde manier werken. Een belangrijke kracht in de legitimering van dit soort tests ligt echter niet alleen in de tests zelf maar in de (eenvoudige) toepassing ervan in onderwijs en onderzoek. Dit verklaart voor een belangrijk deel dat wetenschappelijke ontwikkelingen die in de loop van de vorige eeuw wijzen op de gebrekkige waarde van intelligentietests, niet hebben geleid tot een rem op de ontwikkeling en het gebruik ervan. Belangrijk in dit verband is het Coleman rapport uit 1966. Dit rapport laat zien welke mogelijkheden psychologische tests bieden voor nieuwe vormen van onderzoek in de sociale wetenschappen.

## 8.2.2 Het Coleman-rapport

In 1966 verschijnen de uitkomsten van een grootschalig onderzoek naar kansengelijkheid in het onderwijs in de Verenigde Staten onder leiding van de socioloog Coleman, uitgevoerd in opdracht van de regering van de VS. Het rapport 'Equality of Educational Opportunity' (1966) van meer dan 700 pagina's is de geschiedenis ingegaan als het 'Coleman-rapport'. Het onderzoek moet antwoord geven op de vraag of er in publiek onderwijs sprake is van gelijke onderwijskansen ongeacht ras, kleur, religie of nationaliteit. Het onderzoek heeft grote relevantie in het licht van het debat over segregatie in de VS in de jaren zestig, ook in onderwijs. In die tijd is niet veel informatie beschikbaar over de scholen in de VS, noch waar publieke middelen aan worden besteed, noch wat de resultaten zijn van onderwijs. Coleman is de eerste die, los van de opdracht van onderzoek naar segregatie, wil weten wat de opbrengsten zijn van het onderwijs en wat daarop van invloed is. Mede door de vooruitgang in de technologie is het Coleman en zijn onderzoekers gelukt om een onderzoek uit te voeren waarbij meer dan 650.000 leerlingen zijn betrokken. Er wordt voor het eerst in sociologisch onderzoek op grote schaal gebruik gemaakt van een testbatterij bestaande uit reeds gebruikte, gestandaardiseerde tests ontwikkeld door het instituut 'Educational Testing Service', ETS. De tests worden geacht inzicht te geven in de opbrengsten van het onderwijs, van een school en van het geheel van scholen. Tot voor het Coleman-rapport gaat de discussie over kwaliteit van onderwijs veel meer over de uitgangssituatie en het proces en wordt de kwaliteit gemeten op basis van variabelen als het aantal boeken in de schoolbibliotheek of het aantal uren dat wordt besteed aan verschillende vakken. In dit onderzoek worden de gemiddelde scores van groepen (clusters) van leerlingen met behulp van correlatief onderzoek vergeleken met tal van variabelen. De uitkomst dat er in 1966 sprake is van segregatie, is niet wat het rapport spraakmakend maakt. Voor het eerst wordt met data ondersteund dat het milieu waaruit iemand afkomstig is, gecombineerd met een mix van sociaal-economische factoren, het meest bepalend is voor wat een kind leert. Het rapport is ook het eerste waarin een 'prestatiekloof' tussen 'African-American' kinderen en blanke kinderen wordt gerapporteerd. De uitkomsten van dit rapport zijn de geschiedenis ingegaan met als belangrijkste conclusie dat de bijdrage van de school op het al dan niet realiseren van gelijke kansen beperkt is. De Nederlandse historicus De Rooy vat het rapport als volgt samen: 'De wel zeer verrassende conclusie luidde: de school deed er niet zoveel toe. De verschillen in schoolresultaten werden in hoofdzaak bepaald door sociaal-economische en culturele factoren.' (De Rooy, 2018: 212). Ook de kwaliteit van scholen of van docenten blijkt zeer beperkt de uiteindelijke prestaties van leerlingen te bepalen.

De uitkomsten van het onderzoek hebben wereldwijd veel impact gehad. Los van de uitkomsten heeft vooral de manier van onderzoek doen een grote invloed op de manier van denken over onderzoek naar effecten van onderwijs: 'Regardless of its influence on policy, the Coleman Report had an extremely productive influence on education research. It stimulated a host of studies about school effects.' (Condliffe Lagemann, 2000: 199). Met zijn onderzoek heeft Coleman volgens de Canadees Jacobs de weg vrijgemaakt voor een bepaalde invulling én vaststelling van het concept kansengelijkheid. Coleman heeft breed geïntroduceerd dat er sprake is van kansengelijkheid als de uitkomsten van het onderwijs voor verschillende groepen leerlingen gelijk zijn: 'The Coleman Report was a milestone in how we conceive of equality of educational opportunity because it coincided with the emergence of, indeed crystalized, the model of equality of educational opportunity focused on student outcomes.' (Jacobs, 2016: 319). Opbrengsten van

onderwijs zijn de graadmeters van kansengelijkheid en daarmee schatters van maatschappelijk succes. En onderwijsresultaten zijn meetbaar: 'The rationale for evidence-based policy making is that student outcomes are something that can be measured in tangible ways that enable the gathering of empirical evidence.' (Jacobs, 2016: 319). Het rapport maakt de weg vrij voor onderzoek waarbij opbrengst van onderwijs centraal gesteld wordt én in het licht van kansengelijkheid én voor het idee dat er sprake is van kansengelijkheid als groepen niet verschillen wat betreft de uitkomsten van het onderwijs: opbrengst als indicator van kansengelijkheid. Een opbrengst die met behulp van objectieve criteria geformuleerd én gemeten kan worden. Jacobs stelt dat tevens een basis is gelegd voor het competitie-denken dat inherent verbonden is geraakt met onderwijs: 'For when equality of educational opportunity is framed around the idea of providing ladders of opportunity for upward mobility, competition is inherent to what is valuable about the education system and is thus beyond challenge.' (Jacobs, 2016: 314). En dat we zijn gaan accepteren dat onderwijs een competitie is met als belangrijk kenmerk dat winnaars en verliezers daarbij horen: 'The school system is structured so that some children win, while some lose.' (Jacobs, 2016: 313). Het Coleman-rapport heeft de weg vrijgemaakt voor een andere manier van kijken naar onderwijs. Het Coleman-rapport koos schoolsucces als uitgangspunt, zo stelt de Rooy: 'Dit betekent dat de 'eerlijkheid' van een onderwijsbestel in het algemeen en een school in het bijzonder wordt afgemeten aan de resultaten die door de leerlingen worden behaald. Deze resultaten worden dan in feite beschouwd als een maat voor de 'effectiviteit' van de leerkrachten. Factoren als stimulering door ouders, ziekte en geringe motivatie worden daardoor systematisch buiten beschouwing gelaten.' (De Rooy, 2018: 223).

In de jaren na het verschijnen van het rapport komt er kritiek op de tests die Coleman in zijn onderzoek heeft gebruikt. De tests zoals die in het onderzoek zijn gebruikt, sluiten niet aan op het onderwijs dat de leerlingen hebben gehad. Een belangrijke criticaster is Carver. Hij stelt dat het vergelijken van scholen met elkaar op basis van tests die juist bedoeld zijn om verschillen tussen leerlingen te maximaliseren, niet geschikt zijn om schoolprestaties mee vast te stellen. Omdat er geen gebruik is gemaakt van criterium-gerelateerde tests, die aansluiten op het onderwijs zoals dat in de scholen wordt gegeven, kent het onderzoek van Coleman volgens Carver een belangrijke vorm van bias: 'The Coleman Report used tests and statistics that were biased against finding large differences associated with schools which might be interpreted as attributable to educational treatment differences. Those individuals in charge of the Coleman study should have used criterion-referenced tests or edumetric tests if they were interested in the impact of school differences upon achievement. Instead they used psychometric tests and found what psychometric tests are designed to reflect, i.e., large individual differences.' (Carver, 1975: 85).

Het Coleman-rapport is van grote invloed geweest, niet alleen op het gevoerde beleid in tal van westerse landen, maar ook op de manier waarop onderzoek in onderwijs wordt opgezet en uitgevoerd. Met het rapport is een trend gezet om met tests die vergelijkenderwijs meten, rendement van onderwijs vast te stellen en om opbrengsten van onderwijs voor verschillende groepen met elkaar te kunnen vergelijken. Deze destijds nieuwe manier van denken heeft een enorme vlucht genomen in (onderwijs)sociologisch onderzoek, terwijl ook de onderwijskunde leunt op testcores als basis voor vergelijkend onderzoek. Zo zijn er sinds het Coleman-rapport tal van onderzoeken verschenen waarin scores op tests het 'vaste' gegeven vormen waarmee tal van variabelen worden vergeleken. Coe en Taylor Fitz-Gibbon vatten de problemen van het meten

van onderwijsopbrengsten, 'School Effectiveness Research', samen en komen tot de slotsom dat het doen van dit soort onderzoek niet veel heeft opgeleverd: 'Much research on school effectiveness has been characterised by largely overstated claims and poor modelling.' (Coe&Taylor Fitz-Cibbon, 1998: 421).

Een vergelijkbaar effect als van het Coleman-rapport destijds gaat nu uit van de onderzoeken naar opbrengsten van onderwijs zoals de OECD die uitvoert. Ook hier staan de scores op internationaal vergelijkende tests centraal maar deze zijn tegelijkertijd boven elke discussie verheven geraakt. Het debat gaat met name over de resultaten en zelden over de waarde van de resultaten, gegeven de tests die worden gebruikt. De waarde ervan wordt bijna bij voorbaat serieus genomen. Bij de interpretatie van gegevens raakt de test en wat wel en niet aan de test kan worden ontleend naar de achtergrond en de uitkomsten worden feiten.

### 8.2.3 No Child Left Behind

In de Verenigde Staten is het president Bush die bij zijn installatie in 2001 zegt serieus werk te willen maken van onderwijshervormingen en komt met de 'No Child Left Behind Act' (NCLB-act) en de 'Education Sciences Reform Act'. De eerste is bedoeld om rekenschap te geven over het peil van het onderwijs via tests en de tweede act voorziet in het verzamelen van statistisch materiaal en 'scientifically based research' via het Institute of Education Sciences. Concreet leidt dit beleid tot het jaarlijks afnemen van 45 miljoen gestandaardiseerde, in opdracht van de staat ontwikkelde tests. De ontwikkeling, afname en verwerking gebeurt door enkele grote spelers op de testmarkt, waaronder het ETS, die daarmee een aanzienlijke groei in hun omzet weten te realiseren (Standaert, 2014: 59). Ravitch, onderwijshistoricus en voormalig lid van de National Assessment Governing Board van de Verenigde Staten, beschrijft hoe tests het Amerikaanse onderwijs in hun greep hebben gekregen en wat daarvan de gevolgen zijn. De Rooy stelt dat Ravitch gezien kan worden als een betrouwbare getuige omdat zij aanvankelijk een warm voorstandster van de NCLB-act was (De Rooy, 2018: 244). De roep om rekenschap over het peil van het onderwijs ontstaat in de Verenigde Staten als reactie op het rapport 'A nation at risk' (ANAR), een onderzoek waarin wordt gerapporteerd over de kwaliteit van het onderwijs in de VS naar aanleiding van onderzoeken die wijzen op een achteruitgang in SAT-scores die worden gebruikt bij de toelating tot universiteiten. De conclusies van het rapport 'A nation at risk' leiden tot onrust, wat ook de bedoeling is van het rapport: 'The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people.' (Ravitch, 2010: 27). Het rapport stelt vervolgens dat 'our educational institutions seem to have lost sight of the basic purposes of schooling, and of the high expectations and disciplined effort needed to attain them' (Ravitch, 2010: 27). Het rapport komt vooral met aanbevelingen gericht op het bereiken van het ideaal van 'equality of educational opportunity': de natie zou economisch en sociaal onder druk komen te staan als het onderwijs niet dramatisch zou verbeteren voor alle kinderen. Dit wordt in het rapport vertaald in aanbevelingen die vooral gericht zijn op strengere vereisten voor 'high school grading', hogere standaarden voor academische prestaties, verbeteringen van leerlinggedrag en verbetering van het niveau van docenten door strengere opleidingseisen en meer salaris. Ravitch stelt dat het rapport vooral belangrijk is geweest voor wat er niet staat (Ravitch, 2010: 28). Na het verschijnen van het rapport is het gemakkelijk geworden sociale en economische problemen in de schoenen van het onderwijs te schuiven. Het rapport suggereert volgens Ravitch dat een sterk curriculum en hoge academische standaarden voldoen-

de zijn om de economische en sociale ziekten van de natie aan te pakken (Ravitch, 2010: 32). Ravitch zegt dat wetenschappers sinds het verschijnen van het rapport zich afvragen of dit rapport een accurate beschrijving geeft van de situatie in het Amerikaanse onderwijs, of dat het een voorbeeld is van alarmerende retoriek van de regering Reagan (Ravitch, 2010: 27). In ieder geval maakt het rapport de geesten rijp voor de maatregelen waar de regering Bush junior een kleine tien jaar later mee komt. NCLB is volgens Ravitch ontdaan van onderwijskundige aanpakken en is te karakteriseren als een technocratische aanpak van de schoolproblemen. Succes van onderwijs is teruggebracht tot de scores op gestandaardiseerde toetsen gericht op de twee basisvaardigheden taal en rekenen met het idee dat het beperken tot die twee vaardigheden zou leiden tot een natie die de economische competitie met de rest van de wereld aan zou kunnen (Ravitch, 2010: 33). Democraten én republikeinen scharen zich achter de NCLB-act: de eersten zijn vooral blij met de federale aanpak en de tweede groep waardeert het dat accountability bij wet wordt vastgelegd. En zo is er breed draagvlak ontstaan voor het idee dat tests en accountability leiden tot betere scholen en een verhoging van de kwaliteit van onderwijs (Ravitch, 2010, Standaert, 2014). Hoewel ANAR en NCLB wezenlijk van elkaar verschillen hebben ze volgens Ravitch met elkaar gemeen dat beide zich alarmerend uitlaten over de staat van het onderwijs in casu de scholen. Dat is naar haar overtuiging niet terecht gebleken maar heeft volgens haar zeker wel de aandacht kunnen afleiden van het nemen van maatregelen die wel direct de kansen(on)gelijkheid hadden kunnen aanpakken (Ravitch, 2010: 34). De Rooy stelt dat Ravitch tot het inzicht komt dat het beleid tot averechtse effecten leidt. 'Er verspreidt zich een afrekencultuur die de professionaliteit van leerkrachten aantast, het leidt tot een overvloed aan tests die moeten worden afgenomen en die feitelijk niet veel zeggen over de vorderingen van leerlingen. De leerresultaten verbeteren niet en de ongelijkheid neemt toe.' (De Rooy, 2018: 244). Koretz stelt dat toetsing, 'achievement testing', aanvankelijk de functie had om het lesgeven te verbeteren op basis van de scores. Maar het resultaatgedreven meten dat in gang is gezet met NCLB heeft volgens hem de boel omgedraaid, waarmee tests leidend zijn geworden: 'Tests would now lead. Improving performance on the specific test was to be the explicit goal, and higher-quality instruction would be the consequence. This was the tail wagging the dog.' (Koretz, 2017: 26).

Putnam gebruikt meer recent in zijn boek 'Our Kids, The American Dream in Crisis' (2015) scores op de SAT-test om te onderbouwen dat de VS een land zijn geworden waarin kansengelijkheid niet is toegenomen maar minder is geworden. Ook hij gebruikt de waarde van de score als 'vast' gegeven en de test als leverancier van dat vaste gegeven. Dergelijke onderzoeken versterken het beeld van een test als een betrouwbare en stabiele indicator van een werkelijkheid. Dit gebruik van tests in onderzoek heeft de positie en waarde van tests verder versterkt. Het feit dat onderzoek gebruik maakt van een bepaalde test geeft tests in het algemeen maar die specifieke test in het bijzonder meer belang, waarde en waarheid. Tests, met name intelligentietests, hebben in de Verenigde Staten en het Verenigd Koninkrijk vanaf het begin van de vorige eeuw snel en gemakkelijk hun weg gevonden naar het onderwijs. In Nederland is de weg anders en langer geweest.

### **8.3 Empirisch-analytische wetenschapsontwikkeling in Nederland**

Het behaviorisme is een stroming die Nederland in de loop van de twintigste eeuw bereikt, maar hier in het onderwijs niet gelijk zijn weg vindt, in tegenstelling tot de Verenigde Staten. In 1928 geeft J.B. Watson de volgende omschrijving van het behaviorisme: 'De behaviorist vraagt zich af, waarom maken we niet tot voorwerp van de psychologie wat wij kunnen observeren? Laten wij



ons beperken bij dingen die kunnen geobserveerd worden en wetten formuleren over hetgeen wij observeren' (Standaert, 2014: 113). Invloedrijke pedagogen in Nederland, onder leiding van onder anderen Kohnstamm en Langeveld, hebben zich verzet tegen een behavioristische manier van kijken en onderzoek doen in onderwijs. Onder psychologen zijn het vooral testpsychologen die zich aangetrokken voelen tot een meer natuurwetenschappelijke houding ten aanzien van de menswetenschappen, zoals testpsycholoog Luning Prak: 'In de psychotechniek wil men door middel van proeven de toekomst voorspellen, zoals een keuring op trekvastheid van een metaal [...] dat ook doet.' (Dehue, 1990: 199). In het bedrijfsleven groeit in Nederland aanvankelijk wel de belangstelling voor een meer natuurwetenschappelijke benadering van het selectieproces als bijdrage aan een efficiënte bedrijfsvoering onder het motto van 'de juiste man op de juiste plek'. De resultaten in termen van succesvolle voorspellingen over (functie)geschiktheid vallen tegen en daarmee worden de beperkingen van de psychotechniek voor de psychologische praktijk zichtbaar. Dit leidt ertoe dat volgens Haas binnen de selectiepraktijk de aandacht langzaam verschuift richting een 'bredere interpretatieve psychodiagnostische benadering' (Haas, 1995: 45), waarbij het gebruik van apparaten en gestandaardiseerde tests als intelligentietests naar de achtergrond verdwijnt als het gaat om het bepalen van geschiktheid. Intelligentietests worden hulpmiddelen voor de psycholoog die de scores gebruikt in zijn psychodiagnostische benadering: 'De uitkomsten van deze tests trachtte de psycholoog te interpreteren binnen het raamwerk van de gehele persoonlijkheid. De psycholoog zelf was het belangrijkste instrument in het beoordelen van de persoon die hij tegenover zich had.' (Haas, 1995: 45). In de selectiepraktijk is het afnemen van intelligentietests ingebed in een set van instrumenten waarmee ook andere aspecten worden onderzocht. Zo zijn intelligentietests nooit verdwenen en hebben tijd en ruimte gekregen zich verder te ontwikkelen.

Voor de aansluiting van basis- en voortgezet onderwijs zijn intelligentie en het testen ervan volgens Haas lange tijd ontoereikend en ondoenlijk geacht om ingezet te worden bij de bepaling van schoolgeschiktheid. Dit geldt echter niet voor het speciaal onderwijs. Daar zien met name artsen het als een van hun taken om de psychologische selectie voor die vormen van onderwijs onder hun hoede te nemen (Haas, 1995: 36) en maken daarbij gebruik van intelligentietests. Haas suggereert dat dit mogelijk verband houdt met het feit dat de grondlegger van de psychologische test, Alfred Binet, ook een arts was. Daarmee is het buitengewoon (of later: speciaal) onderwijs het eerste toepassingsgebied waarbinnen psychologische tests worden ingezet (Haas, 1995: 120). Dit onderwijs is een relatief dure vorm van onderwijs, wat maakt dat selectie nodig wordt geacht. Dit gebeurt met behulp van psychologische tests die zo, met de deskundigen in hun kielzog, wel langzaam hun intrede doen in het onderwijs (Haas, 1995:123).

Vanaf de jaren vijftig van de vorige eeuw, neemt het verzet toe tegen de interpretatieve psychodiagnostische benadering bij psychologisch onderzoek. Voor tests in de professionele psychologenpraktijk moet kunnen worden aangetoond dat gebruikte tests daadwerkelijk meten wat ze zeggen te meten. Onderzoek moet dit aantonen. Dit onderzoek gebeurt door de uitkomsten van tests te vergelijken met een tweede oordeel. Er wordt gezocht naar een tweede oordeel 'dat als criterium kon dienen om de waarde van de test aan af te meten' (Dehue, 1990: 201). In de praktijk van werving en selectie voor het bedrijfsleven komt dat erop neer dat directies of leidinggevenden wordt gevraagd een beoordeling te geven van de geschiktheid van hun personeel en vervolgens te onderzoeken of een test tot een vergelijkbare beoordeling komt. In dat

licht is het van belang dat beide oordelen onafhankelijk van elkaar tot stand komen. Daarmee groeit een praktijk waarbij testcores worden gezien als 'voorspellers' van criteriumcores: 'om de waarde van de test zo overtuigend mogelijk te kunnen verdedigen, moesten de eerste testuitslagen worden gezien als verantwoording van een verwachting omtrent criteriumcores. In termen van de hedendaagse methodologie moesten deze testcores het criterium voorspellen.' (Dehue, 1990: 202). Deze zogenaamde 'correlationele' werkwijze vindt men lastig te verenigen met dan gangbare uitgangspunten van de geesteswetenschappelijke psychotechniek, met name de fenomenologie, omdat 'er geen ruimte bestaat om juist de unieke aspecten en vermogens van deze ene persoon bloot te leggen' (Dehue, 1990: 203). Daarmee komt de fenomenologie onder druk te staan. De wens groeit om ook de geesteswetenschappen een meer natuurwetenschappelijk aanzien te geven.

De Groot, leerling van Luning Prak, weet de in Nederland gegroeide tegenstelling tussen de interpretatieve psychologische praktijk en de psychometrische benadering dichter bij elkaar te brengen, en zelfs met elkaar te verbinden. De Groot werkt zijn denken over het doen van voorspellingen op basis van op objectieve wijze verkregen data uit tijdens zijn sabbatical van 1959 tot 1960 in de VS in zijn boek 'Methodologie, grondslagen van onderzoek en denken in de gedragswetenschappen'. Ware kennis en voorspellen zijn voor hem nauw met elkaar verbonden: 'Het criterium bij uitstek voor "ware kennis" is gelegen in het kunnen voorspellen van de uitkomst van een toetsingsprocedure.' (De Groot, 1961: 20). De Groot wijst 'op de noodzaak van objectiviteit en daarmee op de wetenschappelijkheid van de psychodiagnostiek' (Haas, 1995: 46). En die objectiviteit kan met een meer fenomenologische aanpak in zijn optiek niet voldoende worden geborgd. De Groot doet de volgens Dehue 'methodologische vondst' om de – aan de correlationele benadering ontleende – opvatting van een diagnose als een verwachting over een criterium, te combineren met de geesteswetenschappelijke leerstelling dat een persoon 'in zijn uniciteit' moet worden bestudeerd' (Dehue, 1990: 204). De Groot stelt dat de uitspraken die middels de psychodiagnostiek worden verworven 'niet als conclusies opgevat mogen worden, maar zouden moeten worden beschouwd als gereedschap om te komen tot hypothesen waaruit toetsbare voorspellingen moesten worden afgeleid' (Dehue, 1990: 204). De fenomenologische methode wordt op die manier gezien 'als een hulpmethode van de hypothesevorming' (Busato et al., 2014: 280). De Groot heeft daarmee een bestaande rationaliteit in een nieuwe weten in te voegen, zodanig dat acceptatie ervan in Nederland op grote schaal mogelijk wordt. De Groot verheft 'de voorspelling tot het methodologisch sleutelbegrip voor de hele psychologie en voor de sociale wetenschappen in het algemeen' (Dehue, 1990: 206). Zo weet hij uiteindelijk de empirisch-analytische wetenschapsbenadering bij psychologen, maar ook bij pedagogen en later onderwijskundigen op de kaart te zetten. De 'voorspellingsmethodologie' waar De Groot mee komt is volgens Dehue niet 'puur een afspiegeling of een rechtstreekse concretisering van het algemene empirisch-analytische denken over wetenschap. Deze methodologie is dus bij nadere beschouwing geen prototype van wat overal onder wetenschappelijke rationaliteit wordt verstaan, maar vertoont eigenzinnige trekken.' (Dehue, 1990: 198). Het is de vraag of hier sprake is van een 'eigenzinnige' interpretatie van de vigerende wetenschappelijke rationaliteit of eerder van een 'strategisch slimme zet' maar feit is wel dat De Groot een grote stempel zet op de manier waarop men in de gedragswetenschappen 'onderzoek' gaat zien, interpreteren en uitvoeren.

Zijn denken over toetsconstructie werkt De Groot samen met Van Naerssen uit in het boek 'Stu-

dietoetsen' dat in 1969 verschijnt. Een eerste stap in de toetsconstructie is volgens De Groot de zoektocht naar het criterium. De Groot en Van Naerssen maken onderscheid in 3 soorten criteria die oplopen in mate van concretisering: uiteindelijke, intermediaire en onmiddellijke criteria. Uiteindelijke criteria zijn zo vaag dat een 'vertaling' van uiteindelijke criteria, via intermediaire criteria naar onmiddellijke criteria nodig is. Pas met een onmiddellijk criterium 'hebben we als het ware vaste grond onder de voeten gekregen. We kunnen nu meten en voorspellen' (De Groot & Van Naerssen, 1969: 241). Ook later, in 1986, in zijn boek 'Begrip van Evalueren' wijst hij op het belang van goed uitgewerkte onderwijsdoelen voor meetcontexten: 'Onderwijsdoelstellingen moeten operationeel gedefinieerd kunnen worden als criteriumgedrag van leerlingen' (De Groot, 1986: 64). Hij voegt daar aan toe dat 'zulk criteriumgedrag op de een of andere wijze meetbaar moet zijn, althans objectief constateerbaar' (De Groot, 1986: 64).

De Groot wordt bij een breder publiek bekend als in 1966 zijn boek 'Vijven en Zessen' verschijnt, dat volgens Busato eerder getypeerd kan worden als 'een opruiend pamflet, als een provocerende aanklacht tegen het Nederlandse Onderwijssysteem' (Busato et al., 2014: 290). De Groot karakteriseert daarin het onderwijs als een permanent of zelfs 'immanent' systeem van selectie (De Groot, 1966: 189). Dit alles wil zeker niet zeggen dat De Groot tegen selectie is. Integendeel. Hij vindt selectie, op zichzelf, 'geen euvel'. Hij ziet selectie als een onvermijdelijk en noodzakelijk gebeuren, een 'onontkoombare en in principe nuttige sociale activiteit' (De Groot, 1966: 189). Het competitieve element dat gepaard gaat met cijfers geven heeft meer voordelen dan nadelen in zijn optiek als 'voorbereiding op maatschappelijke competitie, het sportieve element ook, en de stimulering tot het leveren van prestaties' (De Groot, 1966: 26). De problemen die De Groot heeft met selectie in het onderwijs zijn gekoppeld aan de praktijk van het beoordelen, en dan vooral het beoordelen zonder heldere maatstaven. Want het feit dat het onderwijs niet over heldere doelen beschikt, maakt dat heldere maatstaven voor beoordeling ontbreken en dus dat cijfers feitelijk geen absolute waarde hebben en dus niet zonder meer met elkaar te vergelijken zijn. En dat laatste is nu juist wel wat er gebeurt in het onderwijs en dat is onrechtvaardig in de optiek van De Groot: 'Zulke beoordelingsmethoden mogen, voor de belangrijke beslissingen waar het in het onderwijs om gaat, niet worden gebruikt in een maatschappij die pretendeert ieder gelijke kansen te geven: zij moeten door betere worden vervangen.' (De Groot, 1966: 128). Het argument dat De Groot hier inbrengt over het bieden van gelijke kansen past bij de tijdgeest, maar niet zozeer bij het denken van De Groot. Met dat debat houdt hij zich niet bezig omdat hij niet denkt dat er veel te doen is aan ongelijkheid van kansen: 'Middelen om aan de milieu-handicap te ontkomen zijn alléén: het kind vroegtijdig uit zijn milieu halen òf het milieu zelf te veranderen.' (Samenwerkende Instituten, 1967: 89). Voor de oplossing van de problemen in het onderwijs wijst De Groot naar de Verenigde Staten omdat daar al jarenlange ervaring zou zijn met de oplossing van het probleem dat De Groot schetst, namelijk het gebruik van objectief scorebare studietoetsen. Tijdens zijn sabbatical in de VS heeft hij kennisgemaakt met het testinstituut daar, het ETS. De Groot pleit voor de oprichting van 'een Nationaal Instituut voor de Ontwikkeling van Studietoetsen' (De Groot, 1966: 232). Omdat De Groot naar eigen zeggen zelf te druk is voor het realiseren van een dergelijke ambitie en zichzelf ziet als iemand die 'te veelzijdig in onderwijszaken geëngageerd' is om dit op te pakken, eindigt hij zijn boek Vijven en Zessen met de zin: 'Wie bindt de kat de bel aan?' (De Groot, 1966: 246). Mertens karakteriseert Vijven en Zessen 'als een scherpe aanklacht tegen de 'gangbare praktijk' in het onderwijs en de leraar was van die praktijk 'de drager' (Mertens, 2016: 12). De neergang van de docent als professional is bij

Vijven en Zessen begonnen, zo stelt Mertens naar eigen zeggen 'wat polemisch' (Mertens, 2016: 11). De Rooy concludeert dat het boek van De Groot achteraf gezien kan worden als een 'merkwaardig' boekje omdat het feitelijk een herhaling van zetten was van het artikel van Posthumus uit 1940 waarin Posthumus wees op de 'wetmatigheid' waarmee aan elke toets in het onderwijs een verdeling van 25%, 50% en 25% werd toegekend: een kwart zakt, een kwart wordt als zeer goed beoordeeld en de rest, de helft, zit daar tussenin. De Groot voegt volgens De Rooy niet veel meer toe dan 'een opgewonden toon, met veel verontwaardigde uitroepstekens en termen als 'onzedelijk', 'crimineel', 'ondemocratisch!', en een onterechte beschuldiging aan het adres van Kohnstamm' (De Rooy, 2018: 209). Desalniettemin heeft het boek in 1966 veel succes, 'succes dat De Groot vleugels geeft en ervoor zorgt dat hij aanhoudend aanwezig is in het nationale onderwijsdebat' (Mertens, 2016: 17).

In Nederland blijven vormen van geschiktheidstests in lijn met de interpretatieve psychodiagnostische benadering, het domein van (test)psychologen. Een factor van belang hierbij is het Koninklijk besluit waarin vastgelegd wordt dat psychologen het alleenrecht voor het afnemen van psychologische tests krijgen voor de selectie voor de meeste vormen van buitengewoon onderwijs (Haas, 1995: 204). Toch stelt Haas dat met name de pedagogen de groei van psychologische tests in het onderwijs mogelijk hebben gemaakt. Aanvankelijk is er in het onderwijs nogal wat wantrouwen in de richting van psychologen en hun gestandaardiseerde (IQ-)metingen. Juist de pedagogen hebben uiteindelijk een belangrijke bijdrage geleverd aan het 'rijp maken van het onderwijs voor de gedachte dat tests een onmisbaar instrument vormen voor een verantwoord toelatingsbeleid' (Haas, 1995: 199). Tests in het licht van selectie (en afwijzen) stuit in de jaren zeventig op verzet in het onderwijs, maar de argumenten van pedagogen dat met behulp van tests de cognitieve ontwikkeling kan worden bepaald 'pasten beter bij het pedagogisch klimaat' (Haas, 1995: 202). Daarmee hebben pedagogen een belangrijke rol gespeeld in het vrij maken van de weg voor het gebruik van tests in het onderwijs: inzicht in de stand van het ontwikkelingsproces is een argument gebleken dat vanuit didactisch oogpunt moeilijk te weigeren blijkt.

De empirisch-analytische wetenschapsopvatting heeft een grote ontwikkeling doorgemaakt. Het feit dat De Groot zoveel waarde daarbij toekende aan testen, heeft het belang van tests verder doen toenemen en deze een niet meer weg te denken plaats in het onderwijs gegeven, daarbij geholpen door de oprichting van een zelfstandig toetsinstituut in de vorm van Cito.

### 8.3.1 Intelligentietests en De Groot

Busato stelt dat De Groot zich niet erg populair maakt 'als hij zich tevens ontpopt als pleitbezorger van de erfelijkheidshypothese van intelligentie, een onderwerp dat al vroeg in de carrière van De Groot zijn belangstelling heeft. Hij beweert dat meisjes, arbeiderskinderen en diverse rassen dommer zouden zijn dan respectievelijke jongens, kinderen uit hogere milieus, blanken en joden; die verschillen zouden voor een belangrijk deel genetisch zijn bepaald.' (Busato et al., 2014: 264). Ook Mertens wijst erop dat De Groot de opvatting heeft 'dat intelligentie in belangrijke mate aangeboren is. Er zijn verschillen tussen mensen en daar is niet veel anders mee te doen dan ze te erkennen' (Mertens, 2016: 1). Van Kemenade stelt dat De Groot de mening is toegedaan dat de overervingscoëfficiënt van de intelligentie in onze westerse samenleving zeer hoog is, dat er weinig reserve aan talent meer is en dat van onderwijsverbeteringen weinig effect op de begaafdheidsontwikkeling te verwachten is (Van Kemenade, 1981: 142). Als definitie van

intelligentie stelt De Groot dat intelligentie de vaardigheid omvat tot het oplossen van rationele problemen.

Met een uitgekristalliseerde aanpak in inhoud en gebruikte psychometrie voor het ontwikkelen van intelligentietests is er nog altijd geen overeenstemming bereikt over wat intelligentie is, of het meetbaar is en hoe het te meten is. Vroon onderscheidt twee 'stromingen' in de gedragswetenschap die geen van de beide het wetenschappelijke gelijk aan hun kant hebben weten te krijgen: aan de ene kant ziet Vroon de correlationist die gericht is op het willen voorspellen van gedrag en de verschillen tussen individuen en aan de andere kant is daar de cognitivist die vooral wil weten hoe gedrag tot stand komt en welke processen zich daarbij voltrekken (Vroon, 1980: 172). Het gedachtegoed van tests sluit in aard en ontwikkeling nauw aan op de relationele aanpak. Met het toekennen van belang dat wordt gehecht aan tests en de manier waarop gewerkt wordt bij het testen en de onderbouwing van de validiteit ervan, lijkt de relationele stroming het gelijk aan haar kant te hebben gekregen, zij het niet als een wetenschappelijk verkregen gelijk maar wel een in de praktijk gegroeid gelijk. Het definiëren van het begrip intelligentie blijft nog altijd in tautologische formuleringen hangen, wat de aantrekkelijkheid ervan niet lijkt te hinderen, ook niet voor De Groot. Tegen het gebruik van intelligentietests heeft De Groot ook geen principiële bezwaren. En over de discussie die in de wereld van intelligentietests woedt over wat nu precies intelligentie is en hoe dat het beste gemeten kan worden, maakt De Groot zich niet al te druk: 'Elke test-definitie is een betrekkelijk willekeurige specificatie van het begrip [intelligentie]' (De Groot, 1961: 91). De Groot is een pragmaticus en kiest liever voor beperkingen ten aanzien van 'wat' er gemeten wordt dan dat hij kiest voor aanpassingen in 'hoe' er gemeten wordt. Zo kijkt hij ook pragmatisch naar het debat over intelligentie. Uit wat er tot nu toe aan intelligentietests en data beschikbaar is, kan volgens De Groot psychometrisch zo langzamerhand wel achterhaald worden wat in het kader van intelligentie meten relevant is om te meten: 'In de tegenwoordige situatie kan men desgewenst de inhoud van het begrip intelligentie omschrijven door een opsomming van alle acceptabel geachte bepalingsmethoden en door verwijzing naar de algemene factor, die daaruit door factoranalyse empirisch zou kunnen worden getrokken, als een dergelijk omvangrijk testprogramma op een voldoende grote en representatieve steekproef uit de bedoelde populatie uitvoerbaar zou zijn.' (De Groot, 1961: 91). In het licht van zijn nadruk op empirie is het enerzijds verbazingwekkend met welk gemak De Groot omgaat met het feit dat een begrip, aspect of kenmerk lastig te meten is, zoals in het geval van intelligentie. Anderzijds past het bij zijn gerichtheid op het willen kunnen voorspellen: liever een beperkte meting maar dan wel goed dan een goede afspiegeling van hetgeen gemeten moet worden maar dan met een meting waar de nodige vraagtekens bij kunnen worden gezet.

### 8.3.2 Intelligentietests en onderwijs

Omdat testen in onderwijs het domein wordt van psychologen die 'onafhankelijk van het genoten onderwijs' gestandaardiseerd een oordeel moeten geven, is het niet zo verbazingwekkend dat zij hun kennis en ervaring met intelligentietests 'meenemen' naar de context van onderwijs. Deen maakt melding van de duidelijke overlap die er is tussen algemene intelligentietests en schoolprestatietoetsen. Het meten van kennis, onafhankelijk van genoten onderwijs, is lastig en het meten van geschiktheid zonder kennis mee te meten is minstens even lastig. Deen wijst onder andere naar een Leuvens onderzoek waarbij hij stelt dat de 'Algemene Intelligentietest in zijn subtests duidelijk het karakter heeft van een schoolprestatie-onderzoek' (Deen, 1969: 84).

Dat de verschillen tussen intelligentietests en 'achievement tests' lastig te herleiden zijn, heeft voor een belangrijk deel te maken met het feit dat hetgeen gemeten wordt gezien wordt als een indicator van toekomstig functioneren. In het onderwijs toegepast betekent dit dat 'eerder schoolsucces wordt gezien als goede voorspeller van toekomstig schoolsucces'. Daarmee vertonen tests in benaming, in wat er daadwerkelijk gemeten wordt en waar het voor gebruikt wordt, behoorlijke verwantschap. Hoewel het verzet tegen het gebruik van intelligentietests en tegen de waarde ervan toeneemt, 'bleef dit routinematig wel deel uit maken van de werkzaamheden rondom het aansluitingsvraagstuk' (Haas, 1995: 187). Maar, zo stelt Haas, het gebruik ervan verandert langzamerhand. Het gaat niet zozeer om het vaststellen van de intelligentie met behulp van tests op het gebied van taal, rekenen en algemene ontwikkeling maar met behulp van intelligentietests wordt inzicht gekregen in specifieke vaardigheden: 'De intelligentietest diende inzicht te geven in specifieke vaardigheden van het kind op gebieden als rekenen, taal en algemene ontwikkeling.' (Haas, 1995: 187). Daarmee verdwijnen oorspronkelijke intelligentietests niet van het toneel maar verandert de functie ervan: het verandert van doel naar middel. Bovendien gaan de intelligentietests volgens Haas deel uitmaken van bredere testbatterijen om zo een meer volledig beeld van een kind te krijgen. De neiging te willen zwijgen over de onderwerpen 'eugenetica' of 'sociaal-darwinisme' heeft volgens Vroon gevolgen voor het kritisch kunnen kijken naar uitgangspunten die samenhangen met het gebruik van tal van tests die uitgaan van het meritocratisch gedachtegoed en waarbij meten en voorspellen centraal staan, gekoppeld aan een correlatieve benadering van de gedragswetenschap. Er wordt volgens Vroon voortgebouwd op dit bestaande gedachtegoed zonder dit kritisch te beschouwen (Vroon, 1980: 172). Dit maakt dat het vak psychologie voor Vroon 'een niet ongevaarlijke wetenschap is' (Vroon, 1980: 172).

### 8.3.3 Intelligentietests en onderzoek

In het onderzoek naar verborgen talent (Van Heek, 1968) zijn duidelijk invloeden zichtbaar van het Amerikaanse onderzoek van Coleman. Van Heek gebruikt in zijn onderzoek naar het al dan niet aanwezig zijn van verborgen talent de psychologische testbatterij: de Nationale Differentiatietest (NDT). Voorbereidend werk voor deze test is uitgevoerd door de in 1955 opgerichte Research Werkgroep Schoolpsychologie. Zij willen een testbatterij voor de overgang tussen lager en voortgezet onderwijs. In samenwerking met het Nederlands Instituut voor Praeventieve Geneeskunde wordt een onderzoek ingesteld om een testbatterij samen te stellen uit bestaande testbatterijen (Haas, 1995: 185). Dit wordt de Nationale Differentiatietest waarmee toekomstige prestaties in het onderwijs voorspeld moeten kunnen worden. Wereldwijd, en ook in Nederland, ontstaat het besef dat intelligentie alleen niet een goede voorspeller kan zijn van schoolsucces. Intelligentie, gemeten met een intelligentietest, is geen stabiel, aangeboren gegeven. Milieu en gevolgd onderwijs beïnvloeden intelligentiescores: 'De kritiek op het concept intelligentie en de intelligentietest [...] bereikte aan het einde van de jaren zestig en in het daarop volgende decennium een hoogtepunt.' (Haas, 1995: 186). Er wordt gezocht naar andere vormen van testen, vooral in aanvulling op intelligentietests. Van Heek maakt onderscheid in twee vormen van schoolgeschiktheid die met deze discussie samenhangen: parate en potentiële schoolgeschiktheid. 'De parate schoolgeschiktheid is de mate van schoolgeschiktheid, die op het moment van onderzoek bij de onderzochte scholier aanwezig is. Zij wordt uitgedrukt in het percentage kans, dat de desbetreffende scholier volgens de onderzoeksresultaten heeft, de volgende fase van het onderwijs te doorlopen.' (Van Heek, 1968: 34). 'Potentiële schoolgeschiktheid is een theoretische gedachtenconstructie, die een schatting van de kans op schoolsucces geeft [...]. Potentiële

schoolgeschiktheid is zeker niet gelijk aan intelligentie. Immers, in de potentiële schoolgeschiktheid zijn tal van elementen opgenomen, gelijk gemotiveerdheid en doorzettingsvermogen, die bij een intelligentietest veel minder tot uitdrukking komen.' (Van Heek, 1968: 34). Hieruit blijkt volgens Van Heek het besef dat een intelligentietest alleen niet meer acceptabel is voor het bepalen van schoolgeschiktheid. Van Heek heeft in zijn onderzoek de mogelijke intellectuele reserve van leerlingen 'vastgesteld met behulp van een testbatterij (waarin zowel intelligentie- als schoolvorderingentests zijn opgenomen), de schoolcijfers en het oordeel van de onderwijzer over de leerling. De testbatterij is een verzameling tests die in het project Nationale Differentiatie Test geijkt is. In dit ijkingsonderzoek is de bepaling van de v.h.m.o.-geschiktheid gebaseerd op een analyse van de test- en schoolvariabelen, in samenhang met de schoolvariabelen, in samenhang met de schoolresultaten in de 1e klasse van het v.h.m.o.' (Van Heek, 1968: 31).

Sinds de jaren zeventig kent Nederland centrale eindexamens die van belang worden geacht niet alleen voor leerlingen maar vooral om van de kant van de overheid de opbrengsten van onderwijs en de waarde van diploma's te kunnen bewaken. Een vergelijkbare ontwikkeling is te zien bij de eindtoets. De verwevenheid die is ontstaan met onderzoek gericht op het vaststellen van de opbrengsten van onderwijs heeft de waarde van tests en de uitkomsten ervan verder laten groeien. Het belang van testen, bijna ongeacht het soort tests, is in de jaren tachtig en negentig dan ook verder toegenomen. Met het terugtreden van de overheid is ruimte gekomen voor een meer instrumentele benadering van de rol van de overheid richting publieke instituties zoals onderwijsinstellingen. De aandacht wordt daarbij verlegd naar 'output-metingen' als instrumenten voor verantwoording. De waarde van tests wordt enerzijds bewezen door het gebruik van tests zelf en het vertrouwen in de uitkomsten ervan voor betrokkenen, en wordt versterkt door het feit dat tests apart van het onderwijs gepositioneerd zijn bij het Cito. Dit heeft bijgedragen aan het belang en de wetenschappelijke waarde van tests.

## 8.4 Testwetenschap

De testwetenschap heeft in Nederland de ruimte gehad de afgelopen vijftig jaar uit te groeien tot een belangrijke factor in onderwijs en onderwijsonderzoek. Relevant daarvoor is de oprichting van een nationaal instituut voor toetsontwikkeling in de vorm van het Cito. In de testwetenschap wordt onderscheid gemaakt tussen de klassieke testtheorie enerzijds en nieuwe, latere ontwikkelingen anderzijds. Lord en Novick beschrijven de stand van de klassieke psychometrie in 1968 in hun standaardwerk 'Statistical Theories of Mental Test Scores'. Nog altijd is de testwetenschap voor een belangrijk deel gebaseerd op deze stand van zaken van de klassieke testtheorie. Ook in Nederland wordt over toetsen en examens nog altijd geschreven en gerapporteerd in termen die aan deze 'klassieke' testtheorie zijn ontleend. Deze ontwikkeling roept de vraag op wat de waarde en kwaliteit is van tests zoals de eindtoets naar maatstaven uit de testpsychologie zelf.

### 8.4.1 Betrouwbaarheid

Klassieke testtheorie werkt op basis van de aanname dat testcores zijn samengesteld uit een 'ware' score plus een fout, en gebruikt observaties om die ware scores te schatten of te bepalen binnen welk bereik ze liggen met welke waarschijnlijkheid. Deze manier van denken past niet bij de sociale wetenschappen waar gebruik gemaakt wordt van tests, vragenlijsten, toetsen en examens die het karakter hebben van steekproeven. Dat zijn geen metingen zoals in de natuurwetenschappen. In de testpsychologie spreekt men ook over meetfouten, waarmee dan is bedoeld

dat er een verschil is tussen testcores en 'ware' scores (Lord & Novick, 1968). De minder dan perfecte betrouwbaarheid die met steekproeven in de gedragswetenschappen te bereiken is, wordt beschouwd als het equivalent van de meetfout in de natuurwetenschappen. Op basis van een steekproef van opgaven worden leerlingen vergeleken met elkaar en dat zorgt voor variatie omdat zowel de steekproef als de doelgroep niet jaarlijks constant is. Bij tests in het onderwijs draait het daarbij wel veelal om beslissingen op het niveau van een individu en vaak zijn de uitkomsten van belang voor een individu. Zoals het geval is bij toelating, plaatsing of selectie waarbij de score op een test wordt gezien als een indicator met een voorspellende waarde. En hoe meer er op het spel staat, hoe groter de kans is dat een individu zich voorbereidt op een test.

Het is de bedoeling dat eindtoetsen, die van jaar tot jaar verschillen, wel van jaar tot jaar gelijkwaardig zijn, evenals dat bij eindexamens het geval is. De eindtoets van 2019 zou de leerlingen zo goed mogelijk op dezelfde manier moeten rangordenen als de eindtoets van 2018 gedaan zou hebben, zouden ze die eindtoets ook gemaakt hebben. Het vergelijken van de ene rangorde met de andere rangorde is wat in de psychometrie het bepalen van de betrouwbaarheid heet. Het hangt van de specifieke situatie af wat volgens testwetenschappers een passende benadering is voor het bepalen van de betrouwbaarheid van test. Een handige methode is om een test in tweeën te delen, en de scores op de ene helft te vergelijken met die op de andere helft: in hoeverre rangordenen de beide testhelften de leerlingen gelijk? Voor onderzoek naar wat eindtoetsen eigenlijk meten, is die testpsychologie ongetwijfeld relevant. Voor de onderwijspraktijk, en voor leerlingen en hun ouders, gaat het om de testcores zelf, niet om de schattingen van onderliggende 'ware' scores. Overigens, in het geval van rangordes, zoals bij eindtoetsen het geval is, zal de rangorde op basis van geschatte 'ware' scores dezelfde zijn als die op basis van de testcores.

Bij de eindtoets is er formeel geen sprake van zakken of slagen. De eindtoets kan ook niet herkanst worden. Daar gaat een suggestie vanuit dat de eindtoets geen test is waar belangrijke consequenties mee samenhangen. Maar de toets heeft wel degelijk consequenties. Het is nogal van belang in welk hokje een individuele leerling geplaatst wordt. Uit de frequentietabel van de standaardscores is te zien dat er vier cesuurpunten liggen in de bovenste 25 punten van de standaardscoreschaal die loopt van 501-550 (CvTE, Terugblik 2019, tabel 4: frequentietabel). En juist rond de cesuren is volgens de klassieke testtheorie sprake van grote onzekerheid omdat de cesuur het punt is waar het verschil tussen hen die zakken of hen die slagen juist het kleinst is. Dit stelt psychometrici voor de nodige uitdagingen. De gerapporteerde betrouwbaarheid van de eindtoets zegt weinig tot niets over wat er rond de vele cesuren op die eindtoets gebeurt: daar regeert het toeval in veel sterkere mate dan de overall-betrouwbaarheid van de toets lijkt te suggereren. In het geval van de eindtoets is daarbij ook nog sprake van een serie van cesuurpunten omdat de verschillende schooladviezen de feitelijke cesuurpunten vormen binnen de eindtoets. Dit is niet een gevolg van de keuzes van de eindtoets zelf maar van de mate van differentiatie van het stelsel. Die gewenste differentiatie stelt zelfs een perfecte toets of test voor een schier onmogelijke opgave.

De overall-betrouwbaarheid die in de verantwoordingen van de eindtoets gepresenteerd wordt, en de tevredenheid die daarbij wordt uitgesproken als de betrouwbaarheid hoger is dan .90 wekt de suggestie dat de overall-betrouwbaarheid een belangrijke indicator is van de waarde van de eindtoets. Op de waarde van betrouwbaarheid zoals door de klassieke testtheorie wordt vast-



gesteld, is in de afgelopen jaren door psychometrici de nodige nuance aangebracht. Ook Cronbach, wiens maat voor betrouwbaarheid nog altijd gebruikt wordt, is in de loop van zijn leven een grotere waarde aan de theoretische en empirische validatie van test scores gaan toekennen, omdat de test score als afspiegeling van de beoogde 'ware' score nauwelijks te onderbouwen is (Kupermintz, 2003: 295).

#### 8.4.2 Equivaleren en standaardiseren van scores

De scores op de verschillende subtests van de eindtoets worden bij elkaar opgeteld tot een gecombineerde score die wordt omgezet naar een standaard score op een schaal van 501-550. Om zo goed mogelijk te kunnen ordenen, is een spreiding in scores gewenst. De Groot en Van Naersen stellen dat de te beogen (gemiddelde) moeilijkheidsgraad van toetsvragen nauw samenhangt met het doel van de toets: 'Gaaf het om de vraag of de (minimum)doelstellingen met de hele groep bereikt zijn, dan hoopt men op een p-waarde dicht bij 1; gaat het om de differentiatie tussen proefpersonen/leerlingen, dan is zo'n hoge p-waarde ongewenst.' (De Groot & Van Naerssen, 1969: 44). Vanaf het begin van de ontwikkeling van de eindtoetsen tot heden is een hoge gemiddelde p-waarde niet het streven bij de (centrale) eindtoets, waaruit ook statistisch blijkt dat het kunnen differentiëren het primaire doel is van de eindtoets.

Vanaf 1977 is Cito overgegaan op het werken met standaard scores. Cito heeft gekozen voor een 'schaal met een gemiddelde van 535 en een standaarddeviatie van 10. De range van de schaal werd bepaald op 501-550. Deze wat ongewone schaal werd gekozen om iedere verwarring met schoolcijfers, intelligentiequotiënten en dergelijke te voorkomen.' (Wijnstra, 1984:13). Deze werkwijze is ook nu nog de aanpak: 'De totaalscore wordt via lineaire transformatie omgezet in een standaard score. Dit houdt in dat de ruwe score wordt vermenigvuldigd met een getal (A) en dat er daarna een vast getal (B) bij wordt opgeteld. De waarden van A en B worden jaarlijks bepaald. De resulterende standaard score heeft het karakteristieke gemiddelde van omstreeks 535. De verdeling heeft de linksscheve vorm.' (CvTE, 2015:36). Het streven naar een curve met de karakteristieken van een normaalverdeling past bij de functie van de toets om te kunnen differentiëren. Het werken met standaard scores maakt het mogelijk om scores verkregen met toetsen die jaarlijks kunnen verschillen in moeilijkheidsgraad, volgens eenzelfde schaal te rapporteren. Cito voert daartoe een vorm van equivaleren in die het mogelijk maakt dat 'de scores op toetsversies uit verschillende jaren op eenzelfde schaal gebracht kunnen worden' (Wijnstra, 1984:13). De equivaleringsprocedures zorgen ervoor dat de 'gemiddelde score en variantie daarin van jaar tot jaar vrijwel exact vergelijkbaar blijven' (Van Boxtel et al., 2011: 67). Het equivaleren van de scores gebeurt op basis van een lineaire transformatie. Dit is volgens Cito psychometrisch verantwoord onder de volgende voorwaarde: 'We gaan uit van de veronderstelling dat er in de loop van een gering aantal jaren geen dramatische veranderingen plaatsvinden in het landelijke algemene niveau van de leervorderingen. [...] Aannemende dat het gemiddeld niveau in een groep constant is gebleven kunnen de scores van twee eindtoetsen lineair met elkaar worden geequivalereerd.' (Van Boxtel et al., 2011:18). Het is de vraag of deze aanname terecht is. De afgelopen vijftig jaar is de populatie van het basisonderwijs veranderd en dit zou ook zijn weerslag kunnen hebben voor het niveau. De komst van leerlingen met een wisselende migratieachtergrond bijvoorbeeld geeft de bestaande populatie steeds een ander karakter maar ook de autochtone kinderen van nu zijn anders dan die van de jaren tachtig of negentig. Die populatieveranderingen hebben mogelijk gevolgen voor de veronderstelde stabiliteit van scores en scoreverdelingen op taal- en

rekentoetsen. Ditzelfde geldt overigens ook voor veranderingen in didactiek, leermiddelen en het lerarencorps. Met de komst van andere toetsaanbieders en andere eindtoetsen is het daarbij nu ook de vraag of en in hoeverre de groep die gebruik maakt van de Centrale Eindtoets kan worden gezien als 'dezelfde' populatie als daarvoor. Veranderingen in de populatie zetten een proces van equivalenten, zoals tot nu toe gebruikelijk bij de eindtoets, onder druk.

### 8.4.3 Validiteit

In de internationale 'Standards for Educational and Psychological Testing' (2003) wordt gesteld dat het vaststellen van de validiteit van een test het meest belangrijke aspect is voor het bepalen van kwaliteit: 'Validity is the most important consideration in test evaluation.' (Standards 2003). Tegelijkertijd bestaat er geen eenduidigheid over wat er precies onder verstaan moet worden, aan welke criteria een beoordeling ervan zou moeten voldoen en hoe een beoordeling ervan zou moeten worden uitgevoerd (zie ook Hoofdstuk 3). Overeenstemming is er wel over het feit dat een toets die niet valide is, niet geschikt is. Dit maakt de vraag naar de (in)validiteit van een toets weliswaar lastig maar wel relevant om nader te beschouwen. De andere overeenstemming gaat over het feit dat validiteit geen vaste eigenschap of kenmerk is van een meetinstrument, maar dat de vraag naar validiteit gezien moet worden in het perspectief van het gebruiksdoel. In relatie tot de twee beoogde gebruiksdoelen van de eindtoets, selectie en kwaliteit, is de validiteitsvraag tweeledig en zou als volgt geformuleerd kunnen worden:

1. Is er een bruikbaar criterium voor de kwaliteit van de beslissing die met de eindtoets genomen zou moeten worden?
2. Is het op basis van de uitkomsten van deze toets gerechtvaardigd een uitspraak te doen over de kwaliteit van onderwijs?

Uit de diverse verantwoordingen van de eindtoets blijkt dat over validiteit vooral gerapporteerd wordt in termen van correlaties: de uitkomsten van de eindtoets worden gecorreleerd met een tweede gegeven om daarmee de voorspellende waarde, de predictieve validiteit, van de eindtoets aan te tonen. Sinds het begin van de eindtoets wordt het advies van de basisschool 'gebruikt' als extern gegeven waarmee de predictieve validiteit van de eindtoets wordt aangetoond. Met name dat laatste wekt verbazing omdat juist die adviezen, en dan vooral de vermeende subjectiviteit ervan, steeds aanleiding zijn geweest om aan te sturen op een 'objectief' en 'onafhankelijk' oordeel in de vorm van toetsscores ontleend aan gestandaardiseerde metingen. De predictieve validiteit blijkt vanaf het begin van de eindtoets te worden onderbouwd met het gegeven waar de toets een tegenwicht aan zegt te willen bieden, namelijk het oordeel van de leerkracht of school. En uit alle onderzoek uit de afgelopen vijftig jaar blijkt het advies van de leerkracht de plaats in het voortgezet onderwijs na één of meer jaar het beste te 'voorspellen' (Inspectie van het Onderwijs, 2007, Oberon, 2013). In de verantwoording van 2011 meldt Cito dat 'in ongeveer 70% van de gevallen er sprake is van overeenkomst tussen het doorstroomadvies [van de leerkracht/school] en het advies dat wordt uitgebracht op basis van de standaardscore op de eindtoets. Dit betekent dat bij een kwart van de leerlingen de uitkomst van de toets en het advies van de school niet in lijn zijn met elkaar.' (Van Boxtel et al., 2011: 81). Sinds het begin van deze eeuw wordt met enige regelmaat, in samenwerking met het CBS, de relatie onderzocht tussen de scores op de eindtoets, de adviezen van de scholen en de posities die leerlingen vervolgens innemen in het vervolgonderwijs. Dit heet het 'Toelatings- en Doorstroomonderzoek' (TDO). Bij

de overname van de eindtoets door de overheid wordt ervoor gekozen om op een vergelijkbare manier de predictieve validiteit op basis van dit TDO te onderbouwen.

Het is echter de vraag of deze manier van valideren antwoord kan geven op de validiteitsvraag zoals hierboven geformuleerd. Daarbij is een aantal problemen in het geding.

Het voorspellen van toekomstig gedrag en daaraan gekoppeld succes bij toekomstig functioneren is een concept dat aan het begin van de vorige eeuw veel aantrekkingskracht had, op tal van manieren verkend is en weinig succes heeft gekend. Een belangrijk bezwaar én verklaring is volgens Hanson dat het testresultaat vooraf gaat aan het toekomstig gedrag. En omdat niet bekend is hoe iemand in de toekomst zal functioneren, is het lastig zo niet onmogelijk om de test een goede afspiegeling te kunnen laten zijn van wat we te weten willen komen. Er is geen afspiegeling voorhanden: 'Potential is prior to performance.' (Hanson, 1993: 287). Het bepalen van geschiktheid is volgens Hanson een omkering van zaken. Eerst komt het vaststellen van potenties en dan ontstaat zicht op het daadwerkelijke presteren. Een test over toekomstig handelen kan dus niet meer zijn dan een waarschijnlijkheidsinschatting. Dat maakt dat bij tests op zoek wordt gegaan naar wat gedacht wordt dat een goede afspiegeling is van het toekomstig handelen en dat wordt vorm gegeven in een test. Hanson wijst erop dat die afspiegeling gemakkelijk werkelijkheid kan worden. Zo is een test dan geen afspiegeling van een vermeende sociale werkelijkheid maar creëert een test een werkelijkheid (Hanson, 1993: 287). Deze omgekeerde werkwijze, eerst de afspiegeling en dan het daadwerkelijke functioneren, maakt volgens Hanson de kans klein dat iemand bij goed presteren alsnog faalt en het maakt het vrijwel onmogelijk om te ervaren of iemand die op basis van een test is afgewezen het alsnog wel goed zou hebben gedaan. Hanson spreekt in dit verband over een 'transforming capacity' van tests die zowel vooraf als achteraf uitwerkt op de individuen die de test afleggen (Hanson, 1993: 288).

Het doen van voorspellingen op basis van testcores is niet vrij van bezwaren. Want ter onderbouwing van een eventuele validiteit van een instrument zou eerst de vraag gesteld en eenduidig moeten kunnen worden beantwoord of het überhaupt mogelijk is toekomstig (school)succes te voorspellen. En hoe schoolsucces moet worden gezien. Is schoolsucces te zien als geschiktheid in relatie tot verschillende vormen van onderwijs? Hoe ziet die geschiktheid er dan uit? Is het een eigenschap of kenmerk van een persoon of is het eerder te beschouwen als een resultante van genoten onderwijs en dan eerder te zien als verworven kennis of juist als een vaardigheid? En hoe verhoudt het antwoord op die vraag zich tot een instrument voor selectie? In hoeverre kunnen en mogen correlaties, gebaseerd op wat andere leerlingen in het verleden hebben gedaan, worden ingezet als voorspellers van de eventuele geschiktheid van een ander individu dat de toets later maakt?

Vanuit de wetenschappen wordt correlatieve onderzoek in de loop van de vorige eeuw met de nodige scepsis bekeken omdat een samenhang, al dan niet in de vorm van een correlatie, nog geen reden is aan te nemen dat er sprake is van een causaal verband. In de jaren zestig is de correlatieve techniek volop ingezet onder andere bij de onderbouwing van de validiteit van IQ-tests. Dit soort onderzoeken waarbij een gevonden statistisch verband wordt gezien als een oorzakelijk verband, schaarst Sternberg onder de noemer van 'zichzelf waarmakende voorspellingen' (Sternberg, 2002: 9). Een correlatie tussen een score op een test en de latere positie in

onderwijs moet volgens de Australische psychometricus Wilson met de nodige voorzichtigheid bekeken worden: 'Interpreting the meaning of high correlations can be quite tricky. For example, if the rank order of students on a university entry examination in Physics correlates 0.9 with their first year Physics results at University this could be interpreted as an enormously successful outcome in terms of educational prediction. It is also completely consistent with the implication that no new Physics has been learnt, or that the University course has been completely unsuccessful in compensating for initial inequities in knowledge and opportunity.' (Wilson, 1998: 202).

Het feit dat er een correlatie bestaat tussen twee gegevens heeft in de sociale wetenschappen afgedaan als afdoende verklaring of onderbouwing van een gevolgtrekking. Borsboom, Mellenbergh en Van Heerden stellen dat het vertroebelend en lichtelijk beschamend voor de psychologie is dat nog altijd niet met voldoende zekerheid kan worden vastgesteld wat psychologische constructen zoals persoonlijkheidskenmerken, intelligentie of houdingen nu eigenlijk zijn. Zij stellen dat psychologisch onderzoek zich heeft aangepast aan de beperkingen van gangbare statistische procedures in plaats van nieuwe procedures te bedenken die theorieën op een passend niveau kunnen testen (Borsboom et al., 2003: 215). Het toepassen van uitkomsten van vergelijkingen tussen individuen om daarmee voorspellingen te doen voor een individu zijn volgens hen statistisch niet te onderbouwen: 'The within-subjects causal interpretation may be viewed as a fallacious application of between-subjects results to individuals.' (Borsboom et al., 2003: 216). Zij komen tot de slotsom dat als je wilt weten wat er gebeurt in een persoon, dat je dan de persoon zelf moet bestuderen, op het niveau van het individu. En wat er gebeurt op het niveau van individuele processen kan volgens hen niet met de huidige standaard meetmodellen worden vastgesteld (Borsboom et al., 2003: 217).

De manier waarop de validiteit van de eindtoets wordt onderbouwd, is ondanks die ontwikkelingen nauwelijks onderwerp van discussie geweest. In hoeverre de beheersing van taal (en de daarbij aangebrachte beperking van taal door uitsluitend lezen en taalverzorging te hanteren) en rekenen geschiktheid voor de verschillende vormen van voortgezet onderwijs kan voorspellen, is gebaseerd op aannames en uitkomsten van correlatieel onderzoek. Het is daarmee de vraag of taal en rekenen als selectiecriteria geschikt zijn. Onderzoek naar succes- en faalfactoren in het voortgezet onderwijs wijst vooral op andere aspecten die nu niet in de doorverwijzing worden betrokken zoals de hoogte van het gezinsinkomen of de gezinssamenstelling maar die wel verband houden met schoolsucces (Hakkenes & De Wijs, 2012).

'Selffulfilling prophecy' is een ook in onderwijs werkzaam principe. Dit betekent concreet dat de score niet alleen bepalend is voor het advies dat een kind krijgt maar dat de score doorwerkt in het beeld dat een kind van zichzelf heeft en het beeld dat de omgeving heeft van het kind. Daarmee is een score meer dan een advies: het werkt als een etiket, een stempel en kan zich in het hoofd van een kind vastzetten als een gegeven. Het verplichte gebruik van LVS-toetsen versterkt dit fenomeen; steeds vroeger worden etiketten uitgedeeld en gedurende het basisonderwijs is steeds meer tijd om het beeld hiernaar te vormen. Dit geldt ook voor de 'verharding' van de categorieën. Gaven de adviezen in de vorm van de 'poppetjes-grafieken' nog enige ruimte en nuance, de exactheid waarmee de huidige categorisering is omgeven, biedt die niet. In het geval van de eindtoets gaat het om een test waarbij de uitkomsten vergaande gevolgen hebben voor de toekomst van leerlingen en hun zelfbeeld. Het feit dat er geen onderzoek is gedaan naar

de voorspellende geldigheid is volgens psychometricus Wilbrink 'een ernstige zaak' (Wilbrink, 1997: 364).

De validiteit van de eindtoets staat onder druk. Biesta voegt daar nog een extra element aan toe dat hij de 'normatieve validiteitsvraag' noemt, in aanvulling op de technische validiteiten van metingen. Hij stelt de vraag naar de waarde van de meting in relatie tot wat als waardevol wordt beschouwd voor onderwijs: 'meten we wel wat we willen meten? Of meten we wat we makkelijk kunnen meten en waarden we uiteindelijk wat we (kunnen) meten?' (Biesta, 2012: 25). Voor de eindtoets is ook deze 'normatieve validiteitsvraag' niet gesteld of beantwoord. Samenvattend kan gesteld worden dat een goede beantwoording van de vraag of en in hoeverre de eindtoets een valide instrument is voor de selectie voor het voortgezet onderwijs niet of niet voldoende mogelijk is.

Evaluatie van de kwaliteit van onderwijs op basis van resultaten op de eindtoets is een gangbare vorm van evalueren geworden. Vanaf het begin van deze eeuw neemt de Inspectie van het Onderwijs de uitkomsten op de eindtoets mee in haar beoordeling van de kwaliteit van onderwijs. Met regelmaat wordt 'gesleuteld' aan deze beoordeling omdat de fairheid ervan bevestigd wordt. Dit heeft ertoe geleid dat leerlingkenmerken meegenomen zijn in de toetsresultaten ('leerlinggewicht') en daarmee in de beoordeling van kwaliteit van scholen. Welke kenmerken van leerlingen relevant zijn om te worden 'meegewogen' is voortdurend onderwerp van discussie. Daarnaast wordt met regelmaat gewezen op de beperkte opvatting over kwaliteit van onderwijs als uitsluitend de opbrengsten van onderwijs geëvalueerd worden die met de eindtoets kunnen worden gemeten (Onderwijsraad, 2013, Onderwijsraad, 2016). Het nieuwe onderwijsresultatenmodel dat door de Inspectie van het Onderwijs vanaf het schooljaar 2020-2021 zal worden gehanteerd voor de beoordeling van onderwijskwaliteit, kent een nieuwe manier van 'meewegen' van leerlingkenmerken en neemt de scores op de referentieniveaus als uitgangspunt voor de beoordeling van de kwaliteit van onderwijs (zie ook hoofdstuk 5). De validiteit van deze nieuwe manier van evalueren zal op termijn, als de effecten van de nieuwe manier van waarden duidelijk worden, nader kunnen worden bevestigd en onderzocht. De meer fundamentele vraag naar wat gezien wordt als kwaliteit van onderwijs, wat van waarde is voor het bepalen van de kwaliteit van onderwijs en hoe daarop toe te zien, is vooraf niet gesteld.

#### 8.4.4 Test- en itembias

Eerder is meer specifiek ingegaan op de scores op de eindtoetsen in relatie tot bepaalde sociale groepen. Uit hoofdstuk 7 blijkt dat verschillende allochtone groepen gemiddeld lager scoren op de eindtoets. Dit roept de vraag op of en in hoeverre bepaalde doelgroepen in het nadeel zijn bij de eindtoets; ofwel of er sprake is van 'bias'. Bias betekent<sup>27</sup> 'a particular tendency, trend, inclination, feeling, or opinion, especially one that is preconceived or unreasoned'. Van Yperen, Bleichroth en Van de Vijver hebben onderzocht 'of een testscore behaald door een allochtone Nederlander op dezelfde wijze verwijst naar het domein dat de betreffende test wordt geacht te representeren als een gelijke testscore behaald door een autochtone Nederlander. Met andere woorden, is de betekenis van een bepaalde testscore voor een allochtone Nederlander dezelfde als voor een autochtone Nederlander?' (Van Yperen et al., 2002: 7). Zij stellen dat wanneer er sprake is van verschillen in betekenis de testcores 'partijdig' of 'vertekend' zijn en dan is er sprake van bias. Uiterwijk hanteert in zijn dissertatie 'De bruikbaarheid

27] <https://www.dictionary.com/browse/bias>

van de Eindtoets Basisonderwijs voor allochtone leerlingen' (1994) de definitie van Kok (1988) die als equivalent voor 'bias' het Nederlandse begrip 'partijdigheid' kiest en als bijvoeglijk naamwoord het woord 'partijdig' (Uiterwijk, 1994: 7). In 1990 is er door een werkgroep onder leiding van Hofstee onderzoek gedaan naar toepasbaarheid van psychologische tests bij allochtonen. Testgebruikers hebben volgens de werkgroep de verantwoordelijkheid aannemelijk te maken dat verschillen in testcores tussen allochtonen en autochtonen niet aan partijdigheidseffecten kan worden toegeschreven. Een belangrijke eis hierbij is dat de test niet 'ondervoorspelt', dat wil zeggen dat bij een gelijke testscore de verwachte criteriumwaarde voor allochtonen niet lager mag zijn dan voor autochtonen (Hofstee et al., 1990a: 9). De werkgroep stelt vast op basis van het onderzoek dat de onderzochte psychologische tests niet voldoen: 'De eis van onberispelijkheid brengt met zich mee dat veel tests grondig zullen moeten worden herzien. [...] Qua onderbouwing door middel van research inzake test bias scoren alle tests onvoldoende.' (Hofstee, 1990b: 293). Dit rapport zet onderzoek naar het aantonen van onpartijdigheid van tests op de agenda van testontwikkelaars.

Bij de beoordelingen van tests en toetsen door de COTAN worden analyses als vraagpartijdigheid (item bias; differential item functioning) of toetspartijdigheid van belang geacht om de kwaliteit van een test of toets te kunnen evalueren (Evers e.a., 2010). Het vaststellen van het al dan niet aanwezig zijn van 'bias' in tests stuit op de nodige bezwaren en belemmeringen. Shepard geeft een overzicht van problemen die samenhangen met het bepalen van testbias. Ten eerste is er het probleem of er sprake is van bias in een test of van bias bij het gebruik van tests. Met dit onderscheid hangt ook het gebruik van het woord 'fairness' samen in de context van testen. Een test kan 'unfair' zijn in de toepassing maar bias kan ook betrekking hebben op de test zelf. Daarmee zijn beide termen niet als synoniemen te zien. En als er sprake is van bias in een test kan het dan gaan om bias van de test als geheel of om bias van (een deel van de) items. Over het algemeen, stelt Shepard, wordt bias gezien als een vorm van invaliditeit, wat het voor psychometrici gecompliceerd maakt omdat validiteit niet wordt gezien als een eigenschap van een test en bias wel. Shepard stelt dat het verschil maken tussen bias in een test en unfair gebruik van test fundamenteel is en gevolgen heeft voor de manier waarop onderzoek wordt gedaan naar bias (Shepard, 1982: 10). Uitgangspunt voor bias bij een selectietest is volgens Shepard, zich daarbij baserend op Petersen en Novick, dat 'selectie niet gebiased is (met welke definitie dan ook) als het criterium [dat voorspeld moet worden] niet gebiased is' (Shepard, 1982: 19). Drenth en Sijtsma stellen dat cultuurvrije tests niet bestaan. Zij wijzen erop dat veel ervaring is opgedaan met het testen van cultureel zeer sterk uiteenlopende populaties, wat maakt dat zij stellen dat 'de conclusie gerechtvaardigd lijkt dat een zuivere cultuurvrije test niet bestaat en dat het ook zinloos is naar de constructie van een dergelijke test te streven' omdat 'iedere prestatietest het resultaat meet van de inwerking van de omgeving op datgene wat de persoon aan potentieel heeft meegekregen' (Drenth & Sijtsma, 1990: 56). Zij voegen daaraan toe dat 'cultuurvrije intelligentie een onmeetbaar postulaat blijft en wat aanvankelijk werd aangekondigd als cultuurvrije test bij nadere toetsing voortdurend door de mand valt (Drenth & Sijtsma, 1990: 267). Bügel en Sanders stellen in hun 'Richtlijnen voor de ontwikkeling van onpartijdige toetsen' (1998) dat er sprake is van testbias 'als op grond van een toets als geheel de vaardigheid van een bepaalde groep leerlingen ten onrechte systematisch te hoog of te laag wordt ingeschat. Testbias impliceert dat de test niet dezelfde vaardigheid bij de onderscheiden groepen leerlingen meet. Er is sprake van itembias als verschillen in prestaties op een item of opgave veroorzaakt worden door kenmerken van de

opgave die niet relevant zijn voor wat de opgave beoogt te meten. Itembias impliceert dat de opgave niet dezelfde vaardigheid bij de onderscheiden groepen leerlingen meet' (Bügel & Sanders, 1998: 3). In het geval van onderzoek naar itembias gaat men volgens Bügel en Sanders voor elk item na of de gemiddelde itemscore van de even vaardige subgroepen allochtone en autochtone leerlingen al of niet gelijk is: 'differential item functioning' (DIF) (Bügel & Sanders, 1998: 3).

Uiterwijk stelt dat 'het ontbreken van een onpartijdig extern criterium het in feite onmogelijk maakt om te beoordelen of er bij de Eindtoets Basisonderwijs al dan niet sprake is van toetsbias'. Hij formuleert zijn onderzoeksvraag dan ook als 'het nagaan van de predictieve validiteit van de Eindtoets Basisonderwijs voor onderscheiden etnische groepen in vergelijking met die van het advies van de basisschool' (Uiterwijk, 1994: 13). Uiterwijk geeft daarbij aan dat het advies van de basisschool niet kan worden gezien als een 'onpartijdig extern criterium' maar legitimeert zijn keuze omdat 'toetsen en het advies basisschool een functie vervullen bij de schoolkeuze en de toelating tot het voortgezet onderwijs. Daardoor functioneert in de praktijk het bereikte onderwijsniveau wel degelijk als maat voor schoolsucces. Zo zeggen we bijvoorbeeld dat het advies van de basisschool goed is geweest, wanneer een leerling met een VWO-advies zonder doubleren in de derde klas VWO terecht komt. We zeggen echter ook dat de toetsuitslag onjuist was, wanneer een leerling met een score net onder het gemiddelde zonder doubleren eveneens in de derde klas VWO terecht komt.' (Uiterwijk, 1994: 13). Uiterwijk relateert de waarde hiervan door toe te voegen dat het 'uiteraard mogelijk is dat het "zichzelf waarmakend karakter" van het advies van de basisschool andere effecten heeft op de schoolloopbaan dan dat van de toetsuitslag, waardoor het moeilijk is om over juiste en onjuiste adviezen en scores te spreken' (Uiterwijk, 1994: 13).

Ook Shepard geeft aan dat er geen 'foolproof statistical bias detection method' is. Niet voor tests en niet voor testitems. Zij stelt dat technieken om itembias aan te tonen zelf gevalideerd zouden dienen te worden (Shepard, 1982: 22). Met behulp van statistische technieken kunnen anomalieën aan het licht komen maar die vragen vervolgens om expertbeoordeling om nader geïnterpreteerd te worden. Shepard stelt dan ook dat de enige consensus die er is, is, dat statistiek en menselijke beoordeling in combinatie nodig zijn om itembias te kunnen bepalen (Shepard, 1982: 26). Test- en itembias moet volgens Shepard worden geconceptualiseerd als invaliditeit, iets dat de betekenis van testresultaten voor bepaalde groepen verstoort en nader onderzoek vereist in de volle breedte van wat onder bias verstaan kan worden, en naar wat gedaan moet worden voordat het gebruik van een test als fair gekarakteriseerd kan worden (Shepard, 1982: 26).

Het door Uiterwijk uitgevoerde onderzoek, zowel op basis van psychometrische analyses ('detectiefase') als expert-oordelen ('verklaringsfase'), is ook volgens Uiterwijk met veel onzekerheden omgeven. Zo komt uit de detectiefase op basis van de klassieke testtheorie (Mantel-Haenszel-analyses) een veel groter aandeel itembias naar voren dan met behulp van de detectiefase op basis van de itemresponstheorie (IRT): in het eerste geval blijkt voor Turkse en voor Marokkaanse leerlingen zowel in 1987 als in 1989 ongeveer de helft van de 60 taalitems partijdig te zijn (50%), een derde tot een vijfde deel van de 60 rekenitems (33% tot 20%) en een derde deel van de 60 informatieverwerking-items (33%). Bij de analyses met de IRT-procedure blijkt dat het aantal partijdige items voor Turkse en/of Marokkaanse leerlingen beduidend geringer is: 20 van de in totaal 360 geanalyseerde items (6%). Bij een combinatie van beide methodes blijken er in totaal 13 items partijdig te zijn (4%) (Uiterwijk, 1994: 187-8). De inhoudelijke analyse levert volgens

Uiterwijk twee fundamentele problemen op. Enerzijds blijkt het volgens Uiterwijk moeilijk te zijn om met zekerheid aan te geven welk itemelement de biasbron is en anderzijds blijkt dat bij sterk vergelijkbare items de ene keer wel sprake is van itembias en de andere keer niet. Schilt-Mol heeft in een later onderzoek de items uit de eindtoets basisonderwijs nader onderzocht met behulp van 'differential item functioning' (DIF) op een vergelijkbare manier als Uiterwijk in 1994 heeft gedaan. Zij spreekt van itembias als het verschillend functioneren van (groepen van) leerlingen bij een toetsopgave veroorzaakt wordt door construct-irrelevante factoren. De bron van DIF behoort niet tot het kennisdomein dat de toets beoogt te meten, en additionele kennis en vaardigheden anders dan die beoogd worden te meten, spelen een rol bij het correct beantwoorden van het item (Schilt-Mol, 2007: 11). Daar zit volgens Mol de kern van het verschil tussen DIF en itembias: er kan sprake zijn van DIF (op basis van psychometrische data) maar van itembias is alleen sprake als er van construct-irrelevante factoren sprake is. Zij sluit daarmee aan bij de definitie die de 'Standards for Educational and psychological testing' (1999) en de 'Fairness review guides' van het Educational Testing Service (ETS, 2003) formuleren over het testen van construct-irrelevante factoren en waarmee itembias geplaagd wordt in het licht van onderzoek naar de constructvaliditeit. Schilt-Mol heeft aan het gebruikelijke onderzoek met een eerste en tweede fase een aantal fasen toegevoegd: zo zijn de items gemanipuleerd zodat de vermoedelijke bron van DIF is weggenomen, zijn de items opgenomen in nieuwe toetsen waaraan ook niet eerder getoetste items zijn toegevoegd, waarbij een potentiële DIF-bron wordt verondersteld en die zijn vervolgens voorgelegd aan groepen Marokkaanse en Turkse leerlingen, om zo de DIF opnieuw te kunnen onderzoeken en te kunnen bepalen of er daadwerkelijk sprake is van itembias. Schilt-Mol stelt in haar onderzoek dat er voor een deel van de DIF-items wel degelijk ook itembias kan worden vastgesteld. In de laatste verantwoording die Cito zelf heeft opgesteld over de eindtoetsafname van 2010 stelt Cito dat slechts voor 12 van de 200 items er sprake is van een DIF-item en dat het effect op de totale toetsscore minimaal is (Van Boxtel et al. 2011: 80).

Uit de gekozen aanpak blijkt dat voor het vaststellen van itembias psychometrisch onderzoek aan het inhoudelijk onderzoek voorafgaat. Voor het bepalen van DIF wordt rekening gehouden met de vaardigheid van leerlingen. Er is sprake van toetsbias als leerlingen die overall een gelijke vaardigheid hebben zich in een test 'anders' gedragen, waarbij het verschil in score kan worden toegeschreven aan groepslidmaatschap en niet aan verschil in vaardigheid. Dit betekent dat voor onderzoek naar testbias een criterium beschikbaar moet zijn op basis waarvan de vaardigheid van een leerling kan worden vastgesteld. Bügel en Sanders stellen dat dit alleen kan onder de conditie dat het criterium vrij van bias is (Bügel & Sanders, 1998: 3). Dit maakt onderzoek naar testbias al snel lastig want wat kan gelden als bias-vrij criterium waarmee vergeleken kan worden of er daadwerkelijk sprake is van verschil in vaardigheid tussen groepen? Bij de Mantel-Haenszel-procedure doet volgens Glas en Ouborg (1993)<sup>28</sup> de aanwezigheid van items met DIF afbreuk aan de waarde van de somscore als indicator van de overall-vaardigheid van de leerlingen, omdat de somscore mede bepaald wordt door items die voor de twee groepen een verschillende moeilijkheidsgraad hebben (Glas & Ouborg, 1993: 355). Met een iteratief proces kan hier tegenwicht aan worden geboden door te zoeken naar een somscore waarin geen DIF-items voorkomen. Maar dan blijft dat de somscore niet gevalideerd is als

---

28] <https://www.cito.nl/-/media/files/kennisbank/psychometrie/psychometrie-in-de-praktijk/cito-pidp-h9-vraagonzuiverheid-glas-ouborg-1993.pdf?la=nl-NL&hash=92F906C01CA3D4D1B008772943AA0ABDB750F856>



maat voor de vaardigheid van respondenten. Ook bij IRT-modellen is het lastig om testbias vast te stellen met behulp van een criterium dat vrij van bias is. Hier blijft de kritiek volgens Glas en Ouborg, dat de gewogen somscore als maat voor het vaardigheidsniveau eerst zou moeten worden gevalideerd (Glas & Ouborg, 1993: 364). Daar komt bij dat parameterschatting zoals gebruikelijk bij IRT volgens Glas en Ouborg leidt tot restricties op de toetsingsgrootheden, waardoor een item met DIF ten nadele van de ene groep kan resulteren in een of meer items die schijnbaar DIF vertonen ten nadele van de andere groep (Glas & Ouborg, 1993: 364). Door parameterschatting en andere oorzaken kan volgens Glas en Ouborg de informatie die toetsen opleveren enigszins vertroebelen (Glas & Ouborg, 1993: 364). Een combinatie van de Mantel-Haenszel-procedure in combinatie met IRT vinden zij aan te bevelen (Glas & Ouborg, 1993: 371).

Bügel en Sanders wijzen erop dat wanneer 'talige' contexten worden gebruikt, er een grotere kans op bias bestaat. Niet alleen vanwege de verschillen in kennis en interesse van de leerlingen ten aanzien van die contexten, maar ook door verschillen in taalkennis en leesvaardigheid van de kandidaten. In dat geval kan bias optreden, wanneer taalkennis of leesvaardigheid niet het onderwerp van de toets is. Zij adviseren daarom om bij het gebruik van talige contexten zeer kritisch te zijn als taal niet het object van toetsing is (Bügel & Sanders, 1998: 7). Hickendorff en Janssen hebben de invloed van contexten in rekenopgaven op de prestaties van basisschoolleerlingen onderzocht. Zij vonden dat leerlingen die thuis geen Nederlands spreken een stuk slechter scoren op rekenopgaven in een context dan op kale opgaven wat het beeld van eventuele achterstanden kan vertekenen: 'Als we rekenvaardigheid slechts vaststellen met een toets die voor het grootste deel uit contextopgaven bestaat, wordt de achterstand van leerlingen met een niet-Nederlandse thuistaal overschat.' (Hickendorff & Janssen, 2009: 10).

Uit de scoreoverzichten van de Centrale Eindtoetsen naar migratieachtergrond, wordt duidelijk dat de gemiddelde standardscore verschilt gegeven de achtergrond van leerlingen (zie ook hoofdstuk 7). Het beeld dat oprijst uit de onderzoeken naar itembias wijst op een beperkte systematische invloed van items gegeven de groepen van herkomst van leerlingen. Maar de manier waarop test- en itembias wordt vastgesteld, kan voor een deel verantwoordelijk worden geacht voor het niet kunnen vaststellen van test- en itembias in deze onderzoeken: het uitgaan van vergelijkingen van scores van leerlingen met *gelijke* vaardigheidsscores onttrekt mogelijk een substantieel deel van mogelijke itembias aan het zicht. Bovendien beïnvloedt de manier waarop de groepen worden geordend mogelijk ook bias. De ordening op herkomstlanden is een keuze waarbij kinderen van bijvoorbeeld hoger en lager opgeleide ouders samen onder één noemer worden gebracht. De vraag is of ook niet andere ordeningen van groepen leerlingen op bias moeten worden onderzocht. Gezien de toenemende aandacht voor de kloof die lijkt te ontstaan tussen lager en hoger opgeleiden in het kader van de meritocratisering van de samenleving, is het wellicht relevant onderzoek te doen naar test- en itembias volgens de scheidlijnen van het opleidingsniveau van ouders.

De waarde van geconstateerde itembias in relatie tot eventuele testbias is op basis van bestaand onderzoek niet te achterhalen. Los van de onzuiverheden die samenhangen met procedures om onzuiverheden in items en daarmee in tests op te sporen, hangen de gevolgen van itembias in hoge mate samen met de implicaties die verschillen in scores hebben voor de verkregen schooladviezen. Het is niet te achterhalen hoeveel scorepunten verschil er is tussen de ene standards-

core en de andere. Daarmee is onduidelijk hoeveel impact het aandeel opgaven met itembias heeft op de scorerresultaten van een individuele leerling met een bepaalde migratieachtergrond en daarmee hoeveel effect het aandeel items met itembias heeft op de toekomst van toetsdeelnemers. Hoe belangrijker de toets is in zijn gevolgen voor de toekomst van een leerling, zoals in het geval van de eindtoets, hoe relevanter de vraag naar de effecten van itembias in relatie tot toetsbias en hoe relevanter de manier waarop dit onderzoek gebeurt.

De focus op de psychometrische onderbouwing van itembias van de afgelopen decennia gaat voorbij aan de bredere opdracht die de werkgroep onder leiding van Hofstee in 1990 in haar rapport meegaf aan testontwikkelaars maar vooral ook aan hen die tests willen inzetten bij allochtone doelgroepen. Hofstee bepleit nadrukkelijk ruimte voor wat hij noemt het 'professiebeginsel' vanuit het standpunt dat 'omgaan met mensen mensenwerk is en niet aan procedures kan worden overgelaten' (Hofstee, 1990b: 293). Zelfs wanneer tests aan de hoogste eisen voldoen, stelt Hofstee, 'zullen zich in individuele gevallen altijd bijzondere situaties blijven voordoen die klakkeloze toepassing van regels en procedures absurd maken' (Hofstee, 1990b; 293). Hofstee benadrukt dat testafname en testinterpretatie dienen te zijn ingebed in een proces van professionele oordeelsvorming: 'De noodzaak van verbijzondering geldt in principe bij ieder individu. In die zin is het dubieuze gehalte van psychologische tests in verband met allochtonen een geluk bij een ongeluk. Het bevat een aansporing ernst te maken met het elementair beginsel dat mensen recht hebben op een individuele benadering in aangelegenheden die voor hen van groot belang zijn.' (Hofstee, 1990b; 293).

#### 8.4.5 Belang van testen

Bij tests en de uitkomsten ervan zijn de belangen niet altijd even groot. Tests verschillen in de mate van 'onschuld' die ze hebben. Het maken van onderscheid tussen tests in hun mate van 'onschuld' is een belangrijk aspect in de discussie over de rol van tests. Testen in de betekenis van 'checken' of er overdracht van kennis heeft plaatsgevonden kan op allerlei niveaus worden gerealiseerd. Voor een docent is testen een integraal onderdeel van de dagelijkse praktijk van lesgeven en kan een test vanuit een bepaald perspectief worden gezien als het start- en eindpunt van leren: pas als onderwijs aansluit bij het kennisniveau van een leerling kan er sprake zijn van overdracht van lesstof. En pas als er sprake is van beheersing, kan verder gegaan worden met leren. Maar in de praktijk blijken tests ook anders te worden gezien en ingezet. Daarbij is het aan de test zelf niet te zien hoe de test wordt gebruikt, terwijl het gebruik bepaalt welke functie een test heeft. Voor het voeren van discussies over de waarde van tests is het van belang te onderscheiden welke rol een test vervult in de praktijk en daarbij niet alleen te kijken naar het instrument op zich en wat men met de test beoogt, maar vooral ook naar het belang dat gekoppeld is aan de uitkomst van een toets, op de korte én op de langere termijn (zie ook Hoofdstuk 3). De waarde van tests hangt namelijk samen met wat er met de uitkomsten wordt gedaan en door wie, en met wat er op het spel staat voor wie, kortom met de vraag hoe 'high stakes' een test is.

Cambridge dictionary voor Business English hanteert de volgende definitie van high stakes: 'High stakes is used to describe a situation that has a lot of risk and in which someone is likely to either get or lose an advantage, a lot of money, etc.' Een element van deze definitie is dat high stakesness gerelateerd is aan het verlies van geld of het lopen van grote risico's. Voor de context van onderwijs en toetsen spreekt Cambridge dictionary van een high stakes test als de test 'very

important is for the person who takes it'. In deze definitie beperkt het high stakes karakter zich tot het belang van de test voor degene die de test aflegt. Education Reform geeft een bredere definitie: 'Any test used to make important decisions about students, educators, schools, or districts, most commonly for the purpose of accountability'. Au hanteert de definitie van Madaus uit 1988 die stelt dat een test high stakes is als de uitkomsten gebruikt worden voor het nemen van belangrijke beslissingen die betrekking hebben op studenten, docenten, bestuurders, scholen en overheden (Au, 2007: 258). Nichols en Berliner spreken van high-stakes tests als het gaat om tests waar serieuze consequenties aan verbonden zijn (Nichols & Berliner, 2007: XV). Zij plaatsen high stakes tests nadrukkelijk in het politieke domein door over tests te spreken als een 'policy mechanism' (Nichols & Berliner, 2007: 6). In hun optiek hebben dergelijke tests vaak een dubbele functie: enerzijds worden ze ingezet om veranderingen in het onderwijs in gang te zetten, de functie van 'effector', en tegelijkertijd worden ze gebruikt om vast te kunnen stellen of veranderingen zich ook daadwerkelijk voordoen, de functie van 'detector' (Nichols & Berliner, 2007: 6). Nichols en Berliner zien dat met name bij high stakes tests de validiteit onder druk komt te staan. De inzet van high stakes tests om de algehele kwaliteit van onderwijs te verbeteren blijkt bijvoorbeeld als effect te hebben dat het curriculum zich beperkt tot datgene dat in de test aan bod komt (Nichols & Berliner, 2007: 132). Ook is er dan effect op de onderwijstijd, die substantieel ingevuld zal worden met het trainen voor een test, wat ten koste gaat van leerlingen en docenten en tegelijkertijd ervoor zorgt dat de indicator waar de test oorspronkelijk informatie over zou geven, geen nuttige informatie meer kan verschaffen: het is niet meer zichtbaar of de algehele onderwijskwaliteit vooruitgaat of alleen de testresultaten (Nichols & Berliner, 2007: 123). Zij concluderen dat zodra het belang van een test toeneemt, er veel inspanning en middelen zullen worden besteed aan het goed scoren op die test, wat niet direct leidt tot grondiger leren of breder leren: 'Effort is not necessarily directed to learning at deeper levels nor at learning unrelated but interesting and useful content.' (Nichols & Berliner, 2007: 141). Jacobs stelt dat hoe meer er op het spel staat voor deelnemers aan een test, hoe relevanter het wordt de fairheid van wat er op het spel staat mee te wegen in de beoordeling van de validiteit van de uitkomsten van een test (zie hoofdstuk 7).

De overgang van de eindtoets van een instrument ter ondersteuning van de leerkracht en de school voor het opstellen van het schooladvies voor voortgezet onderwijs naar een instrument in handen van de overheid waarmee toezicht kan worden gehouden op zowel de individuele adviezen als de opbrengsten van onderwijs, maakt de toets voor alle betrokkenen meer 'high stakes'. Het determinerende karakter van het 'advies' op basis van de toets maakt een nader onderzoek naar de relatie tussen de zekerheid van de toetsuitslag en de fairheid van de uitkomsten voor winnaars en verliezers van belang.

## 8.5 Samengevat

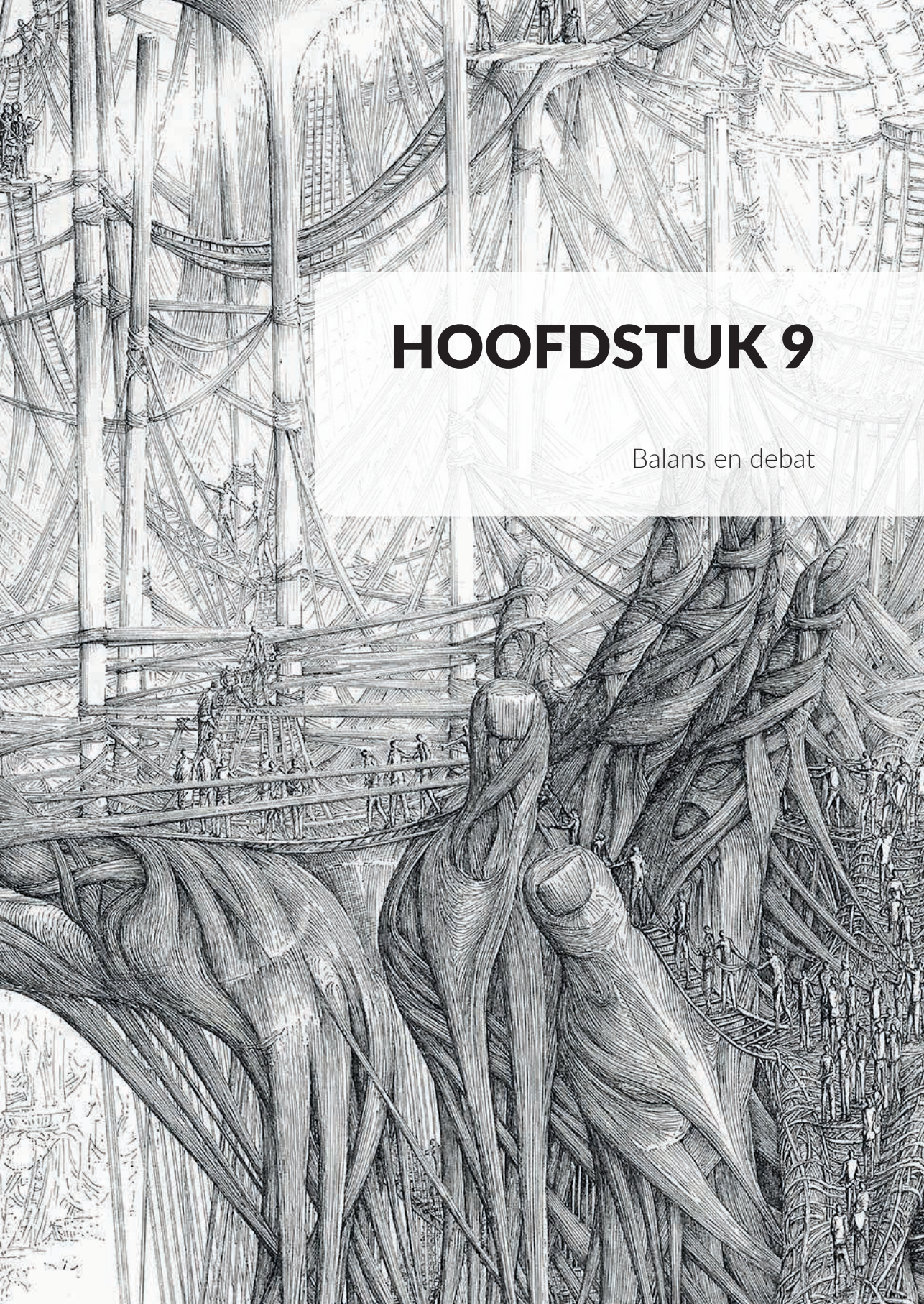
De aanvankelijke focus op de diversiteit en uniciteit van de mens in de menswetenschappen, heeft plaatsgemaakt voor een andere manier van kijken naar mensen: het vergelijken van mensen met elkaar op één bepaald aspect. Dat aspect wordt gezien als een kenmerk en vervolgens zodanig uitvergroot dat typeringen ontstaan: van 'pars pro toto' (deel voor het geheel) naar 'totum pro parte' (geheel voor een deel). Voor het gebruik van intelligentietests in onderzoek kan er waarde liggen in het vergelijken van groepen met een hoge score op een intelligentietest met groepen met een lage score op een intelligentietest in relatie tot een variabele als 'gezondheid' of 'plaats

op de arbeidsmarkt'. Een dergelijk gebruik van scores op intelligentietests verschilt fundamenteel van het gebruik van intelligentiescores om individuele capaciteiten te duiden. De eindtoets is ontwikkeld als een vorm van een intelligentietest met een gecombineerde score op taal en rekenen als inschatter van toekomstig schoolsucces in combinatie met testpsychometrische uitgangspunten zoals gebruikelijk bij intelligentietests. De hoogst scorenden worden nu geselecteerd voor de hoogste vormen van vervolgonderwijs en gezien als de 'slimsten'. De uitkomsten van dergelijke vergelijkingen met behulp van tests zijn zo krachtig geworden dat uitkomsten van metingen synoniem lijken te worden met kenmerken die gezien worden als vaststaande, of zelfs als aangeboren eigenschappen die vergaand classificeren en ordenen legitimeren, daarbij verwijzend naar een gelijke verdeling van kansen die zou worden beoogd. De wetenschap die achter tests schuilgaat, biedt onvoldoende basis om dit te kunnen legitimeren. Er dringt zich een beeld op van de testwetenschap die vooral de belangen dient van hen die met testen maatschappelijke belangen willen dienen en daarbij voorbij gaan aan de belangen van de leerling en waarbij enige empathie voor de individuele leerling en zijn persoonlijke omstandigheden ontbreekt. Een score op een intelligentietest kan niet in individuele gevallen de waarde van een individu voorspellen. Net als Binet hebben gebruikers van een test de professionele en morele verplichting om in individuele gevallen ook te kijken naar mogelijk relevante individuele omstandigheden.

De testwetenschap heeft zichzelf weten te positioneren als een wetenschap die serieus genomen wordt. De institutionalisering ervan met de oprichting van het Cito heeft belangrijke discussies en ontwikkelingen in de wetenschap in het algemeen en in de testwetenschap in het bijzonder in belangrijke mate aan het zicht kunnen onttrekken, onbewust of bewust. Ze hebben niet waarneembaar geresulteerd in aanpassingen van bestaande tests en hun praktische toepassingen. Daarbij is na overname van de eindtoets door de overheid een nieuw element aan de bestaande eindtoets toegevoegd, namelijk het vaststellen en beoordelen van de kwaliteit van onderwijs met behulp van ingevoegde referentieniveaus. Deze dubbele functie van én selectie én beoordelen van kwaliteit van onderwijs vormt een bedreiging voor de validiteit van de eindtoets. Institutionalisering kan ervoor zorgen dat de belangen voor een instituut groot kunnen zijn. Ook kan er sprake zijn van (te) veel vertrouwen in de experts en expertise van een instituut, wat democratisch toezicht of controle in de weg staat. Voor de ontwikkeling van de eindtoets, die al meer dan vijftig jaar op dezelfde leest is geschoeid, kan met terugwerkende kracht gesproken worden van institutioneel 'oververtrouwen'. De overheid heeft de eindtoets overgenomen zonder de test zelf, de waarde en de effecten ervan voor de praktijk van het onderwijs en samenleving te bevragen wat betreft fairheid en kansengelijkheid. De ontwikkelingen in de afgelopen decennia met betrekking tot de belangrijkste kernbegrippen van de testwetenschap, de validiteit en de betrouwbaarheid, vormen daarbij aanleiding huidige vormen van geschiktheidstests zoals de eindtoets opnieuw te beschouwen in het licht van de huidige wetenschap, ook ten aanzien van intelligentie en de ontwikkelbaarheid ervan.







# HOOFDSTUK 9

Balans en debat





## Hoofdstuk 9 Balans en debat

*'Het feit dat de mens in staat is tot handelen, betekent dat het onverwachte van hem kan worden verwacht, dat hij kan waarmaken wat in de hoogste mate onwaarschijnlijk is. En dit weer is slechts mogelijk omdat ieder mens uniek is. Met betrekking tot deze iemand die uniek is kan in waarheid worden gezegd: daar was voordien niemand.'*

*Uit: Hannah Arendt, 'De menselijke conditie' (2009)*

In dit onderzoek is de eindtoets nader onderzocht. Daarbij is ervoor gekozen de eindtoets niet louter te beschouwen als een test en in die hoedanigheid te onderzoeken maar als een beleidsinstrument, dat, gegeven de lange historie en de rol die de eindtoets is gaan vervullen in de inrichting van onderwijs en samenleving, is onderzocht als instituut en als institutie. De rol van de eindtoets op dit moment laat zich het best begrijpen door positief op elkaar inwerkende en elkaar versterkende mechanismen: diepgewortelde tradities van het Nederlandse onderwijs ten aanzien van selectie hebben de weg vrijgemaakt voor een instrument dat een uitweg belooft uit een impasse die dan al honderd jaar duurt, namelijk het vraagstuk van wie toegelaten zou mogen worden tot de maatschappelijk het meest in aanzien staande onderwijstrajecten. Het zoeken naar de meest geschikten past bij het streven naar een meer meritocratische herordening van de sociale structuren waarbij duidelijk is geworden dat de manier van selecteren de bestaande ordening eerder bevestigt dan dat ze ruimte biedt aan een fundamentele herordening. Dit wordt extra duidelijk in de periode dat Nederland zich geconfronteerd ziet met het meer permanente karakter van de immigratie. Het vertrouwen in toetsing kan groeien door de behoefte van de sociale wetenschappen om met een natuurwetenschappelijke benadering voor het doen van onderzoek meer voor vol te worden aangezien in de wereld van onderzoek. Door gebruik te maken van objectieve, gestandaardiseerde en gevalideerde vormen van tests hebben zij bijgedragen aan een vergroting van het vertrouwen in dergelijke vormen van meten en de onderliggende constructen. Dit versterkt vervolgens ook het belang en de geldingsdrang van de testwetenschap zelf. Ook daar is omwille van de instandhouding van het imago van deze tak van wetenschap gezocht naar theoretische constructen en 'oplossingen' die, terugkijkend, met de wetenschap van nu, het nodig maken dit opnieuw te beschouwen. De klassieke testtheorie is in de afgelopen decennia tegen beperkingen aangelopen die het waard zijn om in de huidige toepassingen kritisch te worden onderzocht. Dit geldt ook voor de effecten die wereldwijd te rapporteren zijn op basis van het gebruik van of de inzet van extern gevalideerde vormen van testen. Dit wordt bemoeilijkt door het feit dat er een toename is van behoeften en belangen van overheden om met behulp van tests grip te willen krijgen op opbrengsten en resultaten van onderwijs en daarmee ook op de inrichting van onderwijs en samenleving. In combinatie met een afname van inhoudelijke deskundigheid vanuit een besturingsfilosofie waarbij de overheid regie voert zonder diepgaande, inhoudelijke kennis van zaken, kan dit leiden tot het sturen op resultaten of opbrengsten die ofwel niet of beperkt van waarde zijn, ofwel niet realiseerbaar zijn en het doel van kwaliteitsverbetering mogelijk eerder verder weg dan dichterbij brengen. En daarmee schadelijke effecten genereren voor wat met onderwijs wordt beoogd.

Dit hoofdstuk is te lezen als een samenvatting van de vorige hoofdstukken waarbij de balans van de eerdere hoofdstukken wordt opgemaakt (9.1), er conclusies worden getrokken over de huidige situatie (9.2) en er gekeken wordt naar wat wel en geen reële verwachtingen zijn ten

aanzien van tests als beleidsinstrumenten (9.3). Tot slot wordt gekeken hoe een beter evenwicht kan worden bereikt en welke discussies daarmee open liggen (9.4).

## 9.1 De balans opmakend

De eindtoets die nu aan het eind van groep 8 verplicht moet worden afgenomen, heeft zich ontwikkeld tot een beleidsinstrument in handen van de overheid, waarmee de overheid controle heeft over zowel het vraagstuk van de selectie als over het vraagstuk van de kwaliteit in termen van opbrengsten van onderwijs. De Inspectie van het Onderwijs is belast met de uitvoering van de controle. Hoewel de afname van de eindtoets bewust is verplaatst naar een moment later in het schooljaar met de intentie om daarmee de druk voor leerlingen te verminderen, is het belang van de eindtoets niet verminderd. Integendeel. De aan de toetsing ontleende, gestandaardiseerde scores op basis waarvan leerlingen worden vergeleken met elkaar en met wettelijk vastgestelde adviescategorieën voor het voortgezet onderwijs, zijn in hoge mate bepalend voor de loopbaan in het voortgezet onderwijs. Hierbij zijn de afgelopen jaren steeds meer hindernissen opgeworpen die tussentijds overstappen en doorstromen bemoeilijken. De scores op de eindtoets fungeren bij de advisering voor die overgang als norm waartegen de door de scholen gegeven adviezen worden afgezet als indicator van kwaliteit; zijn de gegeven adviezen in lijn met de uitkomsten van de toets? Voor de beoordeling van de kwaliteit van het gegeven onderwijs op individuele scholen vormen de scores op de eindtoets een belangrijk ankerpunt; zijn de leerresultaten in lijn met dat wat verwacht mag worden gezien de samenstelling van de populatie van een school? Zowel voor de selectiefunctie als voor de kwaliteitsbeoordeling van individuele scholen is onderlinge vergelijking het uitgangspunt. De scores op de verschillende toetsonderdelen taal en rekenen worden bij elkaar opgeteld en via een lineaire transformatie omgezet in jaarlijks vergelijkbare standaardscores op basis waarvan de leerlingen kunnen worden geordend met één getal, van hoog naar laag scorend (van 550 naar 501), waarbij grenzen zijn getrokken die gelden als cesuren voor de verschillende vormen van voortgezet onderwijs. Die aanpak maakt een jaarlijks min of meer stabiele verdeling van aantallen leerlingen over de verschillende vormen van voortgezet onderwijs mogelijk. De invoeging van referentieniveaus voor taal en rekenen in de bestaande eindtoets vormt vanaf 2020 een nieuwe basis voor een kwaliteitsoordeel over een individuele school waarbij de manier waarop achtergrondvariabelen van leerlingen meewegen ook is aangepast: scholen worden vergeleken met elkaar op basis van een gecombineerde score op taal en rekenen, rekening houdend met de populatie leerlingen waarop de scores zijn gebaseerd. Ook in het geval van het nieuwe onderwijsresultatenmodel hanteert de Inspectie van het Onderwijs voor de beoordeling van de kwaliteit van scholen een competitief, op groei gericht, model en zal er sprake blijven van een competitief beoordelingsmodel, waarbij zowel scholen als leerlingen aangezet worden tot presteren in een wedstrijd die hoe dan ook verliezers kent.

### **Verbinding onderwijs en arbeidsmarkt**

Het groeiend belang dat in de loop van de tijd aan de uitkomsten van de eindtoets is toegekend kan op een aantal manieren verklaard worden. Als eerste verklaring geldt een diep in het onderwijs verankerde behoefte om onderwijs te zien als instrument in dienst van arbeidsmarkt en economie. In de honderd jaar die aan de eerste versie van de eindtoets vooraf ging, is in Nederland voortdurend gezocht naar een 'oplossing' voor de uitval uit het onderwijs door te zoeken naar betere selectiemechanismen. Waar veel andere landen oplossingen voor uitval vooral zochten in andere richtingen, zoals het aantrekkelijker maken van het voortgezet onderwijs zelf, het bieden

van meer tijd en ruimte aan de leerling om zich te ontwikkelen (uitstel van selectie) of door meer aandacht te geven aan de vorming van de jeugd, is het Nederlandse beleid vooral gericht geweest op het verbeteren van de selectie om daarmee arbeidsmarkt en samenleving zo efficiënt en effectief mogelijk te bedienen. De toename van het belang van onderwijsdiploma's past in deze visie. Ook in de hoogtijdagen na de democratiseringsgolf van de jaren zestig is de economische gerichtheid van het onderwijs de rode draad gebleven van het onderwijsbeleid. In de jaren tachtig werd de grip op het onderwijs verstevigd vanuit zorg over de enorme jeugdwerkloosheid die met meer en beter op de arbeidsmarkt aansluitende scholing zou kunnen worden aangepakt. Het probleem van 'overscholing' leidt tot ingrijpen in onderwijsbeleid zodat niet te veel jongeren de weg naar het hoger onderwijs kiezen. De jaren negentig hebben de economische gerichtheid van het onderwijs getechnocratiseerd met het zoeken naar kwaliteitsindicatoren waarmee de overheid meer op afstand de opbrengsten van onderwijs kan bewaken. De aanvankelijke balans van meer autonomie voor de scholen in ruil voor nieuwe vormen van verantwoording is nooit goed van de grond gekomen. De overheid zelf heeft zich niet alleen bezig gehouden met het 'wat' van het onderwijs, in termen van doelen en verantwoordingen op de doelen met behulp van metingen, maar heeft zich middels een verplicht gebruik van eindtoets en LVS nadrukkelijk begeven op het pad van het 'hoe' van het onderwijs; door toe te zien op de concrete invulling van 'opbrengstgericht werken' en door met behulp van tests, ook gedurende het onderwijs, te bepalen waar de focus in het onderwijs op gericht zou moeten worden en hoe dat zou moeten gebeuren. Op dit moment is het gangbaar geworden om over het doel van onderwijs te spreken in termen van een 'drieslag'. Daarmee wordt bedoeld dat onderwijs gericht moet zijn op de volgende drie functies: kwalificatie, socialisatie en subjectwording/persoonsvorming. Toetsing blijkt op dit moment met name gericht op de kwalificatiefunctie van onderwijs en voor het basisonderwijs betekent dit vooral een gerichtheid op taal en rekenen. Taal en rekenen, althans die aspecten daarvan die met gestandaardiseerde toetsen als de eindtoets en de toetsen van het LVS gemeten worden, zijn nu prominent in het onderwijs gepositioneerd. Dit is ten koste gegaan van de waardering van ook andere doelstellingen van het onderwijs, zoals de Onderwijsraad de laatste jaren stelt: dat wat meetbaar is, staat meer centraal dan dat wat merkbaar is en aan belang niet per se onderdoet voor dat wat meetbaar is. Deze technocratische onderwijsoriëntatie in combinatie met een zucht naar het meetbaar willen maken van kwaliteit staat onderwijs met aandacht voor brede doelen in de weg. Er kan pas ruimte ontstaan voor het zien van de waarde van de volle breedte van het onderwijs als de gedachte dat van onderwijs een cijfermatige balans kan worden opgemaakt, verlaten wordt.

### ***Gelijke kansen en testen***

Van oudsher is in Nederland het onderwijs gebruikt om de 'besten' of de 'slimsten' te selecteren. Diploma's worden gezien als een eerste waarborg van de selectieprocedure richting arbeidsmarkt. De score op de eindtoets, maar nu ook de scores op de toetsen van het LVS, vormen de eerste stappen in die selectie. De verplichting de uitkomsten van LVS-toetsen te gebruiken bij het opstellen van schooladviezen in combinatie met scores op de eindtoets hebben aan toetscores niet alleen een grotere waarde toegekend maar ook de selectie in feite eerder vervroegd dan verlaat. Van uitstel van selectie is in Nederland dan ook de facto nooit sprake geweest. De beperkingen van mogelijkheden om binnen het voortgezet onderwijs van leerweg te wisselen, vooral opstroom, hebben de gevolgen van die selectie verder vergroot. Steeds vroeger worden de kansen van een leerling op een hoog schooladvies bepaald. Het LVS is uitgegroeid tot een

instrument niet alleen bedoeld om deficiënties vast te stellen op basis waarvan in het onderwijs aandacht naar individuele leerlingontwikkelingen kan gaan, maar is vooral een instrument geworden om verwachtingsmanagement van scholen, ouders en leerlingen mee vorm te geven. Dit verwachtingsmanagement is nodig omdat er in het onderwijs sprake is van een verdelingsvraagstuk waarbij niet voor iedereen elke gewenste vorm van onderwijs beschikbaar is. De helft van de leerlingen krijgt een advies met daarin een verwijzing naar een vorm van vmbo. En slechts 20% van de leerlingen maakt kans op een directe verwijzing naar het vwo. Daarmee heeft onderwijs meer en meer het karakter gekregen van een competitie die in het voordeel van leerlingen uitvalt die het best weten te scoren op een bepaalde afspiegeling van taal- en rekenvaardigheid en daarmee de meeste kans hebben op een vwo-advies. In de praktijk blijken dit vooral de kinderen van autochtone, hoger opgeleide ouders te zijn. Gelijke kansen betekent dan ook zeker niet dat kinderen, ongeacht hun achtergrond, moeten kunnen komen tot gelijke onderwijsresultaten. Het sturen op ongelijke onderwijsuitkomsten is een expliciete doelstelling van de eindtoets en van de toetsen van het LVS. Want de toetsen zijn gericht op het kunnen differentiëren van leerlingen zodat ze geordend kunnen worden van laag naar hoog. Ook in de beoordeling van de kwaliteit van het gegeven onderwijs door de scholen is de bestaande ongelijkheid het uitgangspunt. Afhankelijk van de mate van achterstand worden voor scholen te behalen doelen geformuleerd: voor leerlingen met grote sociaal-economische achterstanden zijn lagere taal- en rekenresultaten acceptabeler dan voor leerlingen met sociaal-economische voorsprong. Bevestiging van verschillen in resultaten voor taal en rekenen, mits te herleiden tot sociaal-economische achterstanden, leidt tot een positief inspectie-oordeel: de school levert dan (voldoende) kwaliteit. Voor de leerling en zijn ouders betekent 'gelijke kansen' vooral dat alle leerlingen dezelfde toets op hetzelfde moment maken zodat ze 'eerlijk' met elkaar vergeleken kunnen worden. Niet de geschiktheid of passendheid ten opzichte van een vorm van vervolgonderwijs staat daarbij centraal, maar de relatieve positie ten opzichte van de rest van het deelnemende cohort leerlingen, op basis van hun gecombineerde score taal en rekenen, is daarbij het uitgangspunt. En die score wordt maatschappelijk geassocieerd met en geaccepteerd als een bepaalde mate van intelligentie waarmee de rechtvaardigheid van de adviezen wordt onderbouwd. De invulling die zo aan het begrip 'gelijke kansen' wordt gegeven is vooral een procedurele. Het selecteren van de hoogst scorenden, ofwel de 'besten', bepaald met een toets die aan iedereen op dezelfde manier wordt voorgelegd, vormt de basis voor de beoogde en gewenste meritocratische ordening van de samenleving. Terwijl uit onderzoek blijkt dat meritocratie in vergelijking met een aristocratische standenmaatschappij tot een andere ordening leidt, maar zeker niet tot minder harde ongelijkheid. De meritocratische uitgangspunten van onderwijs en met name de toetsing bevestigen dit: toetsing en de uitkomsten ervan leiden zowel op het niveau van de school als op het niveau van de leerling meer tot bevestiging van de uitgangssituatie bij intrede in het funderend onderwijs, dan dat het leidt tot het wegwerken van verschillen en het vergroten van kansen voor kinderen die met achterstanden het onderwijs instromen. Als onderwijs de plaats moet zijn waar kinderen kansen geboden worden zich te ontwikkelen en te groeien, zijn de feitelijke vervroeging van de selectie, de vergelijkende manier van meten in combinatie met de schaarste aan onderwijsplekken, de vormen van differentiatie en de manier waarop nu kwaliteit van onderwijs van een school wordt beoordeeld, geen maatregelen waarmee dit doel dichterbij gebracht kan worden. De eindtoets en het daarmee samenhangende kwaliteitsoordeel van scholen door de overheid dragen eerder bij aan een verdere bevestiging van de bestaande maatschappelijke ongelijkheid en maken zo een verdere verharding van ongelijkheid mogelijk.

### ***Vaststaand gegeven of score op een test?***

De score op de eindtoets is vanaf het begin gezien als een indicatie van de intellectuele capaciteiten van een leerling: hoe hoger de score, hoe meer het kind heeft opgestoken van het voorgaande onderwijs, hoe slimmer het moet zijn en hoe geschikter het zal zijn voor algemeen vormende tracks als havo/vwo. Want wie het meest geprofiteerd heeft van het aangeboden onderwijs, zou de grootste kans hebben de investering van het voortgezet onderwijs waar te maken. Ten tijde van de ontwikkeling van de eerste eindtoets legt de wet beperkingen op aan het ontwikkelen en afnemen van intelligentietests. Dat is voorbehouden aan psychologen. De opkomst van schoolvorderingstoetsen in de Verenigde Staten bieden een oplossing, onder de aanname dat resultaten uit het verleden de beste garantie bieden voor de toekomst. Ook de eerste bedenker van de eindtoets, A.D. De Groot, ziet intelligentie als iets dat in aanleg in meer of mindere mate aanwezig is en dat zich al op jeugdige leeftijd uitkristalliseert en dan al gemeten kan worden. Ook in de decennia daarna wordt de eindtoets door de makers van dan het Cito gepositioneerd als een intelligentietest, 'maar dan beter'. De scores van de verschillende testonderdelen worden bij elkaar opgeteld, lineair getransformeerd naar een standardscore zodat er met één getal uitdrukking kan worden gegeven aan de 'hoeveelheid' die iemand ergens van heeft. En dat ergens wordt gedefinieerd als de 'ability', als een combinatie van intelligentie en inzet. Door het ontwerp van de toets en door psychometrisch aan te sturen op een model van een normaalverdeling, wordt nog eens versterkt dat met de eindtoets een eigenschap, een kenmerk wordt vastgesteld die een normaalverdeling kent: je kunt er veel of juist weinig van hebben. Alle leerlingen uit een cohort worden op basis van een enkele variabele met elkaar vergeleken. Door een enkele variabele uit te vergroten tot een alles bepalend kenmerk van een persoonlijkheid, is het individu in al zijn veelzijdigheid naar de achtergrond gedrongen. Hij is gereduceerd tot een score die synoniem is voor zijn slimheid. Daar komt bij dat een individu met zijn individuele score ook vergeleken wordt met de scores die behaald zijn door eerdere cohorten leerlingen, hún gemiddelde standardscores en hún succes in het voortgezet onderwijs. Het succes en falen van die eerdere generaties in het voortgezet onderwijs wordt herleid naar hun eerdere score op de eindtoets, waarmee dat succes en falen bepalend zijn geworden voor de kansen van nieuwe generaties, terwijl succes en falen in het voortgezet onderwijs aan tal van niet aan taal en rekenen gerelateerde aspecten kan worden geweten. De waarde die wordt toegekend aan de standardscore die wordt gezien als een maat van intelligentie, is tot mythische proporties gegroeid: leerlingen met de hoogste scores zijn de besten, de slimsten en hebben recht op de meest begeerde onderwijstrajecten met de langste onderwijstijd en de grootste kansen op de beste posities op de arbeidsmarkt, dat alles aangetoond door leerlingen met dezelfde scores die hen voor zijn gegaan in het voortgezet onderwijs. Terwijl er nauwelijks aandacht is voor werking van de 'selffulfilling prophecy': hoog waarden leidt tot goed presteren. En andersom. Dit geldt ook voor de eindtoets. Alle diversiteit is met deze vorm van toetsing teruggebracht tot een enkele score met grote gevolgen: er zijn winnaars en verliezers gecreëerd waarbij de winnaars zich de winst toe-eigenen en de verliezers vol schaamte wegduiken. De toetsing is zo gericht dat er sprake is van exclusiviteit en buitensluiting terwijl de pedagogische opdracht van onderwijs er een hoort te zijn van inclusiviteit: iedereen doet ertoe en iedereen mag meedoen.

### ***Subjectieve objectiviteit***

De eerste versie van de eindtoets verschijnt in 1966. De eindtoets zoals we die nu kennen is daarmee meer dan 50 jaar oud en is in de basis in hoge mate dezelfde gebleven. Vanaf de ont-

wikkeling is er veel nadruk gelegd op de objectiviteit van de beoordeling die met de eindtoets gepaard kon gaan. Door het gebruik van meerkeuzevragen is de beoordeling van de juistheid van een antwoord met minder subjectiviteit omgeven, sterker, vanaf het begin is er de mogelijkheid de verwerking geheel geautomatiseerd te laten verlopen. Maar een objectief vastgestelde score is niet mogelijk zonder dat daar een keur aan subjectieve besluiten aan vooraf is gegaan, of die gevolgd worden door tal van subjectieve besluiten die samenhangen met de keuzes die gemaakt worden bij de (psychometrische) verwerking van de data, in termen van classificeren en standaardiseren. Het ontrafelen van de stappen en besluiten die met de eindtoets samenhangen, laat zien dat de eindtoets een aaneenschakeling is van subjectieve, door menselijke oordelen en ingrijpen beïnvloede besluiten. Al deze besluiten zijn lang geleden genomen en zijn voor een belangrijk deel niet direct te achterhalen. Door indirect, op basis van de werking van de toets zelf, een substantieel deel van deze subjectieve besluiten aan de oppervlakte te brengen, ontstaat weer een beeld van de subjectieve objectiviteit die daadwerkelijk achter de eindtoets schuilgaat.

### ***Smalle verantwoording met grote gevolgen***

Onder het motto 'niet roeien maar sturen' is de overheid in het kader van NPM voor onderwijs gaan zoeken naar indicatoren waarmee kwaliteit van onderwijs eenvoudig en efficiënt kan worden uitgedrukt en op basis waarvan toezicht op kwaliteit van onderwijs kan worden vormgegeven. De reeds bestaande en door Cito B.V. geëxploiteerde eindtoets is daartoe overgenomen en aangepast: leerstofonafhankelijke niveaus voor taal en rekenen zijn ingevoegd. Door het toezicht te richten op een voortdurende groei van prestaties op (een beperkte afspiegeling van) taal- en rekenvaardigheid heeft de overheid met deze toevoeging het competitieve karakter en de resultaatgerichtheid van het primair onderwijs verder vergroot. De smalle vertegenwoordiging van de veelheid van onderwijsdoelen in de indicatoren zorgt ervoor dat scholen op een relatief klein aandeel van al hun onderwijsinspanningen worden 'afgerekend'. Internationale voorbeelden laten zien dat dit op termijn kan leiden tot een verschraving van het onderwijs en tot een cultuur waarin niet kwaliteit van onderwijs centraal staat, maar hoge scores op gestandaardiseerde toetsen. De focus op 'meetbare' indicatoren kan zo een ernstige bedreiging vormen voor de kwaliteit van onderwijs: leerlingen zijn gebaat bij het verwerven een brede kennisbasis en niet bij het drillen en trainen voor toetsen die maar een smal deel van het curriculum dekken en waarvan niet is aangetoond dat juist dat deel van het curriculum de leerlingen krachtiger zou voorbereiden voor vervolgonderwijs en samenleving.

### ***Toetslast***

De bij de start van de eindtoets in 1966 als moeizaam geziene combinatie van én selectie én beoordeling van onderwijskwaliteit speelt de eindtoets nog altijd parten. Is aanvankelijk de selectie gericht op het bepalen van wie wel en wie niet geschikt zouden zijn voor het v.h.m.o./vwo, daar zijn nu meer beslissingen aan toegevoegd: er zijn nu acht groepen die van elkaar moeten worden onderscheiden, wat betekent dat er zeven cesuurpunten in de toets zijn besloten, terwijl de toets minder opgaven is gaan bevatten en grotere gevolgen heeft wat betreft de uitkomsten voor betrokkenen. Daar is nu de vaststelling van de beheersing van verschillende referentieniveaus aan toegevoegd om op basis daarvan de kwaliteit van scholen te kunnen beoordelen. En zo is aan de steeds grotere waarde die al wordt toegekend aan de score in relatie tot het schooladvies, nu ook het oordeel over de geleverde onderwijskwaliteit van een school toegevoegd. Daarmee is kwaliteit van onderwijs afhankelijk geworden van de scores die leerlingen laten zien op de eindtoets.

Zowel het selectie-element als het schoolkwaliteitsbeoordelingsaspect zijn de afgelopen jaren steeds verder verzaamd zonder grondig onderzoek naar de draagkracht van eindtoets, gegeven de verschillende ambities en gewichttoekenningen. Nader onderzoek naar de ontwikkelingen in de psychometrie van de afgelopen decennia geeft aanleiding om opnieuw naar de eindtoets te kijken. Zeker nu er nieuwe aspecten aan de eindtoets zijn toegevoegd is het van het grootste belang te onderzoeken of de eindtoets hierop figuurlijk is 'berekend'. Uit dit onderzoek is een beeld ontstaan van een intensivering van de aanvankelijke aarzeling van De Groot in 1966 die destijds over de verschillende functies van de eindtoets zei: 'het moet kunnen'. Het beeld dat op basis van dit onderzoek is ontstaan, is eerder: 'Het kan niet, maar het moet wel'.

## 9.2 Institutioneel 'oververtrouwen'

De ontwikkelingen in de natuurwetenschappen hebben de menswetenschappen zodanig beïnvloed dat ook het onderwijs onder invloed is gekomen van een meer 'empirisch-analytische' wetenschapsopvatting. A.D. de Groot, bedenker van de eindtoets en gezien als 'founding father' van het Cito, heeft hierin een belangrijke rol gespeeld. De manier waarop hij toetsing heeft ingevoegd in een empirisch-analytisch wetenschapskader, heeft een sterke basis gelegd voor het geloof in de wetenschappelijke beloften die met toetsing in de context van onderwijs zouden kunnen worden waargemaakt. Zijn rol in de ontwikkeling van de empirisch-analytische wetenschap meer in het algemeen heeft tevens geleid tot vertrouwen in zijn manier van denken over testen. Deze specifieke manier van benaderen van tests heeft zijn beslag gekregen in de oprichting van een zelfstandig, in eerste instantie aan de overheid gelieerd instituut. Daarmee heeft het denken over testen zoals destijds gezien en uitgewerkt door De Groot ruimte gekregen om breed uitgerold te worden in onderwijs, politiek en samenleving. Een bepaalde manier van denken en kijk op de wetenschap is hiermee geïnstitutionaliseerd. Dit denken heeft zijn weerslag gekregen in het Cito als instituut maar ook in de eindtoets, die gezien kan worden als een belangrijke representant van het empirisch-analytisch denkkader waarmee de onderwijswerkelijkheid van het basisonderwijs gevormd is. De eindtoets laat zien welke normatieve doelstellingen de maatschappelijke orde nastreeft of nagestreefd heeft. Elke institutie heeft naast een empirisch ook een normatief aspect. Het empirische uitgangspunt van dat wat gemeten wordt, objectief vastgesteld moet kunnen worden, heeft voor de eindtoets betekend dat de inhoud van de meting beperkt is tot datgene dat met behulp van meerkeuzevragen geoperationaliseerd kan worden, omdat alleen dan een 'objectieve' beoordeling mogelijk zou zijn. Dit streven naar 'objectiviteit' heeft geleid tot een keur aan 'subjectieve' keuzes die het primair onderwijs hebben beïnvloed zonder dit een plek te geven in een breed maatschappelijk debat. Onderwijs is met dit uitgangspunt in de afgelopen vijftig jaar versmald tot dat wat gemeten kan worden met objectief te beoordelen toetsen. Hoewel De Groot destijds nooit heeft gezegd dat andere aspecten dan de meetbare niet van waarde zouden zijn voor onderwijs, is dit wel een belangrijk terugslag-effect gebleken van de manier waarop toetsing in Nederland is vormgegeven. Het is nu terecht de vraag te stellen of we meten wat gemeten kan worden of dat we meten wat van belang is om te meten. Als alleen dat wat met objectief te beoordelen toetsen meetbaar kan worden gemaakt in de praktijk een plek blijkt te krijgen in toetsing, dan krijgt de inhoud van wat gemeten wordt op die manier vanzelf een grotere relevantie voor het onderwijs. Daarmee zijn de kaders voor onderwijs en toetsing niet in handen van democratische processen maar worden deze geregeerd vanuit institutionele belangen. Het institutionaliseren van kennis op het gebied van toetsing bij externe instituten als Cito en CvTE draagt ertoe bij dat de overheid weliswaar verantwoordelijk is voor wat er gebeurt in onderwijs

op het gebied van onderwijsinhoud en toetsing, maar hier nauwelijks regie op kan voeren. Daar komt bij dat de controle en het toezicht op onderwijs voor een belangrijk deel in handen van de Inspectie van het Onderwijs ligt, die middels toetsing moet toezien op en rapporteren over de kwaliteit van onderwijs. Dit uit handen geven en delegeren vanuit de overheid heeft ertoe geleid dat niet alleen op het operationele niveau de overheid geen grip meer heeft op de praktijk, maar ook op het niveau van het maken van collectieve keuzes leunt op de door de instituties aangegeven mogelijkheden en beperkingen ten aanzien van wat op welke manier getoetst kan worden. Zelfs op het constitutionele niveau blijkt ten aanzien van toetsing (en examinering) in de afgelopen decennia in hoge mate gevaren te zijn op het oordeel en de belangen van externe instituten. Daar waar scholen beperkt hebben weten te profiteren van een overheid die met vertrouwen taken delegeert die gaan over de uitvoering van het onderwijs, hebben de toetsen en examens als instituties en de instituten waar deze ontwikkeld worden of waar toezicht wordt gehouden op die ontwikkeling, wel degelijk geprofiteerd van een overheid die met vertrouwen de eigen taken heeft gedelegeerd aan die instituties. Er is sprake van zoveel vertrouwen dat er nauwelijks voldoende kennis op het ministerie aanwezig is om de eigen instituties kritisch te kunnen volgen, te controleren en de gemaakte of te maken keuzes te positioneren in een democratisch debat. Zoals een hoge ambtenaar van OCW terugkijkend stelde: 'Wij hadden geen idee'. Hiermee is een situatie ontstaan waarbij vertrouwen hebben in de instituties geen keuze meer is maar een noodzaak omdat de overheid zelf over onvoldoende expertise beschikt. In dat licht kan er gesproken worden van institutioneel 'oververtrouwen'. Dit hindert een democratische debat over de rol van onderwijs in de samenleving. Er wordt gestuurd en afgerekend op cijfers zonder dat er zicht is op hoe de cijfers tot stand komen, wat de onderliggende keuzes zijn en wat de gevolgen zijn van een dergelijke focus. Cijfers, gebaseerd op scores op de huidige toetsen, worden gezien als een betrouwbare afspiegeling van prestatiemeting- en beoordeling. Maar de cijfers dienen vooral het belang van het inzichtelijk kunnen maken van individuele prestaties, van de leerling, via de leerling van de school en via het geheel van scholen van het onderwijs. De prestaties van leerlingen en scholen kunnen worden afgezet tegen de prestaties van andere individuen en andere scholen. De vraag is echter wat die kennis bij kan dragen aan verbetering van de kwaliteit van onderwijs in de volle breedte? En wat er op deze manier overblijft van de vrijheid van onderwijs? Hoeveel ruimte is er daadwerkelijk voor scholen, ouders en leerlingen om te kiezen? En krijgen de kinderen het onderwijs dat ze nodig hebben?

### 9.3 Eindtoets onder druk

De eindtoets heeft altijd gehinkt op twee ambities: die van de wettelijke opdracht om te komen tot een selectie-instrument dat als aanvulling zou dienen op het oordeel van de school over een passend schooladvies en die van het willen laten zien dat toetsing nuttig kan zijn in het opmaken van de balans van onderwijs in termen van opbrengsten en kwaliteit. Vanaf het allereerste begin hebben de ontwikkelaars zelf getwijfeld over de verenigbaarheid van alle doelen waarbij vooral de wens ze te kunnen combineren de vader van de gedachte is geweest. Het ontbreken van concrete leerdoelen heeft ertoe geleid dat destijds, tamelijk willekeurig en zelfs tegen het gevoel van de toen betrokken ontwikkelaars in, taal en rekenen veel aandacht hebben gekregen in de toets terwijl de gedachte was dat studievaardigheden en algemene ontwikkeling veel meer en beter geschikt zouden zijn voor het bepalen van cognitieve capaciteiten. Het ontbreken van leerdoelen om toetsvragen op te baseren, zoals de empirische cyclus voorschrijft, is ondervangen door te stellen dat met behulp van concrete toetsopgaven de discussie over doelen van het onderwijs

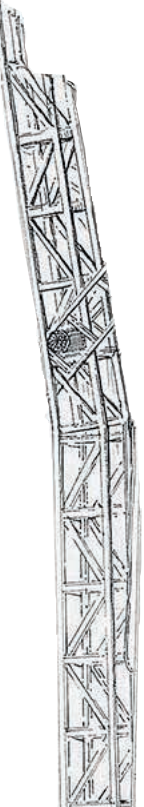


op gang zou kunnen worden gebracht. Dat laatste is niet gebeurd. En taal en rekenen, tenminste die taal- en rekenvaardigheid die met meerkeuzevragen zijn te bevragen, zijn de basis gebleven van de toets. En de opgaven in de toets zijn in hoge mate bepalend geworden voor de concrete invulling van de onderwijspraktijk, inclusief de leermiddelen. Ook ten aanzien van selectie is het anders gelopen dan aanvankelijk gedacht. Met het uitzicht op de invoering van een brugklas waar de selectie zich verder zou uitkristalliseren werd door de eerste ontwikkelaars selectie niet als een belangrijke, nieuwe uitdaging gezien. De ambitie voor het ontwikkelen van een toets had veel meer betrekking op het willen kunnen onderzoeken van 'de stand van het onderwijs'. Zo is het niet gelopen. Selectie is het primaire doel van de eindtoets gebleven en substantieel uitgebreid. Er is nu niet alleen de vraag wie mag naar het vwo, maar er is een complexer verdelingsvraagstuk ontstaan waarbij leerlingen verdeeld worden over acht adviescategorieën. Deze uitbreiding van cesuren heeft aan de aard en opzet van de toets niets veranderd. Nog altijd is de toets cognitief gericht en beperkt tot opgaven voor taal en rekenen die zodanig zijn geconstrueerd dat schifting en ordening mogelijk is. Maar met een verdere beperking van het aantal opgaven, een vergroting van het aantal adviescategorieën en een verzwaring van de belangen van verwijzing, moet de huidige eindtoets meer opdrachten realiseren dan bij de constructie van de eerste eindtoets in 1966 was te voorzien. Aan de aanvankelijke discussie over het dubbele karakter van de toets van zowel individuele selectie als stand-van-zaken-bepaler is in zekere zin een einde gekomen met de wetwijziging van 2011. Uit die wetwijziging blijkt dat de overheid wil kiezen voor een criteriumgerelateerde meting aan de hand van referentieniveaus als leerstofonafhankelijke norm voor het vergelijken van scholen met elkaar. Tegelijkertijd wil de overheid dit realiseren binnen dezelfde toets. Daarmee wordt opnieuw aan de bestaande toets een nieuwe opdracht gegeven. De toets moet differentiëren tussen meer adviescategorieën dan ooit en tegelijkertijd aantonen dat bepaalde vastomlijnde beheersingsdoelen door het gros van de leerlingen, toe te schrijven aan het effectieve handelen van scholen, worden behaald. Dit is een contradictoire opdracht die huidige toetsontwikkelaars voor hetzelfde dilemma stelt als het team van De Groot destijds in 1966. Ook nu zijn de belangen groot. De eindtoets blijft daarmee een hybride-variant op de schaal van toets naar test. Een combinatie van een criterium- én tegelijkertijd normgerelateerd meetinstrument gaat ten koste van de kwaliteit van het geheel, al was het maar omdat de validiteit ervan niet te bepalen is. Door te kiezen voor een validering in de vorm van een correlatie met de positie in het vervolgonderwijs maakt het selectieprobleem tot een tijdelijk probleem; het opleggen van een strikte toepassing van de adviescategorieën maakt van het advies op basis van de toets op termijn een selffulfilling prophecy. Na een aantal jaar werken met dit systeem zullen er nauwelijks afwijkingen meer gevonden kunnen worden. Op dit moment zijn er nog altijd leerlingen die in weerwil van hun score op de eindtoets het voordeel van de twijfel krijgen en daarmee laten zien dat de toets zijn 'voorspelling' lang niet altijd waarmaakt. Het lijkt een kwestie van tijd dat dit zich niet meer voordoet. Daarmee wordt de toets op termijn een zichzelf waarmakend instrument. Toetsing, begonnen als empirisch-analytisch instrument waarmee evidence-based beslissingen tot stand zouden kunnen komen, is in handen van de overheid uitgegroeid tot een instrument dat gekarakteriseerd kan worden als 'evidence-transforming'. Met de eindtoets wordt niet langer gepoogd een 'proxy' te hebben van een bepaalde interpretatie van de werkelijkheid, maar er wordt een werkelijkheid 'waar' gemaakt die een eigen leven gaat leiden.

#### 9.4 Democratisch debat

Kansengelijkheid, het wegwerken van (onderwijs)achterstanden en sociale mobiliteit zijn vraag-

stukken die veel taaier zijn gebleken dan aanvankelijk de gedachte is. De vraagstukken zijn te groot om neer te leggen bij onderwijs om daar 'op te lossen' en gaan de huidige eindtoets zeker ver te boven. Bovendien staat de manier waarop onderwijs nu geacht wordt te streven naar het aanpakken van deze vraagstukken op gespannen voet met andere invalshoeken die op onderwijs en samenleving betrekking hebben. Naar succes en falen in onderwijs wordt op dit moment op een eendimensionale manier gekeken. Dit heeft gevolgen. Onderwijs in dienst van economisch georiënteerde selectiebelangen zorgt niet voor inclusiviteit maar sluit uit. De vraag is of onderwijs waar pluriformiteit het uitgangspunt is en waar vrijheid en ruimte geboden wordt aan mensen om verschillend te zijn en om hun eigen capaciteiten te ontdekken, een samenleving niet veel meer brengt. Het is van belang het vraagstuk van de 'fairness' van het huidige competitie- en selectiemodel in dat licht te bevragen. Dit proefschrift heeft de bedoeling gehad de onderliggende uitgangspunten van de eindtoets weer zichtbaar te maken en te plaatsen in het licht van de huidige stand van wetenschap en samenleving. In 1966 roept A.D. de Groot op tot een 'democratisch protest' tegen de beoordelings- en beslissingspraktijken in ons paternalistische onderwijs' (De Groot, 1966: 245). Het is de hoogste tijd dat, na ruim vijftig jaar eindtoetsing, er nu een vergelijkbare oproep komt met betrekking tot de praktijken van het beoordelen en beslissen bij de overgang van basis- naar voortgezet onderwijs, zodat een democratisch debat gevoerd kan worden over selectie en kwaliteit in het perspectief van het bevorderen van een meer rechtvaardige samenleving en meer rechtvaardig onderwijs.





## Nawoord

Hoe kennis vergaard en opgedaan wordt, dat is waar methodologie over gaat. De term 'methodologie' stamt af van het Griekse meta (na), hodos (weg), logos (leer), en betekent letterlijk 'leer van de af te leggen weg'. De af te leggen weg voor onderzoek is niet op voorhand gegeven. Het is eerder omgekeerd: het te onderzoeken object dwingt de te volgen weg af. Deze manier van kijken naar onderzoek en de te volgen methodiek geeft ruimte en lucht voor het onderwerp zelf en voor de onderzoeker. Nu het proefschrift er ligt, wil ik op deze plek nog eens reflecteren op de manier waarop het onderzoek is verlopen en de tekst tot stand is gekomen.

Hoewel veel handboeken voor het doen van onderzoeken en het schrijven van proefschriften suggereren dat onderzoek doen een gefaseerd, stapsgewijs en lineair proces is, geldt dit zeker niet voor mij en dit proefschrift. Dit onderzoek en het schrijven van dit proefschrift laten zich meer typeren als een 'odyssee', een zoektocht die me, zonder kaart, naar verre en vreemde uithoeken heeft gebracht om te leren en te begrijpen hoe de wereld die ik wilde ontdekken, er feitelijk uitzag en hoe die is ontstaan. Vanuit verwondering en verbazing over de rol en impact van de eindtoets basisonderwijs op alle niveaus van en voor alle betrokkenen bij het onderwijs, ben ik mijn zoektocht gestart, vanuit vragen om te willen weten en te willen begrijpen. 'Hoe heeft dit zo kunnen gebeuren?' was daarbij de hoofdvraag. Met de focus op het 'hoe', op 'dit', op 'zo' en op 'gebeuren'. De werking van de eindtoets zelf was mij, vanuit mijn jarenlange ervaring als toetsontwikkelaar, wel bekend maar dat gold zeker niet voor alle aspecten die in het politieke en maatschappelijke discours aan de eindtoets werden toegedicht. Ik ben dan ook begonnen met het ordenen van alle aspecten, kenmerken, argumenten, voors- en tegens van de eindtoets en in het kielzog daarvan het LVS, zoals ik die in tal van beleidsdocumenten vanaf het begin van deze eeuw ben tegengekomen. Van al die punten ben ik me gaan afvragen wat de waarde ervan was, hoe die waarde was onderbouwd, waar die waarde aan werd ontleend of op was gebaseerd en waar die vandaan kwam. En of er onderzoek was, dat hier een andere blik op kon werpen. Het is een zoektocht geworden naar de waarde en waardering van de positionering van de eindtoets en in het verlengde daarvan, het LVS, in het maatschappelijk en onderwijskundig debat.

Het maatschappelijk belang in combinatie met de inbedding van de eindtoets (en LVS) in bestuur en beleid hebben gemaakt dat het onderzoek een hoog iteratief gehalte heeft gekend, ingegeven door de ambitie maar ook wens en de behoefte om antwoorden te willen kunnen geven op de fundamentele vragen en om bestaande uitgangspunten en aannames opnieuw te kunnen bevragen. Juist omdat het veelal ging om uitgangspunten en aannames die soms al decennialang het denken over en handelen op basis van de eindtoets vormgeven en sturen heeft dit grote zorgvuldigheid gevraagd bij de uitvoering van het onderzoek. Het betekende een voortdurend proces van 'tegedenken' en 'spiegelen' om a-priori standpunten te voorkomen. Enerzijds door voldoende diepte en breedte in het bronnenonderzoek na te streven en anderzijds door het voeren van gesprekken met betrokkenen. De fundamentele aannames en uitgangspunten die ik wilde kunnen bevragen, vormden de basis voor mijn zoektocht naar bronnen, geschreven bronnen en personen uit de wereld van onderwijs en beleid. Elke bron bracht mij drie nieuwe bronnen, elke gesprekspartner bracht me drie andere gesprekspartners, én nog meer bronnen. Het zoeken naar antwoorden en het willen begrijpen voerde mij kris kras door vele disciplines in een proces, waarbij ik mijzelf ertoe bracht het chaotische ervan te verdragen. Ik wist immers dat

om de eindtoets heen een multidisciplinair web van vraagstukken was ontstaan dat ik open en onbevangen tegemoet zou moeten treden om mijn onderzoek goed te kunnen uitvoeren.

Hoe al die informatie te ordenen? Op momenten voelde dat als een uitzichtloos proces. Het ordenen moest niet alleen logisch zijn voor mezelf, alle antwoorden op al mijn vragen bevatten maar ook voor derden voldoende coherent en systematisch zijn. Het aanbrengen van een logische ordening is een iteratief proces geweest. Ik heb verschillende pogingen ondernomen. In eerste instantie koos ik voor een chronologische ordening maar dit bleek geen goede manier om te komen tot een logisch, samenhangend verhaal. Vervolgens heb ik het materiaal geordend naar mogelijke 'verklaringen' voor het gebruik, belang en waardering van de eindtoets. Ook die ordening bleek, alhoewel beter dan de chronologische, niet te kunnen leiden tot een coherente beschrijving van de onderzoeksbevindingen. Door de eindtoets te beschouwen als een instituut en een institutie en die te benaderen vanuit verschillende perspectieven, ontstond samenhang. De uiteindelijke indeling is te herleiden naar de verschillende perspectieven waarmee de eindtoets is bezien: vanuit een historisch perspectief, een sociologisch perspectief, een politiek-bestuurlijk perspectief, een psychologisch perspectief, een onderwijskundig perspectief en een psychometrisch perspectief. Met die verschillende perspectieven kon ik wat geteld en verteld wordt met de eindtoets in het maatschappelijk en onderwijsdebat duiden en van context voorzien. En ordenen. Zo zijn er negen hoofdstukken tot stand gekomen die ieder vanuit een eigen invalshoek licht werpen op het geheel van aspecten, kenmerken, argumenten, voor- en tegens van eindtoets en LVS zoals we die in het maatschappelijk en onderwijskundig debat tegenkomen. Hiermee is de cirkel rond: de verbazing en verwondering zijn 'vertaald' in begrip en duiding van de eindtoets zelf, de maatschappelijke context waarin de eindtoets is ingebed en de belangen die daarmee samenhangen en de gevolgen die de eindtoets heeft voor de leerlingen zelf, hun ouders, hun leraren, hun scholen en de samenleving als geheel.

Wat ik heb onderzocht is een fenomeen dat in een concrete maatschappelijke werkelijkheid van grote betekenis is geworden. Het betreft een instrument – de eindtoets – dat gaandeweg voor steeds meer doelstellingen is ingezet en waaraan gaandeweg steeds meer betekenissen zijn gehecht. We kunnen dus spreken van een proces van institutionalisering van een instrument waarin allerlei legitimaties zijn gebruikt om die inzet te rechtvaardigen. Vooral de wetenschappelijke en beleidsinhoudelijke legitimaties heb ik onderzocht. Dat betekent dus dat niet in den brede wetenschappelijke theorieën over toetsen en testen zijn verkend of dat er een probleemstelling is ontwikkeld op basis van deze theorieën. Wat ik heb gedaan is een toetsing van de wetenschappelijke argumenten en onderbouwingen die gebruikt zijn in het politieke, het beleidsmatige en het maatschappelijke discours over de eindtoets en zijn voorlopers. De selectie van bronnen en de selectie van wetenschappelijke kennis is dus gebaseerd op de in deze discourses naar voren gebrachte argumenten en bewijsvoering. Een gesloten aanpak van theoretische en conceptuele aard, op grond van een omvattende verkenning van de relevante literatuur zou niet passend en adequaat zijn geweest. In alle hoofdstukken heb ik duidelijk gemaakt welke ambities, doelstellingen, instrumentele verwachtingen en bewijsvoering met betrekking tot de eindtoets vanuit het perspectief van dat hoofdstuk aan de orde zijn. Op grond van geëxpliciteerde wetenschappelijke inzichten en controleerbare argumentaties heb ik deze steeds beoordeeld en met – voor zover beschikbaar – empirische kennis geconfronteerd. Die beoordeling en confrontatie schragen mijn kritische en falsifiërende uitspraken. Ze zijn daarmee controleerbaar en nodigen uit tot

discussie, zoals het in de wetenschap, en zeker ook in het publieke debat, hoort.

Ik kijk met een goed gevoel terug op de gevolgde methodiek, in het besef dat een open methodologie twee kanten heeft. Enerzijds kan een open methodologie bias voorkomen omdat er op voorhand geen wegen van onderzoeken zijn afgesloten, anderzijds ligt bias op de loer omdat de onderzoeker voortdurend keuzes maakt in de wegen die worden ingeslagen en in de manieren waarop kennis wordt verzameld. Maar ook voor een onderzoek volgens vooraf gebaande paden geldt dat bepaalde kennis daarbij buiten de boot valt en zo aan het zicht kan worden onttrokken.

Hoe dan ook, een gekozen methodologie is in hoge mate bepalend voor de kennis die een onderzoek produceert. En is derhalve altijd aan discussie onderhevig. Met dit onderzoek is beoogd een andere dimensie, een andere manier van kijken toe te voegen aan het bestaande discours over de eindtoets in het maatschappelijk en onderwijskundig debat door bestaande, vigerende aannames en uitgangspunten te onderzoeken en opnieuw voor discussie bloot te leggen. In die zin voelt het afronden van dit onderzoek en deze tekst meer als een startpunt en een nieuw begin dan als een afsluiting.

## Literatuur

- Achterhuis, H. (1988). *Het rijk van de schaarste: Van Thomas Hobbes tot Michel Foucault*. Baarn: Ambo.
- Achterhuis, H. (2010). *De utopie van de vrije markt*. Lemniscaat.
- American Educational Research Association, American Psychological Association, Joint Committee on Standards for Educational, Psychological Testing (US), & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. American Educational Research Association.
- Amsing, H. (2002). *Bakens verzetten in het voortgezet onderwijs, 1863-1920*. Groningen: Eburon.
- Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Bakker, W., & Lehning, P. (1991). *Kiezen en verdelen: Selectie van burgers als interventiestrategie*. Meppel: Boom.
- Bartels, A. (1963). *Een eeuw Middelbaar Onderwijs: 1863-1963*. Groningen: Wolters.
- Berding, J. (2016). *Ik ben ook een mens: Opvoeding en onderwijs aan de hand van Korczak, Dewey en Arendt*. Culemborg: Phronese.
- Biesta, G. (2012). *Goed onderwijs en de cultuur van het meten*. Amsterdam: Boom Lemma.
- Biesta, G. (2015). *Het prachtige risico van onderwijs*. Culemborg: Phronese.
- Biesta, G. (2016). *Het leren voorbij: Democratisch onderwijs voor een menselijke toekomst*. Culemborg: Phronese.
- Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins.
- Bisseret, N. (2014). *Education, class language and ideology* (Vol. 167). London: Routledge.
- Blok, H., Karssen, M., Meijer, J., & Roeleveld, J. (2016). Over de validiteit van indicatoren voor schooleffectiviteit. *Pedagogiek*, 36(1), 29-47.
- Bloom, B. S. (1976). *Human characteristics and school learning*. McGraw-Hill.
- Borghans, L., Schils, T., van Langen, A., & Leest, B. (2016). *De invloed van opbrengstindicatoren op het functioneren van scholen in het primair onderwijs*.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological review*, 110(2), 203.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological review*, 111(4), 1061.
- Bovens, M. A. (2006). De diplomademocratie: Over de spanning tussen meritocratie en democratie. *B en M: tijdschrift voor beleid, politiek en maatschappij*, 33(4), 205-218.
- Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT press.
- Bredo, E. (2003). The development of Dewey's psychology. In Zimmerman, B. J., & Schunk, D. H. (Eds.). (2003). *Educational psychology: A century of contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*. New York: Routledge.
- Broadfoot, P. (1992). Measuring the quality of education for parents and students. In Vedder, P. (1992). (Ed.). *Measuring the quality of education*. Swets & Zeitlinger.

Bronneman-Helmers, R. (2011). *Overheid en onderwijsbestel: Beleidsvorming rond het Nederlandse onderwijsstelsel (1990-2010)*. Den Haag: Sociaal en Cultureel Planbureau.

Bügel, K., & Sanders, P. (1998). *Richtlijnen voor de ontwikkeling van onpartijdige toetsen*. Arnhem: Cito.

Bukodi, E., & Goldthorpe, J. H. (2018). *Social mobility and education in Britain: Research, politics and policy*. Cambridge University Press.

Bulterman-Bos, J. A., & Muynck, A. (2014). *Is alles van waarde meetbaar? Toetsing en vorming in het onderwijs*. Buijten & Schipperheijn Motief.

Bureau ICE. (2015). *Toets!4*

Burt, C. (1943). Ability and income. *British Journal of Educational Psychology*, 13(2), 83-98.

Busato, V., van Essen, M., Koops, W. (2014). *Van fenomenologie naar empirisch-analytische psychologie*. Amsterdam: Prometheus.

Carver, R. P. (1975). The Coleman Report: Using inappropriately designed achievement tests. *American Educational Research Journal*, 12(1), 77-86.

Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890-1930*. New York: University Press.

Cito. (1993a). *Het Cito tussen Schiermonnikoog & Maastricht*. Arnhem: Cito.

Cito. (1993b). *Telt alleen het resultaat? Conferentiebundel 25 jaar Cito*. Arnhem: Cito.

Cito. (2019). *50 jaar Cito: Met passie voor het onderwijs van vandaag en morgen*. Arnhem: Cito.

Claassen, A., & Mulder, L. (2003). *Leerlingen na de overstap: Een vergelijking van vier cohorten leerlingen na de overgang van basisonderwijs naar voortgezet onderwijs met nadruk op de positie van doelgroepopleerlingen van het onderwijsachterstandenbeleid*. Nijmegen: ITS.

Centraal Plan Bureau. (2019). *De waarde van eindtoetsen in het primair onderwijs*. Den Haag: auteur.

Coe, R., & Fitz-Gibbon, C. T. (1998). School effectiveness research: Criticisms and recommendations. *Oxford Review of Education*, 24(4), 421-438.

Condliffe Lagemann, E. (2000). *An elusive science: The troubling history of education research*. Chicago: The University of Chicago Press.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2e ed. rev.). New York: Harper & Row.

College voor Toetsen en Examens. (2015). *Verantwoording Centrale Eindtoets PO*. Utrecht: auteur.

College voor Toetsen en Examens. (2015). *Interpretatie van het leerlingrapport bij de Centrale Eindtoets PO 2015*. Utrecht: auteur.

College voor Toetsen en Examens. (2019). *Terugblik 2019: Resultaten Centrale Eindtoets 2019*. Utrecht: auteur.

Dane, J. (2014). Een 5 voor vlijt. *Toets!3*. Culemborg: Bureau ICE.

Danziger, K. (1990). *Constructing the Subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.

de Bruijn, H. (2004). Meetbaarheid en maakbaarheid. Over de perverse gevolgen van prestatie-meting. *De Academische Boekengids*, 45-9/11.

de Groot, A. D. (1961). *Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen*. 's-Gravenhage: Mouton & Co. Derde druk, 1966.



- de Groot, A. D. (1966). *Vijven en zessen: Cijfers en beslissingen: het selectieproces in ons onderwijs*. Groningen: J.B. Wolters.
- de Groot, A. D. (1972). *Selectie voor en in het hoger onderwijs*. Den Haag: Staatsdrukkerij.
- de Groot, A. D. (1986). *Begrip van Evalueren*. 's-Gravenhage: VUGA.
- de Groot, A. D., & van Naerssen, R. F. (1969). *Studietoetsen: Construeren, afnemen, analyseren*. Den Haag: Mouton.
- de Jong, I., Van Twist, M., Bressers, D., & Schram, J. (2018). *Gevoel voor getallen*. Den Haag: NSOB.
- de Regt, A. (2004). Welkom in de ratrace. *Amsterdams Sociologisch Tijdschrift* [31-3].
- de Rooy, P. (2018). *Een geschiedenis van het onderwijs in Nederland*. Amsterdam: Wereldbibliotheek.
- de Vijlder, F. (1996). *Natiestaat en onderwijs: Een essay over de erosie van de relatie tussen westerse natiestaten en hun onderwijssystemen*. VUGA.
- de Volder, I. (2012). *Externe schoolevaluaties in Europa: Een vergelijkend onderzoek*. Antwerpen: Maklu.
- de Vries, G. C. (1993). *Het pedagogisch regiem: Groei en grenzen van de geschoolde samenleving*. Amsterdam: Meulenhoff.
- Deen, N. (1969). *Een halve eeuw onderwijsresearch in Nederland: Het Nutsseminarium voor Pedagogiek 1919-1969* (Vol. 76). Groningen: Wolters-Noordhoff.
- Dehue, T. (1990). *De regels van het vak: Nederlandse psychologen en hun methodologie 1900-1985*. Amsterdam: Van Genneep.
- Dekkers, H. (1999). *Toegankelijk en effectief onderwijs*. Nijmegen: ITS.
- Denessen, E. (2017). *Verantwoord omgaan met verschillen: Sociaal-culturele achtergronden en differentiatie in het onderwijs*. Leiden: Universiteit Leiden.
- Dewan, H. (2020). Education for all: challenges along the way. *Learning Curve*, (6), 8-11.
- Dewey, J. (1929). *The sources of a science of education*. New York: Horace Liveright.
- Didau, D. (2019). *Making Kids Cleverer: A manifesto for closing the advantage gap*. Carmarthen (UK): Crown House Publishing.
- Dijsselbloem, C. (2008). *Parlementair Onderzoek Onderwijsvernieuwingen*. 's Gravenhage: Sdu Uitgevers.
- Drenth, P. J. D., & Sijtsma, K. (1990). *Testtheorie*. Houten: Bohn Stafleu Van Loghum.
- Driessen, G. (2011). *Over-en onderadvisering bij de overgang naar het voortgezet onderwijs*. Nijmegen: ITS.
- Dronkers, J. (2007). *Ruggengraat van ongelijkheid: Beperkingen en mogelijkheden om ongelijke onderwijskansen te veranderen*. Amsterdam: Mets&Schilt uitgevers / Wiardi Beckman Stichting.
- Duyvendak, J. W., Bouw, C., Gërkhani, K., & Velthuis, O. (2013). *Sociale kaart van Nederland: Over instituties en organisaties.-2e geheel herz. dr.* Den Haag: Boom Lemma uitgevers.
- Dweck, C. (2011). *Mindset: De weg naar een succesvol leven*. Amsterdam: Uitgeverij SWP.
- Ehren, M. (2019). *Trust and accountability to improve education systems: The golden duo for education reform*. Amsterdam: Vrije Universiteit Amsterdam.
- Eidhof, B., Van Houtte, M., & Vermeulen, M. (Eds.). (2016). *Sociologen over onderwijs: Inzichten, praktijken en kritieken*. Antwerpen: Maklu.

Elffers, L. (2018). *De bijlesgeneratie: Opkomst van de onderwijscompetitie*. Amsterdam: Amsterdam University Press.

Elster, J. (1992). *Local justice: How institutions allocate scarce goods and necessary burdens*. New York: Russell Sage Foundation.

Emmelot, Yolande, Karsten, Sjoerd, Ledoux, Guuske, Vermeulen, A. (2004). *Ervaringen met het vernieuwde onderwijszoekzicht*. Amsterdam: Universiteit van Amsterdam (SCO-rapport 707, projectnummer 44413)

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). *COTAN beoordelingsstelsel voor de kwaliteit van tests (geheel herziene versie)*. Amsterdam: NIP.

Expertgroep Doorlopende Leerlijnen Taal en Rekenen. (2008). *Over de drempels met taal en rekenen: Hoofdrapport van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen* (p. 107). Enschede: Stichting Leerplan Ontwikkeling.

Feldhusen, J. (2003). Lewis M. Terman: A Pioneer in the Development of Ability Tests. In Zimmerman, B. J., & Schunk, D. H. (Eds.). (2003). *Educational psychology: A century of contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*. New York: Routledge.

Foucault, M. (1989) [1975]. *Discipline, toezicht en straf: De geboorte van een gevangenis*. Groningen: Historische Uitgeverij.

Frissen, P. (2007). *De staat van verschil: Een kritiek van de gelijkheid*. Amsterdam: Van Gennep.

Frissen, P., Steen, M. van der, Peeters, R., Frankowski, A., Jong, I. de, Chin-A-Fat, N. & Schram, J. (2015). *Sturing van onderwijskwaliteit in het primair onderwijs. Op zoek naar een balans tussen gedeelde betekenis en variëteit*. Den Haag: Nederlandse School voor Openbaar Bestuur.

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform*. East Melbourne: Centre for Strategic Education.

Gellner, E. (2008). *Nations and nationalism*. New York: Cornell University Press.

Glas, C. A. W., & Ouborg, M. J. (1993). Vraagonzuiverheid [Differential item functioning]. *Psychometrie in de praktijk*, 349-370.

Goldthorpe, J. H. (2016). Social class mobility in modern Britain: changing structure, constant process. *Journal of the British Academy*, 4(89-111).

Gould, S. J. (1996). *De mens gemeten: De geschiedenis van de intelligentietest*. Amsterdam: Contact.

Grosheide, J. (1971). Onderwijsvernieuwing in een stroomversnelling. *Nieuwe lijnen voor het onderwijs van morgen*. Ministerie van Onderwijs en Wetenschappen. Den Haag: Ando.

Haas, E. (1995). *Op de juiste plaats: De opkomst van de bedrijfs- en schoolpsychologische beroepspraktijk in Nederland* (Vol. 20). Hilversum: Uitgeverij Verloren.

Hakkenes, A., & de Wijs, A. (2012). Van Cito-toets naar brugklas en door naar diploma. *Sociaal-economische Trends*, 65-79.

Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Los Angeles: University of California Press.

Heckman, J. J. (2019). The Race between demand and supply: Tinbergen's pioneering studies of earnings inequality. *De Economist*, 167(3), 243-258.

Heijne, B. (2019). *Onbehagen: Nieuw licht op de beschaafde mens*. Amsterdam: Uitgeverij Prometheus.

Hickendorff, M. & Janssen, J. (2009). De invloed van contexten in rekenopgaven en prestaties van basisschoolleerlingen. *Panama-post*, jaargang 28-4. Arnhem: Cito.

- Hirsch, F. (1977). *Social limits to growth*. London: Routledge.
- Hofstee, W. K. B., Campbell, W. H., Eppink, A., Evers, A., Joe, R. C., Van de Koppel, J. M. H., Zweers, H., Choenni, C. E. S., & Van der Zwan, T. J. (1990a). *Toepasbaarheid van psychologische tests bij allochtonen*. Utrecht: Landelijk Bureau Racismebestrijding. LBR-serie nr. 11.
- Hofstee, W. K. B. (1990b). Toepasbaarheid van psychologische tests bij allochtonen. *De Psycholoog, juni 1990*, 291-294.
- Hood, C. (1986). *The tools of government*. Chatham, New Jersey: Chatham House.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1)
- Hoof, J. J. B. M., & Dronkers, J. (1980). *Onderwijs en arbeidsmarkt: Een verkenning van de relaties tussen onderwijs, arbeidsmarkt en arbeidssysteem*. Deventer: Van Loghum Slaterus.
- Huisman, P. (1997). *Kennis gewogen: Analyse van sociaal-wetenschappelijk denken: kritiek en aanwijzingen*.
- Idenburg, PH. (1960). *Schets van het Nederlandse Schoolwezen*. Groningen: J.B. Wolters.
- Idenburg, PH. (1970). Naar een constructieve onderwijspolitiek. *Peдагогische Studiën* (47) 1-18.
- Inspectie van het Onderwijs (2005). *Toezichtkader primair onderwijs 2005*. Utrecht: auteur.
- Inspectie van het Onderwijs (2007). *Aansluiting voortgezet onderwijs op het basisonderwijs*. Utrecht: auteur.
- Inspectie van het Onderwijs (2008). *De Staat van het Onderwijs: Onderwijsverslag 2006/2007*. Utrecht: auteur.
- Inspectie van het Onderwijs (2009). *De Staat van het Onderwijs: Onderwijsverslag 2007/2008*. Utrecht: auteur.
- Inspectie van het Onderwijs (2010). *De Staat van het Onderwijs: Onderwijsverslag 2008/2009*. Utrecht: auteur.
- Inspectie van het Onderwijs (2015). *De Staat van het Onderwijs: Onderwijsverslag 2013/2014*. Utrecht: auteur.
- Inspectie van het Onderwijs (2016). *De Staat van het Onderwijs: Onderwijsverslag 2014/2015*. Utrecht: auteur.
- Inspectie van het Onderwijs (2016). *ONDERZOEKSKADER 2017 voor het toezicht op de voor-schoolse educatie en het primair onderwijs*. Utrecht: auteur.
- Inspectie van het Onderwijs (2017). *De Staat van het Onderwijs: Onderwijsverslag 2015/2016*. Utrecht: auteur.
- Inspectie van het Onderwijs (2018). *De Staat van het Onderwijs: Onderwijsverslag 2016/2017*. Utrecht: auteur.
- Inspectie van het Onderwijs (2020a). *Onderwijsresultatenmodel PO: De inspectie en leerresultaten, het onderwijsresultatenmodel, wat heb ik hier als school of bestuur aan? en de schoolweging*. Utrecht: auteur.
- Inspectie van het Onderwijs (2020b). *De staat van het Onderwijs: Onderwijsverslag 2018/2019*. Utrecht: auteur.
- Jacobs, L. A. (2016). Dealing fairly with winners and losers in school: Reframing how to think about equality of educational opportunity 50 years after the Coleman Report. *Theory and Research in Education*, 14(3), 313-332.

Jonk, A. (2014). *Visie van de onderwijsinspectie op de Cito eindtoets*. Geraadpleegd op 26-09-2020: <https://wij-leren.nl/cito-toets-onderwijsinspectie.php>

Kangasniemi, E. (1992). National and School Level Assessment in Finland. In Vedder, P. (1992) (Ed.). *Measuring the quality of education*. Lisse: Swets & Zeitlinger.

Karabel, J. (2005). *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton*. Boston: Mariner Books.

Koenis, S. (2016). *De januskop van de democratie: Over de bronnen van boosheid in de politiek*. Amsterdam: Van Gennep.

Koretz, D. (2017). *The testing charade: Pretending to make schools better*. Chicago: University of Chicago Press.

Korsten, A. F. A. (2011). New public management. Retrieved October, 10, 2013.

Korsten, A. F., Abma, K., & van der Meer, A. D. (2019). *Het gras bij de burens: Benchmarking de maat genomen*. Den Haag: Boombestuurkunde.

Korthals, R. A. (2015). *Tracking students in secondary education: Consequences for student performance and inequality*.

KOTAN (1988). *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen*. Amsterdam: NIP.

Kouwer, B. (1957). *Tests in de psychologische praktijk*. Utrecht: Erven J. Bijleveld.

Kuiper, T. (1927). Een testonderzoek op verschillende middelbare scholen. *Paedagogische Studiën*, 8 (321-337).

Kupermintz, H. (2003). Lee J. Cronbach's Contributions to Educational Psychology. In Zimmerman, B. J., & Schunk, D. H. (Eds.). (2003). *Educational psychology: A century of contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*. New York: Routledge.

Langeveld, M. J. (1957). *Inleiding in de Studie van de Paedagogische Psychologie*. Groningen: J.B. Wolters.

Lascoumes, P., & Le Gales, P. (2007). Introduction: Understanding public policy through its instruments - From the nature of instruments to the sociology of public policy instrumentation. *Governance*. 20-1 (1-21).

Latour, B. (1988). *Wetenschap in actie*. Amsterdam: Bert Bakker.

Latour, B. (1993). *The pasteurization of France*. London: Harvard University Press.

Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken: Over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.

Lehning, P. B. (1991). *Beleid op niveau: Over de architectuur van overheidsinterventie*. Meppel: Boom.

Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. New York: Macmillan.

Leune, J. M. G. (2007). *Verstandig onderwijsbeleid*. Rotterdam: Erasmus Universiteit.

Lindemann, B. (2014). *Lost in Translation: How Public Professional Services reconfigure professional practices*.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

- Lorenz, C. (2012). If you're so smart, why are you under surveillance? Universities, neoliberalism, and new public management. *Critical inquiry*, 38(3), 599-629.
- Luyten, H., Oomens, M. & Scholten, F. (2019). *Evaluatie Wet Eindtoetsing PO: Eindrapportage*. Universiteit Twente.
- Madaus, G. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's policy. *Harvard Educational Review*, 64(1), 76-95.
- Mandemakers, K. (1996). *HBS en Gymnasium: Ontwikkeling, structuur, sociale achtergrond en schoolprestaties, Nederland, circa 1800-1968*.
- Markovits, D. (2019). *The Meritocracy Trap: How America's Foundational myth feeds inequality, dismantles the middle class and devours the elite*. New York: Penguin Press.
- Matthijssen, M. A. J. M. (1971). *Klasse-onderwijs*. Deventer: Van Loghum Slaterus.
- Matthijssen, M. (1975). Sociale ongelijkheid van onderwijskansen; vijf interpretaties van sociale werkelijkheid. *Sociologische Gids*, 22(6), 427-445.
- Matthijssen, M. (1982). *De elite en de mythe*. Deventer: Van Loghum Slaterus.
- Mayer, R. E. (2003). EL Thorndike's enduring contributions to educational psychology. In Zimmerman, B. J., & Schunk, D. H. (Eds.). (2003). *Educational psychology: A century of contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*. New York: Routledge.
- Mau. (2017). *Das metrische Wir: Über die Quantifizierung des Sozialen*. Berlin: Suhrkamp.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meijnen, G. W., & Wieringen, A. V. (2000). *Onderwijs in maatschappelijk perspectief*. Leuven: Garant.
- Meijnen, G. W. (2004). Het concept meritocratie en het voortgezet onderwijs. *Pedagogische Studiën*, 81(2), 79-88.
- Meirieu, P. (2016). *Pedagogiek: De plicht om weerstand te bieden*. Culemborg: Phronese.
- Mellink, B. (2014). *Worden zoals wij: Onderwijs en de opkomst van de geïndividualiseerde samenleving sinds 1945*. Amsterdam: Wereldbibliotheek.
- Mertens, F. (2014). *De lange weg naar een wet op het onderwijstoezicht*. Oosterwijk: Wolf Legal Publishers.
- Mertens, F. (2016). 'Vijven en Zessen' van Prof. Adriaan de Groot: Een boekje dat geschiedenis maakte. Oosterwijk: Wolf Legal Publishers.
- Mijs, J. J. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research*, 29(1), 14-34.
- Ministerie van Onderwijs en Wetenschappen (1968). Prioriteitennota. *Uitleg (134)*. Den Haag: auteur.
- Ministerie van Onderwijs en Wetenschappen (1969). *Democratisering van het onderwijs*. Zitting 1968-1969: 10 176. 's-Gravenhage.
- Ministerie van Onderwijs en Wetenschappen (1972). *Nota over het onderwijsbeleid*. 's-Gravenhage: Staatsuitgeverij.
- Ministerie van Onderwijs en Wetenschappen (1975). *Discussienota Contouren van een toekomstig onderwijsbestel*. Zitting 1974-1975: 13 459 nrs. 1-2. 's-Gravenhage: Staatsuitgeverij.

Ministerie Onderwijs en Wetenschap, (1993a). *Niet zomaar gekozen: Over kerntaken, leerwegen en overheidsgeld*. Den Haag: auteur.

Ministerie Onderwijs en Wetenschap (1993b). *Kwaliteitszorg in primair en voortgezet onderwijs*. Den Haag: auteur.

Ministerie Onderwijs, Cultuur en Wetenschap (1995). *De school als lerende organisatie: Kwaliteitsbeleid op scholen voor primair en voortgezet onderwijs*. Den Haag: auteur.

Ministerie van Onderwijs, Cultuur en Wetenschap (2001). *Grenzeloos Leren*. Den Haag: auteur.

Ministerie Onderwijs, Cultuur en Wetenschap (2004). *Koers Primair Onderwijs: Ruimte voor de school*. Den Haag: auteur.

Ministerie van Onderwijs Cultuur en Wetenschap (2010). Invoering referentieniveaus taal en rekenen. Brief van Minister van Onderwijs, Cultuur en Wetenschap aan scholen in Nederland, 7 juli 2010. 220779.

Ministerie van Onderwijs Cultuur en Wetenschap (2011a). Actieplan 'Basis voor presteren'. Brief van Minister van OCW aan Voorzitter van de Tweede Kamer, 23 mei 2011. 303172.

Ministerie van Onderwijs Cultuur en Wetenschap (2011b). Toetsing in het primair onderwijs. Brief van Minister van OCW aan Voorzitter van de Tweede Kamer, 1 maart 2011. Kamerstukken II, 31293, 89.

Ministerie van Onderwijs, Cultuur en Wetenschappen (2012). Memorie van toelichting wijziging wet PO, 31 januari 2012. Kamerstukken II, 33157, 3.

Ministerie van Onderwijs, Cultuur en Wetenschap (2016). Beleidsreactie Onderwijsverslag 2014-2015, 14 april 2016. Kamerstukken VIII, 34300, 140.

Ministerie van Onderwijs, Cultuur en Wetenschap (2017). Resultaten eindtoets 2016-2017. Brief van Minister OCW aan Voorzitter van de Tweede Kamer, 24 november 2017, Kamerstukken II, 31293, 380.

Ministerie van Onderwijs, Cultuur en Wetenschap (2018). Kamerbrief over schooladviezen en resultaten op eindtoets in schooljaar 2017-2018. Brief van Minister OCW aan Voorzitter van de Tweede Kamer, 29 november 2018, referentienummer 1416959.

Ministerie van Onderwijs, Cultuur en Wetenschap (2019a). Kamerbrief over bevordering kansgelijkheid in het onderwijs. Brief van Minister OCW aan Voorzitter van de Tweede Kamer, 13 maart 2019, referentienummer 1474648.

Ministerie van Onderwijs, Cultuur en Wetenschap (2019b). Kamerbrief over ONDERZOEKSKADER 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs. Brief van Minister OCW aan Voorzitter van de Tweede Kamer, 7 juni 2019, referentienummer 8544170.

Ministerie van Onderwijs, Cultuur en Wetenschap, (2019c). Eindevaluatie Wet eindtoetsing po. Brief van de minister OCW aan Voorzitter van de Tweede Kamer, 21 juni 2019, Kamerstukken II, 31293, 477.

Ministerie Onderwijs, Cultuur en Wetenschap (2020). *De overstap van po naar vo: Handreiking schooladvisering*. Den Haag: auteur.

Myrdal, G. (1969). *Objectivity in social research*. London: Gerald Duckworth & Co LTD.

Nahuis, R., & Verbruggen, J. (2009). *Wat is bekend over de effecten van kenmerken van onderwijsstelsels? Een literatuurstudie*. Den Haag: Centraal Plan Bureau.

Nelissen, J. (2014). De 'aaneensluiting': De psychologie van het leren denken ten tijde van Kohnstamm. *Reken en wiskundeonderwijs*, jaargang 33, 18–26.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, Massachusetts: Harvard Education Press.

- Noordegraaf, M. (2004). *Management in het publieke domein: Issues, instituties en instrumenten*. Bussum: Coutinho.
- Nussbaum, M. (2011). *Niet voor de winst: Waarom de democratie de geesteswetenschappen nodig heeft*. Amsterdam: Ambo| Anthos.
- Nussbaum, M. (2012). *Mogelijkheden scheppen*. Amsterdam: Ambo| Anthos.
- Oberon, (2013). *Voorspellende waarde van LOVS toetsen voor schoolsucces*. Utrecht: auteur.
- OECD. (2016). *Reviews of National Policies for Education Netherlands 2016: Foundations for the future*. Parijs: OECD Publishing.
- Onderwijsraad (2004). *Werken aan gezamenlijke doelen*. Den Haag: auteur.
- Onderwijsraad (2006). *Doortastend onderwijstoezicht*. Den Haag: auteur.
- Onderwijsraad (2006). *Versteving van kennis in het onderwijs*. Den Haag: auteur.
- Onderwijsraad. (2010). *Vroeg of laat: Advies over de vroege selectie in het Nederlandse onderwijs*. Onderwijsraad. Den Haag: auteur.
- Onderwijsraad. (2011) *Toetsing in het primair onderwijs*. Den Haag: auteur.
- Onderwijsraad. (2013). *Een smalle kijk op onderwijskwaliteit*. Den Haag: auteur.
- Onderwijsraad. (2016). *De volle breedte van onderwijskwaliteit: Van smal beoordelen naar breed verantwoorden*. Den Haag: auteur.
- Onderwijsraad. (2019). *Doorgeschoten differentiatie in het onderwijsstelsel*. Den Haag: auteur.
- Osborne, D., & Gaebler, T. (1993). *Reinventing government: The five strategies for reinventing government*. New York: Penguin.
- Papadopoulos, G. (1994). *Education 1960-1990: The OECD Perspective*. Parijs: OECD.
- Pearl, J., & Mackenzie, D. (2019). *Het boek waarom: De nieuwe wetenschap van oorzaak en gevolg*. Amsterdam: Maven Publishing.
- Peeters, R., Hofman, R. H., & Frissen, P. H. A. (2013). *Het ongemak van autonomie: Onderwijsbeleid tussen vrijheid en verantwoording*. Den Haag: Nederlandse School voor Openbaar Bestuur.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. Oxford University Press, USA.
- Pols, W. (2016). *In de wereld komen: Een studie naar de pedagogische betekenissen van opvoeding, onderwijs en het leraarschap*. Antwerpen: Garant.
- Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational leadership*, 56, 8-16.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Boston: Pearson College Division.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, Virginia: ASCD.
- PO-raad (2015). *Handreiking schooladvies*. <https://www.poraad.nl/nieuws-en-achtergronden/handreiking-schooladvies>.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, New Jersey: Princeton University Press.
- Posthumus, K. (1940). Middelbaar onderwijs en schifting. *De Gids*, jaargang 104: 24-42

- Pouwels, J., & Jungbluth, R. (1991). *Scholen toetsen op kwaliteit: Studie ten behoeve van het WRR-advies inzake onderwijsverzorging*. Nijmegen: ITS.
- Prak, J. L. (1952). *Tests op school*. Groningen: JB Wolters.
- Putnam, R. D. (2016). *Our kids: The American dream in crisis*. New York: Simon and Schuster.
- Quené, M. (2018). *Voorbij de managementmaatschappij: De invloed van management op werk, democratie en vrijheid*. Rotterdam: Lemniscaat.
- Rancière, J. (2007). *De onwetende meester: Vijf lessen over intellectuele emancipatie*. Leuven: Acco.
- Ravitch, D. (2016) [2010]. *The death and life of the great American schoolsystem: How testing and choice are undermining education*. New York: Basic Books.
- Resh, N., & Sabbagh, C. (2016). Justice and education. In Sabbagh, C. & Schmitt, M. (Eds.) (2016). *Handbook of social justice theory and research* (pp. 349-367). New York: Springer.
- Rist, R. C. (2000). Self Fulfilling Prophecy. *Harvard Educational Review*, 70-3 (131-174).
- Roeleveld, J., Driessen, G., Ledoux, G., & Meijer, J. (2011). *Doelgroepopleerlingen in het basisonderwijs: Historische ontwikkeling en actuele situatie*. Amsterdam: Kohnstamm Instituut. (Rapport 857, projectnummer 20432).
- Rose, T. (2016). *The end of average: How to succeed in a world that values sameness*. London: Penguin UK.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The urban review*, 3(1), 16-20.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York: Teachers College Press.
- Sabbagh, C., & Schmitt, M. (Eds.). (2016). *Handbook of social justice theory and research* (pp. 10-14). New York: Springer.
- Sahlberg, P., Heijmans, J. W., & de Groot, E. (ed. 2016). *Finnish lessons: wat Nederland kan leren van het Finse onderwijs*. Helmond: OMJS.
- Samenwerkende Instituten (1967). *Amsterdamse Schooltoetsen: Verslag van het eerste onderzoek 'L.O.-Schooltoets Amsterdam, 1966'*. Groningen: J.B. Wolters.
- Sandel, M. (2010). *Rechtvaardigheid: Wat is de juiste keuze?* Utrecht: Uitgeverij Ten Have.
- Sandel, M. (2020). *The tyranny of merit: What's become of the common good?* London: Penguin, UK.
- Sanders, P. (red.) (2017a). [2012] *Toetsen op school: Voortgezet Onderwijs*. Arnhem: Cito.
- Sanders, P. (red.) (2017b). *Toetsen op school: Primair Onderwijs*. Arnhem: Cito.
- Schilt-Mol, T. M. M. L. (2007). *Differential item functioning en itembias in de Cito-Eindtoets Basis-onderwijs*.
- Schüssler, E. (red.) (2006). *Weg van de middenschool: Dertig jaar na de start van het middenschool-experiment*. Antwerpen: Garant.
- Schuyt, C.J.M. (1991). *Op zoek naar het hart van de verzorgingsstaat*. Leiden: Stenfert Kroese.
- Schuyt, K. (2009). *Steunberen van de samenleving: Sociologische essays*. Amsterdam: Amsterdam University Press.
- Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven: Yale University Press.



- Sen, A. (2013). *Het idee van rechtvaardigheid*. Rotterdam: Lemniscaat.
- Sharp, C., Hutchison, D., & Whetton, C. (1994). How do season of birth and length of schooling affect children's attainment at key stage 1? *Educational Research*, 36(2), 107-121.
- Shepard, L. A. (1982). Chapter 1: Definitions of Bias. In Berk, R. A. (Ed.). (1982). *Handbook of methods for detecting test bias*. Baltimore: Johns Hopkins University Press.
- Shepard, L. A. (1993). Chapter 9: Evaluating Test Validity. *Review of Research in Education*, 19(1), 405-450.
- Smits, A. (2020). *Onderwijsinnovatie en ICT*. Zwolle: Hogeschool Windesheim.
- Staatscourant (2018). Nr. 39602. 17 juli 2018.
- Standaert, R. (2001). *Inspectorates of education in Europe: A critical analysis*. Leuven: Acco.
- Standaert, R. (2007). *Vergelijken van onderwijssystemen*. Leuven: Acco.
- Standaert, R. (2014). *De becijferde school: Meetcultus en meetcultuur*. Leuven: Acco.
- Standaert, R. (2016). De queeste naar een gedragen curriculum. In De Man, L. & Denys, K. (Eds) (2016). *De toekomst van de eindtermen: Van onderwijsbeleid tot klaspraktijk*. Brussel: Politeia (35-64).
- Standaert, R. (2017). Pisa en zijn verregeende ambities. *Tijdschrift Voor Onderwijsrecht En Onderwijsbeleid (TORB)*, 3, 74-84.
- Stellwag, H. (1955). *Selectie en selectiemethoden: Een inleidende studie in het aansluitingsvraagstuk L.O. en V.H.M.O.* Groningen: J.B. Wolters.
- Sternberg, R. J. (2003). *Succesvolle intelligentie: Hoe praktische en creatieve intelligentie succes bepalen*. Lisse: Swets & Zeilinger Publishers.
- Sternberg, R. J., & Jarvin, L. (2003). Alfred Binet's contributions as a paradigm for impact in psychology. In Zimmerman, B. J., & Schunk, D. H. (Eds.). (2003). *Educational psychology: A century of contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*. New York: Routledge. American Psychological Association.
- Stone, D. (2012)[1998]. *Policy paradox: The art of political decision making*. (3<sup>rd</sup> ed.). New York: Norton & Company.
- Swierstra, T., & Tonkens, E. (Eds.). (2008). *De beste de baas? Verdienste, respect en solidariteit in een meritocratie*. Amsterdam: Amsterdam University Press.
- Terwel, J. (2006). Terwel, J. (2006). *Is de school een sorteermachine? Schoolkeuze en schoolloopbaan van leerlingen van 10-16 jaar*. Amsterdam: Vrije Universiteit Amsterdam.
- Thijssen, T. (1926)[ed 2007]. *De gelukkige klas*. Amsterdam: Stichting CPNB.
- Timmer, W. (2011). *Het doel wel gesteld: Een praktijkonderzoek naar de toepassing van doelregelgeving*.
- Timmermans, A., Kuyper, H., & van der Werf, G. (2012). *Schooladviezen en onderwijsloopbanen. Voorkomen, risicofactoren en gevolgen van onder- en overadvisering*. Groningen: Gronings Instituut voor Onderzoek van Onderwijs.
- Tjepkema, S. (2002). Bouwen aan een lerende organisatie via het opsporen van leerblokkades. *Handboek Schoolorganisaties en Onderwijs Management*, 1-25. Enschede: Universiteit Twente, Toegepaste Onderwijskunde.
- Tonkens, E. H. (2008). *Mondige burgers, getemde professionals: Marktwerking en professionaliteit in de publieke sector*. Amsterdam: Van Gennep.

Turner, R. H. (1960). Sponsored and contest mobility and the school system. *American sociological review*, 855-867.

Uiterwijk, H. (1994). *De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen* [The usability of the Primary Education Final Test for nonnative students]. Arnhem: Cito.

Uiterwijk, J.H. & Theunissen, T. (1996). *Verantwoording Eindtoets Basisonderwijs 1992*. Arnhem: Cito.

Ultee, W. C. (1980). Is education a positional good? An empirical examination of alternative hypotheses on the connection between education and occupational level. *Netherlands (The) Journal of Sociology and Sociologia Neerlandica Amsterdam*, 16(2), 135-153.

van Boxtel, H. & Hemker, B. (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.

van Boxtel, H., Engelen, R. & de Wijs, A. (2011). *Wetenschappelijke verantwoording van de Eindtoets Basisonderwijs 2010*. Arnhem: Cito.

van Dam, M. (2009). *Niemandsland: Biografie van een ideaal*. Amsterdam: De Bezige Bij.

van de Walle, S. (2011). *New Public Management: Restoring the Public Trust Through Creating Distrust?* DOI: 10.2139/ssrn.1594042.

van de Werfhorst, H. G. (2019). Early tracking and social inequality in educational attainment: Educational reforms in 21 european countries. *American Journal of Education*, 126(1), 65-99.

van den Brink, W. P., & Mellenbergh, G. J. (2006). *Testleer en testconstructie*. Amsterdam: Boom.

van der Linden, W. J. (1987). *Het zwalkende niveau van ons onderwijs*. Enschede: Universiteit Twente.

van der Schoot, F. (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPO*. Arnhem: Cito.

van Dyck, M. (Red.) (1997). *Toegankelijkheid van het Nederlandse onderwijs*. Den Haag: Onderwijsraad.

van Gunsteren, H. R. (1984). Wie in zijn graf ligt, maakt geen fouten meer: Een interventieeler te ontwikkelen door ambtenaren. *Beleid En Maatschappij*, 11(6), 159-163.

van Heek, F. e.a. (1968). *Het verborgen talent: Milieu, schoolkeuze en schoolgeschiktheid*. Meppel: Boom.

van Kemenade, J. A., & Kropman, J. A. (1972). Verborgen talenten? Kritische kanttekeningen bij een onjuiste interpretatie. *Sociologische Gids*, 19(4), 219-228.

van Kemenade, J. A. (1981). [red.] *Onderwijs: bestel en beleid*. Groningen: Wolters-Noordhoff.

van Pinxteren, M., & de Beer, P. (2016). *Meritocratie: Op weg naar een nieuwe klassensamenleving* (p. 254). Amsterdam: Amsterdam University Press.

van Rooijen, M., Korpershoek, H., Vugteveen, J., & Opdenakker, M. C. (2017). De overgang van het basis- naar het voortgezet onderwijs en de verdere schoolloopbaan. *Pedagogische Studiën*, 2017 (94) 110-134.

van Strien, P. (2003). De opvoedbaarheid van de intelligentie: Een oud strijdpunt tussen pedagogen en psychologen. *Pedagogiek (Assen)*, 23(2), 122-136.

van Wieringen, A. M. (1996). *Onderwijsbeleid in Nederland*. Alphen aan den Rijn: Samsom HD Tjeenk Willink.

van Wieringen, A. (2000). Onderwijs: Beleid en maatschappij. In: Meijnen, G. & van Wieringen, A. (red.). *Onderwijs in maatschappelijk perspectief*. Leuven: Garant.

- van Yperen, T. A., Bleichrodt, N. & van de Vijver, F. (Red.) (2001). *Diagnostiek bij allochtonen: Mogelijkheden en beperkingen van psychologische tests*. Lisse: Swets & Zeitlinger.
- Vedder, P. (1992). (ed.). *Measuring the quality of education*. Lisse: Swets & Zeitlinger.
- Veldhuijzen, N. H., Goldebeld, P., & Sanders, P. F. (1993). Klassieke testtheorie en generaliseerbaarheidstheorie [Classical test theory and generalizability theory]. *Psychometrie in de praktijk [Psychometrics in practice]*, 33-82.
- Veldhuis, J. (1983). *Een opzienbarend Amerikaans rapport over de kwaliteit van het onderwijs in de Verenigde Staten van Noord-Amerika*. Den Haag.
- Verhaeghe, P. (2012). *Identiteit*. Amsterdam: De Bezige Bij.
- Verhaeghe, P. (2019). *Over normaliteit en andere afwijkingen*. Amsterdam: Prometheus.
- Verseput, J. (1956). *Gedenkboek 1916-1956*. Groningen: J.B. Wolters.
- Visscher, A. J., & Ehren, M. C. M. (2011). *De eenvoud en complexiteit van Opbrengstgericht Werken*. Enschede: Universiteit Twente, Vakgroep Onderwijsorganisatie en -management.
- Vosselman, E. G. J. (2015). De wankeling van een besturingsparadigma: een alternatief perspectief op controllership in de publieke sector. *Tijdschrift voor public governance audit & control*, 13(5), 4-9.
- Vroon, P. A. (1980). *Intelligentie: Over het meten van een mythe en de politieke, sociale en onderwijskundige gevolgen*. Baarn: Ambo.
- Vuyk, K. (2017). *Oude en nieuwe ongelijkheid*. Utrecht: Klement.
- Vuyk, K. (2019). *De feilbare mens: Waarom ongelijkheid zo slecht nog niet is*. Utrecht: Uitgeverij Ten Have.
- Wansink, H. (1992). *Een school om te kiezen: Naar een actuele onderwijspolitiek*. Amsterdam: Bert Bakker.
- Warries, E. (1971) Drie redenen om te toetsen in het onderwijs. *Pedagogische Studiën*, jaargang 48, 1971, 152-161.
- Waslander, S. (1999). *Koopmanschap en burgerschap: Marktwerking in het onderwijs*.
- Wesdorp, A., Bergh, H., Bos, D. J., Hoeksma, J. B., Oostdam, R. J., Scheerens, J., & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek: Een voorstudie op het gebied van het taalonderwijs in de lagere school*. Lisse: Swets & Zeitlinger.
- Wijnstra, J.M. (1984). *Verantwoording Eindtoets Basisonderwijs 1981 (met een historisch overzicht over de periode 1966-1980)*. Arnhem: Cito.
- Wilbrink, B. (1986). Toetsen en testen in het onderwijs. *S.V.O. Jaarverslag/Jaarboek 1985*, 275-288. Den Haag: Stichting voor Onderwijsonderzoek.
- Wilbrink, B. (1995). *Kwalificaties van jongeren: risico's of kansen? Toekomstmogelijkheden*.
- Wilbrink, B. (1996). *Leren Waarderen: de geschiedenis*.
- Wilbrink, B. (1997). Terugblik op toegankelijkheid: meritocratie in perspectief. In M. van Dyck (Red.) *Toegankelijkheid van het Nederlandse onderwijs*. *Studies*, 341-384.
- Wiliam, D. (2013). *Cijfers geven werkt niet*. Meppel: Ten Brink Uitgevers.
- Wilson, N. (1998). Educational standards and the problem of error. *Education Policy Analysis Archives*, 6.

Wools, S. (2015). *All about validity: An evaluation system for the quality of educational assessment*. Enschede: Universiteit Twente.

WRR. (1986). *Basisvorming in het onderwijs*. Den Haag: auteur.

Young, M. (ed. 1994). *The rise of the meritocracy*. New Brunswick, New Jersey: Transaction Publishers.

Young, I. M. (2011). *Justice and the Politics of Difference*. Princeton, New Jersey: Princeton University Press.

## Summary

### **The cat and the bell, Counting and telling with the primary school end test**

The Dutch educational system is classified by the Organisation for Economic Co-operation and Development (OECD) as an 'early tracking' system. Meaning that eleven/twelve year old children are subdivided into different tracks of secondary education. There are four vocational tracks, one mixed track and two general tracks, all with a different level and duration. In the last year of primary education all Dutch children must make a test which is of high importance for the decision of which track they will follow. The test is considered to be a form of an intelligence test. The results on this test are used to predict the track, and its corresponding level, of secondary education that is thought to be most suitable for the child. Fifty percent of the children are advised to follow one of the vocational tracks; the other fifty percent are directed towards either the mixed or general education. Only twenty percent of all pupils has direct access to the highest (vwo level), which upon successful completion (after six years) directly grants access to university. So the primary school end test has a major impact on the future of a child and therefore the stakes are high. This test is perceived as very stressful for the children, their parents and the primary schools. Schools and parents are for this reason willing to spend a lot of energy, time and even money to prepare their children as good as possible for this test.

The Ministry of Education is responsible for the primary school end test. Each year the content of the test is renewed and kept secret until the date of the exam. The test takes two days and contains items regarding reading comprehension, grammar, vocabulary and mathematics. The construction of the test is organised in a way that a normal distribution of scores is realized. The test scores are used to rank children based on their scores from low to high with a scale from 501-550. This is created by a linear transformation in order to create a scale with scores that are comparable to scores in previous years.

The primary school end test is a norm referenced test. The scores of the children are compared to each other and ranked from high to low in a way that the highest twenty percent scores are advised to follow the highest track, the general educational track of vwo. It should be noted that the test is not only used as an advisory tool for the educational future of children, but is also used by the Inspectorate of Education as a quality indicator of primary education in general and of an individual school. The overall test scores are, again, used to compare. The average scores per school are compared, to some amount corrected for the social economic status and other background variables of their children. So, the schools too are ranked and listed from high to low based on their average scores. Schools that get scores that are considered too low given their population, can, in a worst case scenario, be closed by the Minister of Education. The schools are obliged to publish their average scores and parents can use the results when they are choosing a primary school for their child.

This primary school end test already exists for over fifty years and has become a powerful instrument, not only for the children and their parents but for the schools and their teachers as well. This high-stakes test is not only rooted in the history of the Dutch educational system but also in the Dutch society itself. In this thesis the history of the primary end school test, its use, and its consequences are described. In addition, the long-term consequences are compared to

the claims of 'equal opportunity', 'fairness of allocation' and 'meritocratic principles' which are advocated by the initiation of the test and its validity is questioned from different perspectives. In conclusion, the validity of the primary end school test cannot be proved nor supported from any of the chosen perspectives. The test cannot be seen as an instrument offering opportunities to children struggling with adversity but the test appears to be an important instrument in consolidating the existing class society. The test is, above all, a political instrument that serves mainly political and economic purposes.

## Over de auteur

Karen Heij (1963) heeft zich vanaf de afronding van haar studie Toegepaste Taalwetenschappen aan de katholieke universiteit Nijmegen in 1988, beziggehouden met taal en toetsing. Vanaf 1989 heeft zij gewerkt bij Bureau ICE, waarvan zij directeur/eigenaar is geweest van 2002 tot en met 2016. Zij heeft aan de basis gestaan van de ontwikkeling van toetsen voor Nederlands als Tweede Taal, zoals het Staatsexamen Nederlands als Tweede Taal en de diverse generaties inburgeringsexamens, in eerste instantie gebaseerd op de NT2-niveaus. Later heeft zij meegewerkt aan de totstandkoming van het Raamwerk Nederlands als tweede taal. Zij was gesprekspartner voor diverse ministers en ministeries als ook voor het College voor Toetsen en Examens (CvTE) inzake taaltoetsing. Met Bureau ICE heeft zij ook gewerkt aan de ontwikkeling van toetsen voor moderne vreemde talen (MVT) gerelateerd aan het Common European Framework of References for Languages (CEF) en toetsen van taal en rekenen gerelateerd aan de Referentieniveaus, met name voor (v)mbo. In het middelbaar onderwijs heeft zij in opdracht van en samen met diverse mbo-opleidingen nieuwe vormen van examinering ontwikkeld en mbo-instellingen begeleid bij het opzetten van kwalitatief hoogwaardige vormen van examinering en verantwoording daarvan. Vanaf 2012 stond zij aan de basis van de ontwikkeling van de eindtoets voor het basisonderwijs van Bureau ICE, de IEP, één van de eindtoetsen die naast de Centrale Eindtoets, goedgekeurd is als eindtoets voor het basisonderwijs.

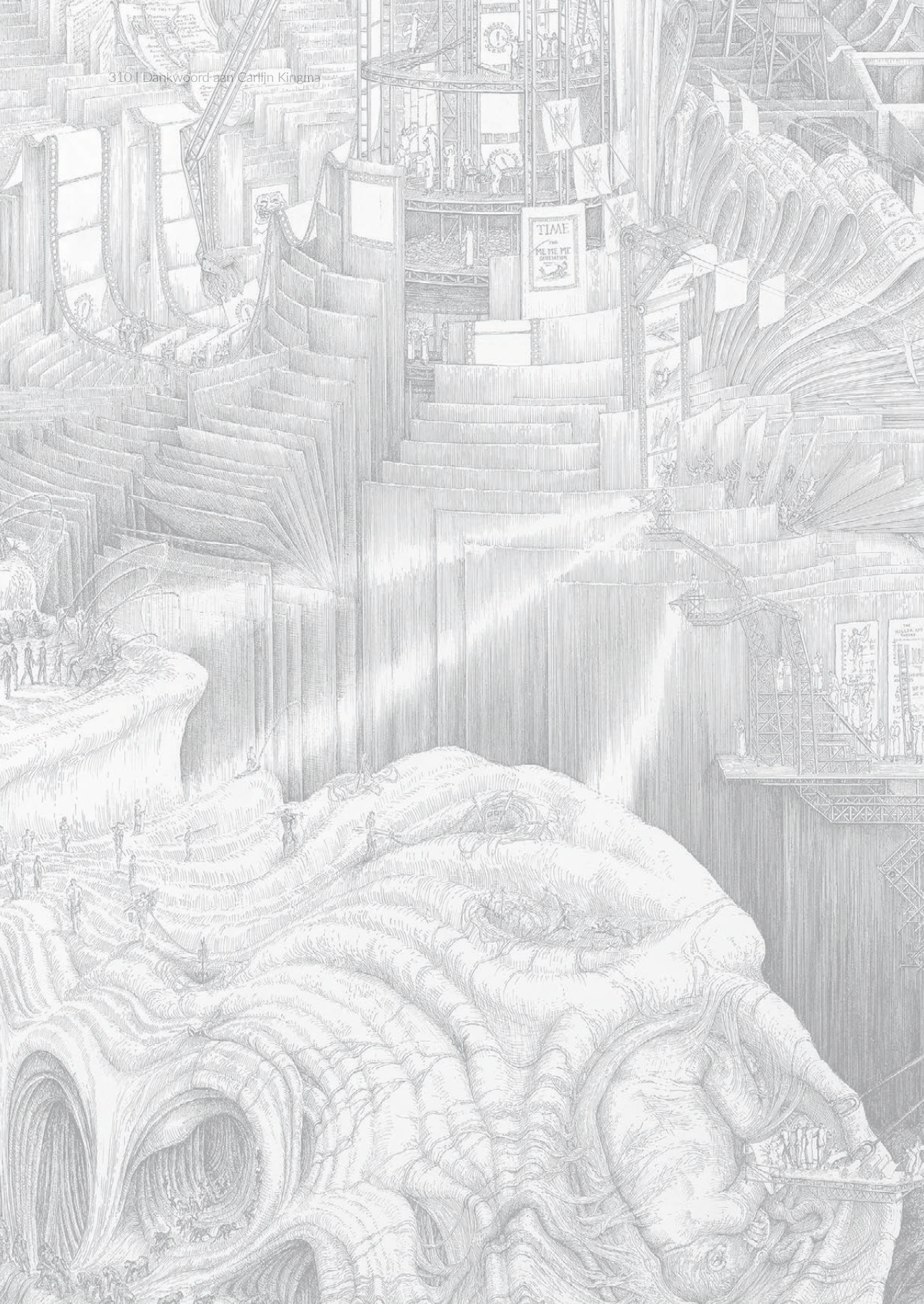


Foto: Bas Uijlings

Taal en toetsing vormen de rode draad van haar werk: Taal als object van toetsing en taal als middel waarmee toetsing gestalte krijgt. Zij was mede auteur van de publicatie 'Schrijven in eenvoudig Nederlands' en 'De Grondwet in eenvoudig Nederlands'. Zij schreef diverse publicaties over taal en taalbeleid (voor vakdocenten, voor taaldocenten en voor schoolleidingen). Over taal en taalbeleid adviseerde zij één van de schoolbesturen op Curaçao in de periode 2011-2019.

Op dit moment werkt zij als zelfstandig toetsexpert bij Parrhesia Onderwijsadvies. In die hoedanigheid verzorgt zij onder andere gastcolleges voor de master Taaldiagnostiek aan de Universiteit Utrecht, lezingen voor scholen voor basis- en voortgezet onderwijs over toetsing in het algemeen en over de basiseindtoets in het bijzonder en ondersteunt zij scholen bij vraagstukken van toetsing en examinering.

Haar promotieonderzoek voerde zij uit als buitenpromovendus bij de Nederlandse School voor Openbaar Bestuur in de periode van december 2017 tot december 2020.





## Het weefsel der mensheid

In dit proefschrift heb ik dankbaar gebruik gemaakt van het kunstwerk 'Het weefsel der mensheid' van cartograaf Carlijn Kingma. In haar werk probeert zij samenlevingen van vroeger en nu te verbeelden, om tot nieuwe vergezichten te komen. Het weefsel der mensheid maakte zij met het journalistieke platform De Correspondent en met historicus Rutger Bregman, schrijver van het boek 'De meeste mensen deugen'.

Het is geen traditionele kaart. Je ziet geen landen en wegen, bergen en rivieren. Het is een kaart met metaforen. Een kaart vol verbeeldingen van ideeën die, als je ze naast elkaar ziet, volgens Kingma, nieuwe inzichten kunnen opleveren. Zij stelt dat het slechts een kaart is van een geschiedenis – die ook nog eens op talloze manieren geïnterpreteerd kan worden. Ze heeft de kaart opgedeeld in hoofdstukken met elk een eigen thema dat in dat deel van de kaart verder is uitgewerkt in subthema's. Zij heeft hierop een uitgebreide toelichting gegeven, ook in de vorm van een audiotour, die respectievelijk te lezen en te beluisteren is via haar website of via de website van de correspondent.

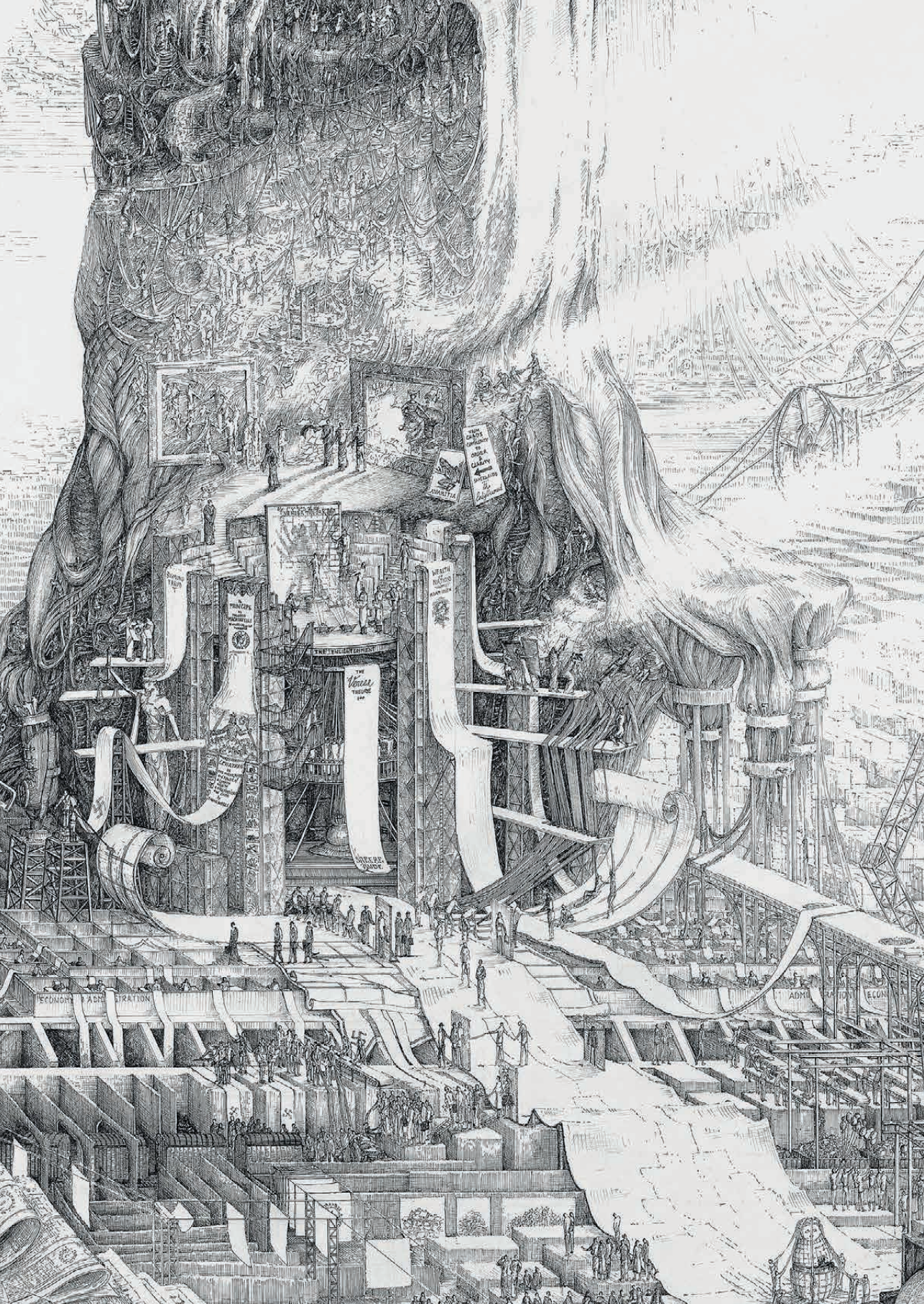
Voor de verschillende hoofdstukken in dit boek heb ik delen van haar kaart uitgelicht die in mijn optiek goed aansluiten op de onderwerpen zoals die centraal staan in de verschillende hoofdstukken van dit boek. Op de voorkant van de kaart is de toren van de macht te zien zoals Kingma die verbeeldt. Die toren is verbonden met de verbeelding van Kingma van de hedendaagse wetenschap, en vormt de achterkant van de kaart.



Voor de verschillende hoofdstukken van dit boek zijn andere uitsneden te zien van het weefsel der mensheid. Voor het eerste hoofdstuk is gekozen voor een nadere beschouwing van de top van de 'toren van de macht' (1). Voor hoofdstuk 2 is gekozen voor het deel van de kaart dat Kingma aanduidt als 'in de greep van de staat' (2). Voor hoofdstuk 3 is gekozen voor het deel van de kaart dat 'Utopia' weergeeft (3). Bij hoofdstuk 4 is het deel van de kaart te zien dat gaat over het 'gehokte leven' (4). Hoofdstuk 5 wordt geïllustreerd door het deel van de tekening dat het 'cabinet of rules, guidelines & measures' laat zien (5). Voor hoofdstuk 6 is gekozen voor het deel van de kaart dat betrekking heeft op 'meritocracy' (6). Het thema 'het sociale weefsel' en dan in het bijzonder de 'wereld zonder muren' is gekozen voor hoofdstuk 7 (7). De toren van de wetenschap, met de uil als symbool, is gebruikt als illustratie bij hoofdstuk 8 (8). Voor het laatste, afsluitende hoofdstuk, is een deel gekozen van de kaart dat betrekking heeft op de toekomst. Hier is gekozen voor de handen die weefsel dragen (the cat's cradle) waar onderwijs een prominente plek heeft (9).

Ik hoop van harte dat het inzicht geven in en delen van dit werk ook anderen inspireert tot nadenken over de verbondenheid van de mensheid en hoe daaraan bij te dragen.





THE GREAT  
GALLERY  
OF  
ARTS  
AND  
SCIENCE



THE GREAT  
GALLERY  
OF  
ARTS  
AND  
SCIENCE

THE  
TREASURY  
OF  
ARTS

THE  
GALLERY  
OF  
ARTS  
AND  
SCIENCE

THE  
GALLERY  
OF  
ARTS  
AND  
SCIENCE

EDUCATION & ADMINISTRATION

ADMINISTRATION & ECONOMY

