

# De pabo- toelatingstoetsen getoetst ▲

Op zoek naar onbedoelde moeilijkheden  
voor aspirant-pabostudenten met een  
niet-westerse migratieachtergrond

ONDERZOEK IN OPDRACHT VAN HET MINISTERIE VAN OCW

Anne Kerkhoff, Gijs van Gaans,  
Job Noordeloos en Tom Wils  
Fontys Lerarenopleiding Tilburg

met medewerking van

Silvia Matijssen en Bas Hemker  
Cito

20 mei 2021  
Fontys Lerarenopleiding Tilburg



LERARENOPLEIDING  
TILBURG



# Colofon

**De pabo-toelatingstoetsen getoetst. Op zoek naar onbedoelde moeilijkheden voor aspirant-pabostudenten met een niet-westerse migratieachtergrond.**

*Auteurs:* Anne Kerkhoff, Gijs van Gaans, Job Noordeloos en Tom Wils (Fontys Lerarenopleiding Tilburg) met medewerking van Silvia Matijssen en Bas Hemker (Cito)

*Redactie:* Anne Kerkhoff

*Opdrachtgever:* Vereniging Hogescholen namens het Ministerie van Onderwijs, Cultuur en Wetenschap

*Opmaak:* Happy Cactus, Eindhoven

*Informatie:* Vereniging Hogescholen/Fontys Lerarenopleiding Tilburg

*Datum:* 20 mei 2021

Deze publicatie is beschikbaar onder de Creative Commons Licentie  
<http://creativecommons.org/licenses/by-nd/4.0/>



Aanbevolen citatie bij hergebruik: Kerkhoff, A., Van Gaans, G., Noordeloos, J. & Wils, T. (2021). De pabo-toelatingstoetsen getoetst. Op zoek naar onbedoelde moeilijkheden voor aspirant-pabostudenten met een niet-westerse migratieachtergrond. Fontys Lerarenopleiding Tilburg, Tilburg



# Inhoud

<b>Managementsamenvatting</b> .....	<b>9</b>
<b>1. Inleiding: achtergronden en leeswijzer</b> .....	<b>13</b>
1.1 Achtergronden.....	13
1.2 De onderzoeksopdracht .....	14
1.3 Leeswijzer .....	14
1.4 Voordat u verder leest .....	15
<b>2. De toelatingstoetsen voor de pabo</b> .....	<b>17</b>
2.1 Achtergronden en doelen .....	17
2.2 De inhoud van de toelatingstoetsen.....	17
2.3 De opzet en de ontwikkeling van de toelatingstoetsen .....	18
2.3.1 Constructiegroep met vakdocenten .....	18
2.3.2 Vaststellingscommissie met vakexperts .....	18
2.3.3 Bepaling cesuur door standaardsetting met pabo-, vo- en mbo-vakdocenten.....	18
2.4 De afname van de toelatingstoetsen .....	19
<b>3. Een nadere verkenning van de onderzoeksvraag</b> .....	<b>21</b>
3.1 Wat betekent 'bias'? .....	21
3.2 Toetsbias en itembias.....	21
3.3 Onderzoek naar bias.....	22
3.3.1 Onderzoek naar itembias .....	22
3.3.2 Onderzoek naar toetsbias.....	22
3.4 Wat is er bekend over de validiteit van de toelatingstoetsen?.....	23
<b>4. Onderzoeksvragen en uitvoering van het onderzoek</b> .....	<b>25</b>
4.1 Inleiding .....	25
4.2 Onderzoeksvragen.....	25
4.3 De uitvoering van het onderzoek en de samenwerking tussen FLOT en Cito.....	26
4.3.1 Gevolgen van coronamaatregelen .....	26
4.3.2 De aanpak van het onderzoek.....	27
4.4 Onderzoeksinstrumenten .....	28
4.4.1 De vragenlijst.....	28
4.4.2 De interviews.....	29

<b>5 De verzamelde gegevens en de toetsresultaten.....</b>	<b>31</b>
5.1 Betrouwbaarheid .....	32
5.2 Respons en kenmerken van de onderzochte groep.....	32
5.3 De variabelen van de vragenlijst .....	33
5.4 Slagingspercentages en verdelingen .....	35
5.5 Hoofdeffecten van migratieachtergrond .....	36
5.6 Hoofdeffecten van andere achtergrondvariabelen.....	37
5.7 Samenhang tussen migratieachtergrond en andere achtergrondvariabelen .....	39
5.8 Conclusies op het niveau van de toets .....	41
<b>6 Onderzoek naar DIF .....</b>	<b>43</b>
6.1 Inleiding .....	43
6.2 Data voor itembiasonderzoek .....	44
6.3 Methode en resultaten van itembiasonderzoek.....	45
6.3.1 Opvallende items bij Geschiedenis .....	46
6.3.2 Opvallende items bij Aardrijkskunde.....	47
6.3.3 Opvallende items bij Natuur & Techniek .....	48
6.4 Verklaringen voor DIF: kijkwijzer .....	48
6.5 Analyse van opgaven Geschiedenis.....	49
6.5.1 Inhoudelijke kenmerken van de opgaven Geschiedenis.....	50
6.5.2 Itembias of moeilijkheidsgraad?.....	51
6.5.3 Wat zeggen de experts?.....	52
6.5.4 Conclusie .....	53
6.6 Analyse van opgaven Aardrijkskunde.....	53
6.6.1 Inhoudelijke kenmerken van de opgaven Aardrijkskunde.....	54
6.6.2 Itembias of moeilijkheidsgraad .....	54
6.6.3 Wat zeggen de experts?.....	55
6.6.4 Conclusie .....	56
6.7 Analyse van opgaven Natuur & Techniek .....	56
6.7.1 Inhoudelijke kenmerken van de opgaven Natuur & Techniek.....	56
6.7.2 Itembias of moeilijkheidsgraad? .....	57
6.7.3 Wat zeggen de experts? .....	57
6.7.4 Conclusie .....	58
6.8 Samenvatting en conclusies .....	59

<b>7</b>	<b>Mogelijke biasbronnen in de toetsdoelen en het voorbereidende materiaal</b>	<b>61</b>
7.1	Inleiding	61
7.2	Powerful knowledge en referentiekaders	61
7.3	Het verwerven van powerful knowledge	62
7.4	Taalvaardigheid en leren	62
7.5	Een kijkkader voor de inhoud van de toelatingstoetsen	63
7.6	Bronnen voor toetsbias in de toelatingstoets Geschiedenis	64
7.7	Bronnen voor toetsbias in de toelatingstoets Aardrijkskunde	65
7.8	Bronnen voor Toetsbias in de toelatingstoets Natuur & Techniek	66
7.9	Bias in de voorbereiding op de toetsen	68
7.10	Opvattingen van aspirant-studenten en andere experts	68
7.11	Samenvatting en conclusies	70
<b>8</b>	<b>Samenvatting en conclusies</b>	<b>71</b>
8.1	Partijdigheid in de toetsopgaven	71
8.2	Partijdigheid in de inhoud van de toelatingstoetsen	72
8.3	Partijdigheid bij de voorbereiding	73
8.4	Bespreekpunten	74
	<b>Nawoord</b>	<b>75</b>
	<b>Literatuurlijst</b>	<b>77</b>
	<b>Bijlagen</b>	<b>79</b>
	Bijlage 1 Informatie voor aspirant-studenten en vragenlijst 'kenmerken aspirant-studenten'	79
	Bijlage 2 Informatie voor pabo-contactpersonen	82
	Bijlage 3 - Zes opgaven	83
	Bijlage 4 Etnische achtergrond	89
	Bijlage 5 Hoofdeffecten en significantie bij opleiding kandidaat	90
	Bijlage 6 Geschiedenis	91
	Bijlage 7 Aardrijkskunde	93
	Bijlage 8 Natuur & Techniek	95
	Bijlage 9 Het bepalen van gecorrigeerde effectgroottes	97
	Bijlage 10 Eisen vaststelling	101
	Bijlage 11 Een aantal aandachtspunten voor toetsconstructeurs	102







# Management- samenvatting

Met ingang van het studiejaar 2015-2016 moeten studenten die op grond van het bezit van een havo- of een mbo 4-diploma toelaatbaar zijn tot het hbo, aan aanvullende eisen voldoen om toegelaten te worden tot de pabo. Aspirant-pabostudenten moeten aantonen dat ze op de leergebieden aardrijkskunde, geschiedenis en natuur & techniek elementaire kennis in huis hebben op niveau havo-3/vmbo-t4. Deze 'bijzondere nadere vooropleidingseisen' zijn vastgelegd in de WHW. Ze zijn uitgewerkt door de SLO en worden in opdracht van de Vereniging Hogescholen en OCW door Cito geoperationaliseerd in de vorm van toelatingstoetsen. De toetsen bestaan per vak uit zestig meerkeuzevragen. Havisten die geen eindexamen hebben gedaan in een of meer van de genoemde schoolvakken, moeten in die vakken een toelatingstoets afleggen voordat ze met de pabo kunnen starten. Mbo'ers moeten voor alle drie de toelatingstoetsen slagen.

Uit onderzoeken van het Centraal Planbureau en ResearchNed is naar voren gekomen dat de nadere vooropleidingseisen in de eerste jaren na de invoering een drukkend effect hadden op de instroom van studenten in de pabo. Die terugloop, die intussen is afgenomen, was voorzien en was bij mbo-studenten groter dan bij havisten. Het gesignaleerde negatieve effect van de invoering van de toelatingstoetsen bleek verder ook samen te hangen met het geslacht van de aspirant-studenten, hun profielkeuze op de havo respectievelijk hun opleidingsrichting in het mbo en met de sociaaleconomische status en de geboorteplaats van hun ouders. Naar aanleiding van de laatste bevinding heeft de Minister van OCW aan de Tweede Kamer toegezegd dat zij de toetsopgaven in samenspraak met de pabo's, Cito en SLO zou laten doorlichten op 'westerse bias'. De Minister heeft de Vereniging Hogescholen (hierna 'VH') gevraagd om het onderzoek te laten uitvoeren.

Dit rapport doet verslag van dat onderzoek. Het onderzoek is in opdracht van de VH uitgevoerd door Fontys Lerarenopleiding Tilburg (hierna: 'FLOT'). Cito heeft medewerking verleend door het verzamelen van aanvullende gegevens over de kandidaten. Daarnaast nam Cito het uitvoeren en beschrijven van de psychometrische analyses op zich. Cito was ook actief betrokken bij de opzet van het onderzoek.

FLOT heeft de inhoud van de toetsen en de manier waarop aspirant-studenten zich daarop kunnen voorbereiden, geëvalueerd. Lerarenopleiders van FLOT hebben daarvoor de relevante documenten bestudeerd en gesprekken gevoerd met een aantal betrokken docenten, studenten en aspirant-studenten. FLOT heeft daarnaast 81 toetsopgaven nader onder loep genomen. Dat waren alle 65 opgaven die in een of meer statistische analyses een of andere vorm van opvallend gedrag vertoonden, aangevuld met 16 willekeurig gekozen andere opgaven. In verband met de noodzakelijke geheimhouding van de opgavenbanken was het niet mogelijk om alle opgaven te bekijken.

De dataverzameling heeft plaatsgevonden in september en oktober 2020. Circa 62% van de aspirant-studenten die gedurende deze periode één of meer toetsen hebben afgelegd, heeft ingestemd met deelname aan het onderzoek. Deze steekproef is wat betreft achtergrondkenmerken en slagingspercentages representatief voor de totale doelgroep van de toelatingstoetsen. Ook de verschillen tussen de slagingspercentages van aspirant-studenten met verschillende vooropleidingen en verschillende etnische achtergronden sluiten aan bij wat op grond van de onderzoeksresultaten van ResearchNed en Centraal Planbureau en eerdere toetsafnames verwacht kan worden.

Kandidaten met een niet-westerse migratieachtergrond behalen lagere scores en slagen minder vaak dan kandidaten zonder migratieachtergrond.

Slagingspercentages in september en oktober 2020			
De cesuur, dat wil zeggen de grens tussen zakken en slagen, ligt bij alle drie de toetsen rond 55%.			
	Kandidaten met twee in Nederland geboren ouders	Kandidaten met minstens één ouder uit een niet-westers land	Vershil
<b>Geschiedenis</b>	68% (van 668 kandidaten)	51% (van 93 kandidaten)	17%
<b>Aardrijkskunde</b>	80% (van 703 kandidaten)	60% (van 90 kandidaten)	20%
<b>Natuur &amp; Techniek</b>	77% (van 776 kandidaten)	42% (van 85 kandidaten)	25%

Het effect van het geboorteland van de ouders van de aspirant-studenten die aan dit onderzoek hebben meegewerkt, blijft zichtbaar wanneer wordt gecorrigeerd voor het effect van de vooropleiding van de aspirant-student (havo of mbo).

Slagingspercentages in september en oktober 2020						
(tussen haakjes aantallen kandidaten per groep)						
	havo		mbo-anders		GZW-onderwijsassistent	
	Geen migratieacht.	niet-westers	Geen migratieacht.	niet-westers	Geen migratieacht.	niet-westers
<b>Geschiedenis</b>	78% (139)	69% (13)	66% (141)	58% (24)	64% (258)	39% (36)
<b>Aardrijkskunde</b>	86% (220)	79% (19)	84% (138)	68% (25)	69% (247)	46% (35)
<b>Natuur &amp; Techniek</b>	84% (270)	72% (18)	83% (138)	48% (23)	67% (261)	24% (38)

Ook een correctie voor de effecten van het opleidingsniveau van hun ouders, de taal of talen die ze thuis respectievelijk met vrienden spreken en hun leesgedrag (boeken, tijdschriften), neemt het effect van de migratieachtergrond van kandidaten niet weg. Het verschil in de slagingspercentages kan ook niet worden verklaard door verschillen in verblijfsduur in Nederland: de meeste aspirant-studenten met een niet-westerse migratieachtergrond die aan het onderzoek hebben meegewerkt, zijn in Nederland geboren. Bijna 90% van de studenten met een niet-westerse migratieachtergrond heeft gedurende de hele schoolloopbaan op Nederlandse scholen gezeten.

De belangrijkste onderzoeksresultaten kunnen als volgt worden samengevat.

- De gemiddeld lagere scores en de lagere slagingspercentages van aspirant-studenten met een niet-westerse migratieachtergrond, kunnen niet verklaard worden door kenmerken van de afzonderlijke toetsopgaven. Noch de statistische analyses door Cito, noch de inhoudelijke analyses van FLOT wijzen op noemenswaardige culturele of talige itembias in het nadeel van aspirant-studenten met een niet-westerse migratieachtergrond.

Het onderzoek leidt tot andere, onderling samenhangende, bevindingen die wel bijdragen aan een mogelijke verklaring voor de lagere gemiddelde scores van aspirant-studenten met een niet-westerse migratieachtergrond.

- De toelatingseisen betreffen per vakgebied een grote hoeveelheid leerstof die op een zeer compacte manier is vastgelegd in Handreikingen die via [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl) met de aspirant-kandidaten worden gedeeld. Voor het zelfstandig verwerken van de informatie in de Handreikingen hebben de kandidaten een leesvaardigheidsniveau nodig dat niveau 3F van het Referentiekader Taal,

het eindniveau Nederlands van havo en mbo 4, overstijgt. Aspirant-studenten die de Nederlandse oriëntatie van de toetsdoelen niet van huis uit hebben meegekregen, zullen daar waarschijnlijk extra veel last van ondervinden.

- Het brede ondersteuningsaanbod voor aspirant-studenten dat intussen voorhanden is, volstaat niet voor iedere aspirant-student. Vooral de aspirant-studenten die de structuur die impliciet aanwezig is in de leerstof niet herkennen en die een deel van de veronderstelde voorkennis ontberen, zullen vaak meer begeleiding nodig hebben.

Op grond van hun onderzoeksresultaten formuleren de onderzoekers een aantal suggesties voor punten van overleg tussen beleidsmakers en de verschillende bij de toelatingstoetsen voor de pabo betrokken partijen. Zij benoemen vier punten die speciale aandacht verdienen: de inhoud van de toetsen, de manier waarop die inhoud beschreven wordt voor aspirant-studenten, de vorm van de toetsen en de manier waarop aspirant-studenten geacht worden zich op de toetsen voor te bereiden.

In hun nawoord verbinden de onderzoekers de uitkomsten van hun onderzoek aan de discussie over kansenongelijkheid in het Nederlandse onderwijs.





# Inleiding: achtergronden en leeswijzer

## 1.1 Achtergronden

Met ingang van het studiejaar 2015-2016 moeten aspirant-pabostudenten die in het bezit zijn van een havo- of een mbo 4-diploma, voordat ze tot de opleiding worden toegelaten, aantonen dat ze op leergebieden aardrijkskunde, geschiedenis en natuur & techniek kennis in huis hebben op een niveau dat vergelijkbaar is met havo-3 of vmbo-t4 (Staatsblad 2014, 284). Deze 'nadere vooropleidingseisen' zijn ingevoerd om de overladenheid van het curriculum van de pabo te verminderen en daarmee de kwaliteit van de opleiding te helpen verhogen. Havo-studenten die geen eindexamen hebben gedaan in een of meer van deze vakken en alle mbo'ers moeten om tot de pabo toegelaten te worden, slagen voor landelijk ontwikkelde toetsen in de genoemde vakken.

Uit onderzoek van het Centraal Planbureau (Bolhaar en Van Ruijven, 2019) en ResearchNed (Van Casteren, 2019) is naar voren gekomen dat de invoering van de nadere vooropleidingseisen een nadelig effect heeft op de instroom van studenten in de pabo. De terugloop, waar bij de invoering van de toelatingseisen al op werd gerekend, blijkt bij mbo-studenten substantieel groter dan bij havisten. Binnen de groepen met havo respectievelijk mbo-4 als vooropleiding komen óók verschillen naar voren. Het negatieve effect van de invoering van de toelatingstoetsen blijkt samen te hangen met de volgende kenmerken van de aspirant-studenten:

- geslacht;
- de profielkeuze op de havo c.q. de opleidingsrichting in het mbo;
- de sociaaleconomische status van hun ouders en
- het geboorteland van hun ouders.

In de periode nadat het CPB en ResearchNed hun onderzoek afronden, is het aantal jaarlijks afgenomen toelatingstoetsen flink gegroeid (Cito, 2021). In het schooljaar 2014-2015 waren er 10.734 afnames, inmiddels is dat aangetrokken tot 16.118 afnames in schooljaar 2018-2019. Ook in 2019-2020 zijn er meer dan 16.000 afnames geweest. Het aantal afgenomen toelatingstoetsen heeft een gunstig effect op de instroom in de pabo. Het afschrikkende effect van de toetsen lijkt in zijn algemeenheid af te nemen. Het is niet duidelijk of dit effect in alle onderscheiden groepen - studenten met verschillende etnische achtergronden, studenten met diverse vooropleidingen - op dezelfde manier optreedt.

## 1.2 De onderzoeksopdracht

Op 16 december 2019 schreven de bewindslieden van OCW een brief aan de Tweede Kamer (Ministerie van OCW, 2019) over de intensivering van de aanpak van de lerarentekorten in het onderwijs. Een van de maatregelen die de ministers in deze brief aankondigden, heeft betrekking op de toelatingstoetsen voor de pabo: op aanbeveling van ResearchNed zullen de toetsopgaven in samenspraak met de pabo's, Cito en SLO worden doorgelicht op westerse bias.

De Minister heeft de Vereniging Hogescholen gevraagd om dat onderzoek te laten uitvoeren.

De Vereniging Hogescholen heeft het onderzoek opgedragen aan Fontys Lerarenopleiding Tilburg (hierna: 'FLOT'). FLOT heeft bij de uitvoering van het project nauw samengewerkt met Cito dat, op verzoek van de VH en FLOT, haar medewerking heeft verleend<sup>1</sup>. FLOT en Cito hebben gezamenlijk de opzet van het onderzoek uitgewerkt. De bijdrage van Cito bestond daarnaast uit de organisatie van de toetsafnames, de dataverzameling en het uitvoeren en beschrijven van de psychometrische analyses.

Het voorliggende rapport doet verslag van het uitgevoerde onderzoek en beschrijft de opbrengsten ervan. De bijdrage van Cito komt in het rapport tot uiting in de hoofdstukken 2, 5 en het eerste deel van hoofdstuk 6. Deze hoofdstukken zijn door de betrokken medewerkers van Cito geschreven.

## 1.3 Leeswijzer

In het volgende hoofdstuk worden de achtergronden van de toelatingstoetsen nader uitgewerkt en worden de inhoud, opzet, ontwikkeling en afname van de toetsen beschreven.

Hoofdstuk 3 bevat een nadere verkenning van het centrale thema van dit onderzoek. Er wordt nader ingegaan op de betekenis van het concept 'bias' of 'partijdigheid' van toetsen en toetsopgaven, en op het verschil tussen 'bias' en 'moeilijkheidsgraad'.

In hoofdstuk 4 worden de onderzoeksvragen nader uitgewerkt en worden de opzet en uitvoering van het onderzoek beschreven. Daarbij wordt ook dieper ingegaan op de samenwerking en taakverdeling tussen de onderzoekers van Cito en FLOT.

Hoofdstuk 5, 6 en 7 behandelen de uitkomsten van het onderzoek. In hoofdstuk 5 beschrijft Cito de steekproef: de studenten die in september en oktober 2020 hebben deelgenomen aan de toelatingstoetsen en die bereid waren hun medewerking te verlenen aan dit onderzoek. Hun toetsresultaten worden in verband gebracht met gegevens over het geboorteland van hun ouders en andere mogelijke relevante achtergrondkenmerken.

Hoofdstuk 6 beschrijft de uitkomsten van het onderzoek naar mogelijke bias in de toetsopgaven. In het eerste deel van dit hoofdstuk doet Cito verslag van de uitgevoerde psychometrische analyses. In het tweede deel beschrijft FLOT de inhoudelijke analyse van in totaal 81 opgaven. Daarbij horen alle 65 opgaven die bij een of meer van de statistische analyses zijn opgevallen.

In hoofdstuk 7 gaat FLOT op zoek naar andere mogelijke verklaringen voor de achterblijvende toetsresultaten van aspirant-pabostudenten met een niet-westerse achtergrond. Daarbij wordt de inhoud van de toelatingstoetsen nader verkend. Er wordt ook gekeken naar de aanwijzingen die de aspirant-studenten krijgen voor de voorbereiding op de toets en de hulpmiddelen die hen daarbij worden aangereikt.

---

<sup>1</sup> Het onderzoeksteam bestond uit:

- vanuit Fontys Lerarenopleiding Tilburg (FLOT): Gijs van Gaans (lerarenopleider geschiedenis), Anne Kerkhoff (lector talentdidactiek), Job Noordeloos (lerarenopleider aardrijkskunde) en Tom Wils (lerarenopleider aardrijkskunde en N&T) en  
- vanuit Cito: Hanneke Bouwmans (logistiek toelatingstoetsen), Bas Hemker (psychometricus) en Silvia Matijssen (projectleider toelatingstoetsen).  
Verder hebben meegewerkt: Mieke van Vemden (lerarenopleider aardrijkskunde, Fontys Lerarenopleiding Tilburg), Desirée Joosten-ten Brinke (hoogleraar Leren van Volwassenen, Open Universiteit) en Max van der Velde (stagiair Cito, masterstudent Universiteit Utrecht).  
Verder hebben meegewerkt: Mieke van Vemden (lerarenopleider aardrijkskunde, Fontys Lerarenopleiding Tilburg), Desirée Joosten-ten Brinke (hoogleraar Leren van Volwassenen, Open Universiteit) en Max van der Velde (stagiair Cito, masterstudent Universiteit Utrecht).

In hoofdstuk 8 worden de onderzoeksresultaten samengevat en besproken. Het rapport wordt afgesloten met een kort nawoord van de onderzoekers,

## 1.4 Voordat u verder leest

De toelatingstoetsen operationaliseren de 'bijzondere nadere vooropleidingseisen voor aardrijkskunde, geschiedenis en het domein natuur & techniek'. Deze eisen zijn door de SLO in opdracht van de Minister beschreven en vastgelegd in het adviesrapport 'Naar een instapniveau pabo' (Diephuis et al, 2013). De SLO heeft bij het uitwerken van de vooropleidingseisen drie 'oriëntatiepunten' gebruikt:

- de kennisbasis die nodig is voor de pabo;
- het curriculum van het primair onderwijs waarvoor de studenten worden opgeleid en
- het curriculum in klas 1 tot en met 3 van de havo en/of het examenprogramma voor de bovenbouw van het vmbo theoretische leerweg.

Die drie oriëntatiepunten zijn duidelijk herkenbaar in de toelatingseisen voor de pabo. De beschreven kennis is in de eerste plaats georiënteerd op een Nederlands perspectief op de geschiedenis en de wereld om ons heen. Die oriëntatie wordt hierna niet ter discussie gesteld. Dit onderzoek is niet opgezet om de politieke besluitvorming omtrent de toelatingseisen voor de pabo te evalueren, maar om na te gaan of een bepaalde groep aspirant-studenten onbedoeld benadeeld wordt door de manier waarop die besluiten in de toelatingstoetsen vorm hebben gekregen.







# De toelatings- toetsen voor de pabo

## 2.1 Achtergronden en doelen

Sinds het studiejaar 2015-2016 zijn er bijzondere nadere toelatingseisen voor iedereen die de opleiding tot leraar basisonderwijs (pabo) wil volgen.

Om het basisniveau van de studenten die met de pabo beginnen te verhogen, moeten aspirant-studenten aantonen dat ze over een bepaald kennis- en vaardigheidsniveau beschikken voor de vakken aardrijkskunde, geschiedenis en natuur & techniek. Vóórdat ze worden toegelaten tot de pabo, moeten ze slagen voor toelatingstoetsen in deze drie vakken. Mbo'ers moeten in alle drie de vakken een toelatingstoets doen. Havisten krijgen vrijstellingen voor de vakken die in hun eindexamenpakket zaten. Vwo-abituriënten krijgen vrijstelling voor alle drie de toelatingstoetsen, ook als zij er géén examen in hebben gedaan.

Voor de vakken aardrijkskunde, geschiedenis en natuur & techniek geldt dat die naast rekenen en taal een belangrijke plaats innemen in het curriculum van de pabo. Het is van belang dat aankomende leerkrachten in het basisonderwijs de benodigde 'denk- en beoordelingskaders' ontwikkelen om de wereld om ons heen toegankelijk en begrijpelijk te maken voor basisschoolleerlingen. Het gaat bij de toelatingstoetsen daarom niet zo zeer om het toetsen van feitenkennis, maar om het toetsen van begrip en beoordelingsvermogen, om het inschatten van feiten en contexten op hun waarde, op het belang ervan voor het heden en de toekomst en om het leggen van verbanden tussen feiten, contexten en gebeurtenissen (Diephuis e.a., 2013).

De toelatingstoetsen zijn ingevoerd om op efficiënte wijze te toetsen of aspirant-pabostudenten voor genoemde vakgebieden over voldoende basiskennis en toepassingsvaardigheden beschikken. Het zijn toetsen waarmee een slaag-zak-beslissing genomen kan worden over de vakgebieden aardrijkskunde, geschiedenis en natuur & techniek. De kandidaat die behoort te slagen op de toelatingstoets, is die kandidaat die precies het niveau heeft ofwel de minimale hoeveelheid kennis van het vak heeft om toegelaten te worden tot de pabo. De toelatingstoetsen zijn uitdrukkelijk niet bedoeld als diagnostische toetsen. De gezakte kandidaten krijgen wel een uitslag op domeinniveau, zodat zij zich beter kunnen voorbereiden op een herkansing.

## 2.2 De inhoud van de toelatingstoetsen

De inhoudelijke basis van de toelatingstoets is het SLO-adviesrapport 'Naar een instapniveau pabo' (Diephuis e.a., 2013). Dit advies bevat een inhoudelijke beschrijving van de bijzondere nadere vooropleidingseisen voor aardrijkskunde, geschiedenis en natuur & techniek. Het laatste domein

bestaat uit de vakgebieden biologie en natuurkunde. Bij deze inhoudelijke beschrijvingen is er per vak door medewerkers van de SLO een Handreiking geschreven (Sluijsmans, 2014a; 2014b; 2014c). Deze Handreikingen zijn bedoeld voor aspirant-studenten voor de pabo en bevatten een inhoudelijke toelichting bij de bijzondere nadere vooropleidingseisen. In de Handreikingen is ook een toetsmatrijs opgenomen, waarin staat hoeveel vragen van elk domein/tijdvak/kernconcept er opgenomen zijn in de toets. De inhoud van de toelatingstoetsen ligt hiermee vast.

## 2.3 De opzet en de ontwikkeling van de toelatingstoetsen

### 2.3.1 Constructiegroep met vakdocenten

De toelatingstoetsen worden ontwikkeld door Cito. Cito gaat samen met een groep vakdocenten van verschillende pabo's met de Handreikingen aan de slag. De docenten bedenken de opgaven. Deze opgaven worden in de constructiegroep uitgebreid besproken met een toetsdeskundige met vakinhoudelijke en vakdidactische kennis. De opgaven worden gecontroleerd op inhoud, bijgeschaafd, aangevuld, verhelderd en vastgelegd in de opgavenbank. Voor de toelatingstoetsen worden op dit moment alleen meerkeuze-opgaven ontwikkeld.

### 2.3.2 Vaststellingscommissie met vakexperts

De opgaven die opgeleverd zijn door de constructiegroep, gaan daarna naar een vaststellingscommissie. Dat is een expertgroep waarin weer andere vakdocenten zitting hebben. Zij checken of een opgave voldoet aan de gestelde eisen (niveau en inhoud).

### 2.3.3 Bepaling cesuur door standaardsetting met pabo-, vo- en mbo-vakdocenten

Uit de uiteindelijke set vastgestelde opgaven worden de toelatingstoetsen samengesteld op basis van de toetsmatrijs. Dertig mbo-, vo- en pabodocenten bepalen hoeveel opgaven een student goed moet hebben om te slagen. Zij leggen gezamenlijk, volgens een gestandaardiseerde en wetenschappelijk onderbouwde methode, de cesuur vast. De cesuur is absoluut bepaald, en is dus niet afhankelijk van de prestaties van andere kandidaten.

De opgavenbank en de bepaalde cesuur vormen de basis voor de uitgegeven toelatingstoetsen. De opgavenbank maakt gebruik van een vaardigheidsschaal die het mogelijk maakt om verschillende toetsen te maken. Per afnameperiode worden er vier toetsversies per vak afgenomen. Deze versies zijn onderling vergelijkbaar op die vaardigheidsschaal: van iedere score in iedere versie is het mogelijk te bepalen welke vaardigheid (kennis) nodig is om die score te behalen. Hierdoor zijn per vakgebied alle kandidaten met elkaar te vergelijken, ongeacht de versie die ze gemaakt hebben, zowel binnen een afnamejaar, als over afnamejaren heen. Het is ook mogelijk om op basis van de informatie in de opgavenbank ieder jaar versies van de toetsen te maken die zowel inhoudelijk als wat betreft moeilijkheid sterk op elkaar lijken.

De cesuur is ook geplaatst op deze vaardigheidsschaal en geeft de minimale vaardigheid weer die een student moet bezitten om te kunnen slagen voor een toets. Iedere toets is aan die vaardigheidsschaal te relateren, zodat ongeacht de moeilijkheid van de toets, iedereen hetzelfde minimale vaardigheidsniveau moet laten zien om een voldoende te halen. Na het eerste jaar van toetsen met deze vaardigheidscesuur is met een tweede vergelijkbare standaardsettingssessie gecontroleerd of de cesuur gehandhaafd kon blijven. Bij deze tweede sessie met deels andere docenten bleek dat het geval. Hiermee is de cesuur vastgesteld, die ieder jaar twee keer statistisch wordt geëvalueerd. Een aspirant-student moet op dit moment om te slagen ongeveer 55% van de meerkeuze-opgaven goed maken. Dat betekent dat een aspirant-student gemiddeld zo'n 33 van 60 meerkeuzevragen met 3 of 4 antwoordmogelijkheden goed

moet beantwoorden om te slagen.

Naast de evaluatie van de cesuur vindt elk jaar ook nog een uitgebreide psychometrische analyse plaats om de kwaliteit van de opgaven te waarborgen. Opgaven die ondanks de verwachtingen van de vakdocenten en experts niet goed doen wat ze zouden moeten doen, namelijk discrimineren tussen kandidaten met voldoende kennis en vaardigheid en kandidaten met onvoldoende kennis en vaardigheid, worden uit de opgavenbank verwijderd.

## 2.4 De afname van de toelatingstoetsen

Een havo-leerling of een mbo-student die naar de pabo wil, meldt zich aan bij een pabo. De pabo verstrekt de aspirant-pabostudenten een link waarmee zij zich als ze nog niet aan de aanvullende toelatingseisen voldoen, in kunnen schrijven voor de toelatingstoetsen. De student mag de toelatingstoetsen op elke pabo doen; dit hoeft niet de pabo te zijn waar hij het onderwijs wil gaan volgen.

Voor de toelatingstoets zijn er de afgelopen jaren steeds zes 'toetsvensters' per studiejaar geweest. Een toetsvenster is de afnameperiode waarbinnen de student een toelatingstoets kan maken. Er zijn toetsvensters in april, juni, juli, augustus, november en januari. In het jaar 2020 zijn er aangepaste toetsvensters geweest in verband met de maatregelen rondom corona.

De toelatingstoetsen zijn digitale toetsen die afgenomen worden op de pabo's. De pabo zorgt voor een ingerichte toetsruimte en surveillanten. Zij hebben hiervoor een vastgesteld afnameprotocol ontvangen.

De toetsen worden afgespeeld via de examensoftware van Cito en de antwoorden van kandidaten worden rechtstreeks naar Cito gestuurd. Een pabo heeft hier geen inzage in. Bij Cito vindt de uitslagverwerking plaats. Aspirant-studenten die geslaagd zijn, ontvangen een uitslagbrief en een certificaat. Zij zijn zelf verantwoordelijk voor het aanleveren van het certificaat op de pabo. Aspirant-studenten die niet geslaagd zijn, ontvangen ook een uitslagbrief, met daarin de uitslag op de toets en een uitslag op hoofddomeinen. De uitslag op hoofddomeinen helpt hen bij de voorbereiding op een eventuele herkansing. Per instroommoment in de pabo mogen studenten elk vak één keer te herkansen. Bij een volgend instroommoment krijgen de studenten opnieuw twee kansen per vak.





## Een nadere verkenning van de onderzoeksvraag

### 3.1 Wat betekent 'bias'?

Onderzoek naar bias in toetsen komt heel vaak voort uit bevindingen zoals die van CPB en ResearchNed: er worden verschillen vastgesteld tussen de prestaties van verschillende groepen kandidaten op dezelfde toets. Vooral wanneer die verschillen onverwacht dan wel ongewenst zijn, leiden ze nogal eens naar vragen om onderzoek naar bias. Daarmee wordt dan een van de meest basale eisen aan de kwaliteit van een toets ter discussie gesteld: de validiteit van de meting. Een toets is valide als hij meet wat hij pretendeert te meten, oftewel als de conclusie die getrokken wordt op basis van de score, juist is. Een andere basiseis, betrouwbaarheid, kan worden beschouwd als een voorwaarde voor die validiteit: van een toets waarvan we niet zeker weten of hij bij een volgende afname dezelfde resultaten zal opleveren, weten we ook niet zeker wat er de eerste en de volgende keer gemeten wordt.

Een toets is valide als hij meet wat hij moet meten. Als prestatieverschillen tussen kandidaten of groepen kandidaten veroorzaakt worden door andere factoren dan verschillen in de kennis of vaardigheid die een toets moet meten, dan is de toets niet valide (Bügel & Sanders, 2006). We spreken dan van 'bias' of, met een duidelijker maar minder vaak gebruikt woord, van 'partijdigheid'.

Een voorbeeld kan het begrip 'bias' verduidelijken. Stel dat een onderzoeker iets wil weten over de leesvaardigheid van leerlingen. Uit onderzoek is bekend dat het begrip van een leestekst nauw samenhangt met de kennis die een lezer over het onderwerp van de tekst had vóórdat hij/zij begon te lezen (Perfetti & Stafura 2014, O'Reilly, Wang & Sabatini, 2019). De onderzoeker ontwikkelt een leestoets die uitsluitend bestaat uit teksten over gamen, een onderwerp dat meer jongens dan meisjes interesseert en waarvan verondersteld kan worden dat de meeste jongens er ook meer over weten dan meisjes. Als de toets bedoeld is om te meten hoe het is gesteld met de leesvaardigheid van de kandidaten, dan zal deze toets 'gebiased' zijn voor meisjes die niet veel met gamen hebben. 'Lezen van teksten over gamen' is immers niet hetzelfde als 'lezen' of 'begrijpend lezen'. Als het doel van de toets was geweest om 'leesvaardigheid van teksten over gamen' te meten, dan zou er géén sprake zijn van 'bias' of 'partijdigheid': alle opgaven meten immers dat specifieke toetsdoel.

### 3.2 Toetsbias en itembias

Onderzoek naar toetspartijdigheid kan betrekking hebben op de toets als geheel of op afzonderlijke toetsopgaven. Er is sprake van 'toetsbias' als de toets als geheel resultaten oplevert waarin de kennis of vaardigheden van bepaalde groepen kandidaten systematisch worden over- of onderschat. Het voorbeeld uit de vorige paragraaf heeft inderdaad betrekking op toetsbias: een toets die alleen teksten over gamen bevat, is geen valide meetinstrument om 'leesvaardigheid' te meten.

Naast onderzoek naar 'toetsbias' staat onderzoek naar 'itembias'. Onderzoek naar itembias heeft geen betrekking op de toets als geheel, maar op afzonderlijke opgaven. Ook hier gaat het om validiteit: meet een opgave wat hij moet meten of zijn er omstandigheden die ertoe leiden dat kandidaten die

de gemeten vaardigheid beheersen, de opgave toch fout beantwoorden? Bügel & Sanders (2006, p. 25) definiëren itembias als "(...)verschillen in prestaties op een item of opgave die veroorzaakt worden door kenmerken van die opgave die niet relevant zijn voor wat die opgave beoogt te meten". Uiterwijk (1974, p. 14) formuleert het als: "Een item is partijdig wanneer leerlingen uit onderscheiden subgroepen maar met een gelijke vaardigheid, een ongelijke kans hebben om het item goed te beantwoorden".

Itembias kan alleen van toetsbias onderscheiden worden als er *binnen* een toets sprake is van verschillen. Als alle opgaven c.q. 'items' van een toets hetzelfde meten, dan vinden we geen itembias. Als we naar ons voorbeeld over de leestoets teruggaan, dan zien we pas itembias als de leestoets voor een deel bestaat uit teksten over onderwerpen die voor de ene groep kandidaten interessanter of bekender zijn dan voor de andere. Die items zullen voor de onderscheiden groepen verschillend functioneren. Daarom wordt itembias ook wel 'differential item functioning' genoemd. Ook voor het onderhavige onderzoek naar mogelijke itembias in de toelatingstoetsen voor de pabo geldt dat we geen itembias zullen aantreffen als alle opgaven hetzelfde meten.

## 3.3 Onderzoek naar bias

Onderzoek naar bias in toetsen start vaak bij het vaststellen van onverwachte en vooral ongewenste uitkomsten van 'high-stake toetsen', dat wil zeggen van toetsen waarvan de resultaten ingrijpende gevolgen hebben voor de belanghebbenden. Bij die belanghebbenden horen vaak niet alleen de kandidaten, maar ook andere betrokkenen. Zo hebben de uitkomsten van de toelatingstoetsen voor de pabo niet alleen grote gevolgen voor de kandidaten - mogen ze starten met de opleiding van hun keuze? - maar ook voor de pabo's en de rest van onze samenleving. Bias in de toelatingstoetsen kan ongewenste gevolgen hebben voor het beleid gericht op het wegwerken van het nijpende tekort aan leerkrachten in het basisonderwijs. De bias kan ook in de weg staan van het streven van overheid, pabo's en scholen naar een 'kleurrijk' docentencorps waarin alle leerlingen en hun ouders zichzelf kunnen herkennen.

### 3.3.1 Onderzoek naar itembias

Onderzoek naar itembias vergt in tegenstelling tot onderzoek naar toetsbias niet per sé een extern criterium. In onderzoek naar itembias worden de scores van de verschillende groepen kandidaten op dezelfde opgaven gebruikt als een soort van 'interne criteria'. De kandidaten worden verdeeld in de groepen die vergeleken moeten worden, bijvoorbeeld 'kinderen die thuis alleen Nederlands spreken' en 'kinderen die thuis minstens twee talen spreken'. Die twee groepen worden dan ieder apart weer onderverdeeld in groepen met vergelijkbare gemiddelde toetsscores. Door per opgave de antwoorden van ééntalige leerlingen met een bepaalde totaalscore te vergelijken met de scores van de meertalige leerlingen die dezelfde totaalscore hebben gehaald, kan vervolgens worden vastgesteld of een bepaalde opgave voor beide groepen even moeilijk is. Als blijkt dat een bepaalde opgave voor de ééntalige leerlingen met een bepaalde gemiddelde toetsscore significant moeilijker is dan voor de tweetalige kandidaten met dezelfde gemiddelde score, dan wordt geconstateerd dat er sprake is van 'DIF': 'differential item functioning'. Daarmee is nog niet vastgesteld dat er sprake is van 'bias'. Om die conclusie te mogen trekken, zal nagegaan moeten worden hoe het komt dat de ééntalige leerlingen de vraag vaker fout beantwoord hebben dan de tweetalige. Van 'bias' is pas sprake als duidelijk is dat het verschil in moeilijkheidsgraad voortkomt uit verschillen tussen de groepen die los staan van het doel van de toets.

In hoofdstuk 6 wordt dieper ingegaan op onderzoek naar itembias in de toelatingstoetsen en de resultaten daarvan.

### 3.3.2 Onderzoek naar toetsbias

Het ontwikkelen van hypothesen over en het bedenken van mogelijke verklaringen voor het bestaan van toetsbias is doorgaans veel gemakkelijker dan het aantonen van de houdbaarheid van dat soort

vermoedens en veronderstellingen. Om de validiteit van een toets aan te tonen, heb je hoe dan ook een relevant criterium nodig: een gegeven waarmee je uitkomsten van de onderzochte toets bij verschillende groepen kandidaten kunt vergelijken. Meestal gaat het dan over de uitkomsten van een toets die hetzelfde moet meten en waarvan de kwaliteit onomstreden is, maar het kan ook een ander gegeven zijn waaraan je de uitkomsten van een toets kunt ijken. Afhankelijk van de aard van het gebruikte criterium kun je zo twee aspecten van de validiteit van een toets onder de loep nemen: de 'convergente validiteit' en de 'predictieve validiteit'. Onderzoek naar convergente validiteit behelst dat je de uitkomsten van de van bias verdachte toets vergelijkt met de uitkomsten van een toets waarvan je aanneemt dat die hetzelfde construct meet als die verdachte toets. In onderzoek naar de 'predictieve validiteit' ga je na in hoeverre je op grond van toetsresultaten voorspellingen kunt doen. Dat gebeurt dan door de uitkomsten van de toets te vergelijken met gegevens over aan de toetsinhoud gerelateerd toekomstig gedrag van de kandidaten. Een uitstapje naar de voormalige Cito Eindtoets Basisonderwijs helpt waarschijnlijk om deze twee soorten van validiteit te verduidelijken. Als de uitkomsten van de Cito Eindtoets Basisonderwijs, zoals vaak is gebeurd, worden vergeleken met de doorstroomadviezen van scholen, dan gaat het om 'convergente validiteit'. Vergelijken we de Citoscores van leerlingen met gegevens over de onderwijspositie die die leerlingen drie jaar later innemen in het voortgezet onderwijs, dan hebben we het over de 'predictieve validiteit' van het instrument.

Het is duidelijk dat er voor onderzoek naar toetsbias valide en betrouwbare gegevens nodig zijn over wat een toets moet meten. Dat is doorgaans geen sinecure en dat geldt ook voor het onderhavige onderzoek.

We gaan hier uit van de in de betrokken wet- en regelgeving vastgelegde hoofddoelen van de invoering van de toelatingstoetsen: de toelatingstoetsen moeten dienen om het curriculum van de pabo niet onnodig te belasten. Door alleen studenten toe te laten die kunnen aantonen dat ze de vereiste basiskennis over aardrijkskunde, geschiedenis en natuur & techniek beheersen, moet er in het curriculum van de pabo ruimte vrijkomen voor andere onmisbare aspecten van de opleiding. Om de predictieve validiteit van de toetsen te kunnen meten, zouden we in principe een extern criterium nodig hebben waarmee we kunnen vaststellen in hoeverre die doelen worden bereikt: wordt het curriculum van de pabo door invoering van de toetsen inderdaad minder belast met noodzakelijke bijspijkeractiviteiten en draagt dat bij aan een hoger niveau van de studenten bij hun afstuderen? Zo'n criterium is moeilijk om valide te operationaliseren en vergt hoe dan ook om een longitudinaal onderzoek. Daarvoor is meer tijd nodig dan er beschikbaar is voor het onderhavige project. Een ander fundamenteel probleem bij zo'n onderzoek naar predictieve validiteit zou zijn dat we van de studenten die niet zijn toegelaten, niet weten of ze de opleiding succesvol zouden hebben afgerond als ze wél waren toegelaten. In dit project beperken we ons daarom tot een analyse van de inhoud van het te toetsen materiaal en de wijze waarop studenten zich op de toets voorbereiden. In hoofdstuk 7 ontwikkelen we een 'kijkkader' dat ons moet helpen om mogelijke bronnen voor toetsbias in de toelatingstoetsen in het nadeel van aspirant-pabostudenten met een niet-westerse achtergrond op het spoor te komen.

## 3.4 Wat is er bekend over de validiteit van de toelatingstoetsen?

Er is tot op dit moment nog weinig bekend over de validiteit van de toelatingstoetsen voor de pabo. Ook de onderzoeken van het CPB en ResearchNed die de aanleiding vormden voor het onderhavige onderzoek, hebben op dit punt niet veel informatie opgeleverd. Deze onderzoeken leveren samen een schat aan gegevens over de eerste effecten van de invoering van de nadere vooropleidingseisen, maar de opzet van de projecten laat niet toe om de waargenomen effecten direct te relateren aan de inhoud van de toetsen. De verbeterde doorstroom naar het tweede jaar (Bolhaar & Van Ruijven, 2019) kan, zoals de onderzoekers zelf ook aangeven, ook het effect zijn van 'zelfselectie' onder studenten als gevolg van het gesignaleerde afschrikeffect van de toelatingstoetsen. Hetzelfde geldt voor de effecten die worden gemeld door de respondenten die door ResearchNed zijn bevraagd. Ook effecten als "(...) een hoger niveau in het eerste jaar, een homogener studentengroep, minder uitval tijdens de pabo (...)" (Van Casteren et al, 2019, p. 8) hoeven niet per sé samen te hangen met kenmerken van de toelatingstoetsen zelf.







# Onderzoeks- vragen en uitvoering van het onderzoek

## 4.1 Inleiding

Het uitgangspunt van deze studie is de toezegging van de bewindslieden van OCW aan de Tweede Kamer om de opgaven van de toelatingstoetsen voor de pabo te laten doorlichten op westerse bias. Het onderzoek moet antwoord geven op de vraag of verschillen in toetsscores tussen studenten met verschillende etnische achtergronden samenhangen met kenmerken van de toelatingstoetsen die niet relevant zijn voor het doel van die toetsen en die onbedoeld negatief uitpakken voor studenten met een niet-westerse achtergrond. Uitgaande van de vraag van de Minister zijn bij de start van dit onderzoek met de Vereniging Hogescholen de volgende afspraken gemaakt over wat het project zou moeten opleveren:

- informatie over mogelijke onbedoelde 'partijdigheid' in de toelatingstoetsen pabo;
- conclusies met betrekking tot de mogelijke aanwezigheid van biasbronnen in de kennis die in de toelatingstoetsen geoperationaliseerd wordt, dat wil zeggen de toetsinhoud en toetsdoelen die de Handreikingen van SLO beschrijven voor de leergebieden aardrijkskunde, geschiedenis en natuur & techniek ([www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl)); en
- handreikingen aan de toetsconstructeurs die helpen om bias in nieuw te ontwikkelen opgaven te voorkomen.

In de volgende paragraaf werken we de onderzoeksopdracht uit in vier onderzoeksvragen. Daarna bespreken we in paragraaf 4.3 de uitvoering van het onderzoek en presenteren we in 4.4 de gehanteerde onderzoeksinstrumenten.

## 4.2 Onderzoeksvragen

De eerste vraag die in ons onderzoek moet worden beantwoord, is de vraag naar itembias:

**Onderzoeksvraag 1:** *Bevatten de toelatingstoetsen voor de pabo opgaven die studenten met een niet-westerse achtergrond ten onrechte benadelen?*

Om deze vraag te beantwoorden, moeten we onderzoeken in hoeverre de opgaven van de toelatingstoetsen op een onpartijdige wijze meten wat ze moeten meten: beheersing van de in de Handreikingen bij de toets beschreven kennis en vaardigheden op de terreinen van geschiedenis, aardrijkskunde en natuur & techniek. Eerder onderzoek op dit terrein biedt duidelijke aanwijzingen omtrent de te hanteren methoden voor de dataverzameling, het opsporen van items met DIF en de inhoudelijke analyse van de opgespoorde 'verdachte' opgaven. In hoofdstuk 6 zullen we de uitvoering en opbrengsten van dit deel van het onderzoek presenteren.

Wanneer niet kan worden aangetoond dat een bepaalde toets items bevat die studenten met een niet-westerse achtergrond ten onrechte benadelen, dan betekent dat niet per sé dat die toets daarmee ook vrij is van toetsbias. Het is immers mogelijk dat de toets en de beschreven toetsinhouden kenmerken hebben die maken dat alle opgaven partijdig zijn. Als dat het geval is, dan zullen we geen 'differential item functioning' aantreffen terwijl er wel sprake is van 'toetspartijdigheid'. We zullen het onderzoek naar itembias in de toelatingstoetsen dan ook moeten combineren met onderzoek naar de mogelijke aanwezigheid van toetsbias: hebben de toetsen zélf kenmerken die ertoe leiden dat alle opgaven partijdig zijn en die de achterblijvende toetsresultaten van studenten met een niet-westerse achtergrond kunnen helpen verklaren?

We hebben daartoe de volgende twee onderzoeksvragen geformuleerd:

**Onderzoeksvraag 2:** *Bevat de inhoud van de pabotoelatingstoetsen Aardrijkskunde, Geschiedenis en Natuur & Techniek zoals die beschreven wordt in de Handreikingen, kenmerken die aspirant-pabostudenten met een niet-westerse achtergrond, onbedoeld en ten onrechte kunnen benadelen?*

**Onderzoeksvraag 3:** *Heeft de manier waarop studenten zich moeten voorbereiden op de toelatingstoetsen en hebben de hulpmiddelen die ze daarvoor krijgen aangereikt, kenmerken die aspirant-pabostudenten met een niet-westerse achtergrond, onbedoeld en ten onrechte kunnen benadelen?*

De antwoorden op de eerste drie onderzoeksvragen moeten leiden tot een antwoord op de vierde en laatste onderzoeksvraag en resulteren in de in de onderzoeksopdracht gevraagde aanbevelingen voor beleidsmakers en een handreiking voor betrokken toetsconstructeurs:

**Onderzoeksvraag 4:** *Hoe kan voorkomen worden dat aspirant-pabostudenten met een niet-westerse achtergrond systematisch benadeeld worden door de invoering van de toelatingstoetsen van de pabo?*

## 4.3 De uitvoering van het onderzoek en de samenwerking tussen FLOT en Cito

### 4.3.1 Gevolgen van coronamaatregelen

De maatregelen die in 2020 moesten helpen om de verspreiding van het coronavirus tegen te gaan, hebben ertoe geleid dat sommige voorgenomen activiteiten niet uitvoerbaar waren. We moesten daardoor afwijken van het plan zoals dat voor de start met de opdrachtgever was overeengekomen.

De voornaamste wijziging heeft betrekking op de inhoudelijke analyse van de items. In het oorspronkelijke onderzoeksplan was vastgelegd dat de medewerkers van FLOT alle opgaven die in de toelatingstoetsen voorkomen, vóór afname zou analyseren om mogelijke westerse bias te voorspellen. Om de noodzakelijke geheimhouding van de itembanken te bewaken, zouden de FLOT-medewerkers daarvoor naar Cito gaan om daar, onder toezicht van medewerkers van Cito, inzage te krijgen in de itembank. De coronamaatregelen maakten dergelijke bijeenkomsten echter niet uitvoerbaar. Daarom is besloten dat FLOT alleen inzage zou krijgen in de opgaven die na afname DIF vertoonden.

Om vergelijkingen mogelijk te maken, heeft Cito het mogelijk gemaakt dat de FLOT-medewerkers ook een beperkt aantal opgaven zonder DIF kon bekijken. In totaal heeft FLOT inzage gekregen in 81 opgaven, circa 25 opgaven per toelatingstoets. Met deze opzet sluit het uitgevoerde onderzoek aan bij wat in veel biasonderzoek gebruikelijk is.

De coronamaatregelen hebben het ook bemoeilijkt om informanten – pabo-studenten, aspirant-pabostudenten, docenten - te vinden die bereid zouden zijn om deel te nemen aan de interviews. Het aantal afgenomen interviews is daardoor beperkt gebleven. De uitkomsten van de gevoerde gesprekken die de rest van dit verslag gepresenteerd gaan worden, laten het dan ook niet toe om vérgaande conclusies te trekken. De interviews bieden niet meer dan een 'illustratie' van de gegevens die uit de rest van het onderzoek naar voren zijn gekomen.

### 4.3.2 De aanpak van het onderzoek

FLOT is in maart 2020 gestart met een verkenning van mogelijke biasbronnen in de toelatingstoetsen voor de pabo. De FLOT-medewerkers zijn vertrokken vanuit bestaande theorieën en opvattingen over enerzijds factoren die tot culturele bias kunnen leiden en anderzijds over kennis, de 'structuur' van kennis en het verwerven van kennis. Daarvan uitgaande hebben ze de Handreikingen bij de drie toelatingstoetsen en de andere informatie die gedeeld wordt via [www.10voordeleraar.nl](http://www.10voordeleraar.nl), bekeken en besproken. Parallel aan deze eerste inhoudelijke analyse vond overleg plaats met de Cito-collega's over de explicitering van de betekenis van 'bias' in de specifieke context van de toelatingstoetsen voor de pabo. Deze activiteiten leverden ook de eerste versies op van 'kijkkaders' ten behoeve van de analyse van de items met DIF die uit de statistische analyses van de toetsresultaten naar voren zouden komen. Uiteraard werd daarbij ook gebruik gemaakt van de resultaten van eerder verricht onderzoek naar culturele bias in toetsen.

Kort na de start van het project is ook begonnen met de voorbereiding van de dataverzameling. In dat kader zijn de volgende activiteiten uitgevoerd:

- FLOT en Cito hebben de planning van het project uitgewerkt. Daarbij werd vastgesteld dat de dataverzameling ten behoeve van de uit voeren itemanalyses pas kon starten in september 2020;
- Cito heeft de systemen voor de afname van de toelatingstoetsen zó aangepast dat het mogelijk werd aanvullende data over kandidaten te verzamelen en te koppelen aan hun toetsscores, zonder dat hun privacy in het geding zou komen;
- FLOT en Cito ontwikkelden gezamenlijk een schriftelijke vragenlijst (zie par. 4.4) om voor het onderzoek relevante kenmerken van informanten te verzamelen;
- via het aanmeldsysteem voor de toetsen werden aspirant-kandidaten geselecteerd die via mail benaderd konden worden met informatie over het onderzoek. Zij werden geïnformeerd over de achtergronden en doelen van het onderzoek. Daarbij werd aangekondigd dat zij op de toetsdag gevraagd zouden worden om medewerking te verlenen en een aanvullende vragenlijst in te vullen (zie bijlage 1);
- contactpersonen bij de diverse pabo's zijn via een mail geïnformeerd over het onderzoek en over de informatie die de aspirant-studenten daarover hadden ontvangen (zie bijlage 2);
- toen duidelijk was geworden dat er aan het toetsvenster van september niet voldoende kandidaten met een niet-westerse achtergrond hadden deelgenomen om de gewenste DIF-analyses uit te voeren, is besloten om ook tijdens het toetsvenster van oktober 2020 data te verzamelen. Dat besluit resulteerde opnieuw in noodzakelijke aanpassingen aan de toetssystemen van Cito en brieven aan aspirant-studenten.

Na de afname van de toetsvensters van september en oktober 2020 hebben in de daaropvolgende maanden november en december de volgende werkzaamheden plaatsgevonden:

- de medewerkers van Cito analyseerden, in overleg met de collega's van FLOT, de tijdens de toetsvensters verzamelde data. Het betrof statistische analyses van de achtergronden van de betrokken informanten en van de relaties tussen de over die kenmerken verzamelde gegevens en de toetsscores. Verder waren er verschillende DIF-analyses;
- de medewerkers van FLOT kregen inzage in de 65 opgaven die uit een of meer van de analyses naar voren kwamen doordat ze DIF vertoonden. Cito vulde de 'verdachte' opgaven aan met een 16 willekeurig gekozen opgaven die géén DIF vertoonden. Alle opgaven werden door FLOT-medewerkers geanalyseerd om te komen tot verklaringen voor de gesignaleerde DIF;
- in januari 2021 hebben de betrokken medewerkers van FLOT contact gezocht met een aantal ervaringsdeskundige studenten en andere betrokken experts. Daarbij werd gebruik gemaakt van het eigen netwerk en van de hulp van contactpersonen van de Vereniging Hogescholen bij de pabo's. Bij een deel van de benaderde personen zijn interviews afgenomen. In totaal betreft het de volgende personen:
  - drie studenten van de mbo-opleiding 'onderwijsassistent' die zich ten tijde van het onderzoek voorbereidden op doorstroom naar de pabo. De studenten hadden alle drie een niet-westerse migratieachtergrond;
  - vijf eerstejaars pabostudenten met twee in Nederland geboren ouders;
  - drie eerstejaars pabostudenten met een niet-westerse migratieachtergrond;
  - twee roc-medewerkers. Een van hen is werkzaam als docent aardrijkskunde. De andere informant heeft veel ervaring als docent Nederlands en NT2 en is momenteel werkzaam als adviseur onderwijskwaliteit;
  - twee vmbo-docenten aardrijkskunde met een niet-westerse migratieachtergrond.

Tijdens de gesprekken zijn zes opgaven aan de informanten voorgelegd (bijlage 3). Vijf van de zes opgaven waren uit de DIF-analyses naar voren gekomen. De zesde opgave was geselecteerd uit een van de voorbeeldexamens op basis van overeenkomsten met de opgave die geselecteerd was op grond van de gegevens over DIF. Deze opgave kon nog niet openbaar worden gemaakt.

In februari 2021 werd het verslag van het onderzoek afgerond. Een eerste versie van delen van het rapport is voorgelegd aan een externe deskundige met expertise op het terrein van toetsing. De verkregen feedback is verwerkt in de voorliggende versie van het rapport.

## 4.4 Onderzoeksinstrumenten

Ten behoeve van de dataverzameling zijn twee instrumenten ontwikkeld die hierna besproken worden: een vragenlijst voor de aspirant-studenten die meewerkten aan het onderzoek en een leidraad voor een interview met een klein deel van die doelgroep. De overige instrumenten, 'kijkwijzers' en statistische analyses, komen in de volgende hoofdstukken aan de orde.

### 4.4.1 De vragenlijst

#### *Uitgangspunten*

Bij de constructie van de vragenlijst die we aan de aspirant-studenten zouden voorleggen om achtergrondgegevens te verzamelen, is rekening gehouden met de volgende uitgangspunten.

- Voorop stond dat de vraag om medewerking aan het onderzoek en de aanvullende vragenlijst gecombineerd moesten worden met de afname van de toetsen. Als de toestemming van de kandidaten, hun antwoorden op de vragenlijst en hun toetsscores op verschillende momenten zouden worden verzameld, dan zou het ingewikkeld worden om die gegevens weer aan elkaar te koppelen zonder daarbij de privacy van de kandidaten in gevaar te brengen. Zo'n procedure paste niet in de opzet en het budget van het project.
- De vragenlijst moest zo kort mogelijk zijn: het was ongewenst om de kandidaten nadat ze de inspanning hadden geleverd om zestig toetsvragen te beantwoorden, meer vragen te stellen dan absoluut nodig waren.
- De combinatie van de vraag om medewerking, de aanvullende vragenlijst en de toets zelf, maakte het extra relevant om te voorkomen dat het onderzoek op één of andere manieren zou interfereren met de toetsprestaties die de kandidaten zouden gaan leveren. De kandidaten moesten erop durven vertrouwen dat hun bereidheid tot deelname en, bij deelname, de inhoud van hun antwoorden, op geen enkele manier van invloed zou zijn op de uitslag van hun toets.

#### *Informatie aan de kandidaten*

Op grond van de overwegingen die hiervoor besproken zijn, hebben we besloten de kandidaten in de week voordat ze de toets zouden gaan afleggen, een mail te sturen met uitleg over het onderzoek (zie bijlage 1). We hoopten dat het daardoor niet meer nodig was om de studenten uitvoerig te informeren op de toetsdag zelf. Om de kandidaten te ondersteunen bij hun beslissing om al dan niet mee te werken aan het onderzoek, voegden we de vragenlijst die we hen zouden willen voorleggen toe aan de mail met informatie over het onderzoek. We wilden daarmee ook bereiken dat de studenten die zouden willen meewerken alvast de gelegenheid kregen om zich te oriënteren op de beantwoording van de vragen.

#### *Inhoud van de vragenlijst*

Voor onderzoek naar culturele bronnen voor itembias is in de eerste plaats valide informatie nodig over de etnische achtergrond van de kandidaten. Daarnaast hebben we behoefte aan informatie over andere kenmerken van de kandidaten die van invloed kunnen zijn op de toetsprestaties.

#### *Etnische achtergrond*

Bij het in kaart brengen van de etnische achtergrond van de aspirant-pabostudenten, zijn we uitgegaan van de definitie die gehanteerd is in het onderzoek waarvan de resultaten de aanleiding vormden voor het onderhavige onderzoek: de studie van het CPB. Volgens die definitie, die CPB weer baseert op die

van het CBS, heeft een student een migratieachtergrond als hij of zij en/of minstens een van zijn of haar ouders in het buitenland geboren zijn. Bij de indeling naar een westerse dan wel een niet-westerse achtergrond volgt het CPB de definitie die het CBS hanteert (zie bijlage 4).

We zijn ons bewust van een aantal mogelijke kanttekeningen bij deze definitie. Die betreffen in de eerste plaats de indeling 'westers' versus 'niet-westers'. Bijlage 4 laat zien hoe divers de groep niet-westerse landen volgens de indeling van CBS is: variërend van St. Maarten en Curaçao tot de Westelijke Sahara. Daarnaast is het, zeker voor onderzoek zoals het onze dat op zoek is naar effecten van de etnische achtergrond van studenten, de vraag in hoeverre het verantwoord is om geen rekening te houden met de invloed die de partner met de westerse achtergrond zal hebben op de cultuur in het gezin dat hij of zij samen met een niet-westerse partner heeft gesticht. De gekozen definitie gaat óók voorbij aan effecten die een bepaalde achtergrond kan hebben op wat kinderen in een derde generatie 'van huis uit meekrijgen'. Het uitgangspunt van onze studie - 'we zijn op zoek naar verklaringen voor de bevindingen van het CPB' - heeft ons echter doen besluiten om de definitie die de onderzoekers van CPB hebben gehanteerd, over nemen. Pogingen die we hebben ondernomen om binnen onze steekproef te differentiëren tussen studenten met verschillende niet-westerse achtergronden, door bijvoorbeeld de drie Caraïbische landen in het Koninkrijk te onderscheiden van de andere niet-westerse landen, leverden geen werkbare indelingen op. Onze steekproef is daarvoor te klein.

#### *Andere gegevens die van invloed kunnen zijn op culturele oriëntatie en toetsprestaties*

Naast de vragen naar het eigen geboorteland en dat van beide ouders, hebben we de aspirant-studenten die de toelatingstoetsen in september en oktober 2020 hebben afgelegd, gevraagd naar de volgende kenmerken waarvan bekend is dat ze samenhangen met culturele oriëntatie en/of met schoolsucces.

- De taal of talen die studenten met hun ouders respectievelijk met vrienden spreken. Volgens sommige onderzoekers (Broeder e.a., 1993) is taalkeuze van mensen een betere indicator voor hun etnische achtergrond dan hun nationaliteit of geboorteland (of dat van hun ouders).
- Hun betrokkenheid bij de regio waar ze wonen, bij Nederland of bij een gebied buiten Nederland. We hoopten dat het antwoord op deze vraag zou helpen om verschillen tussen studenten wat betreft hun culturele oriëntatie beter in beeld te krijgen.
- Het aantal jaren onderwijs dat de studenten in Nederland hebben genoten. Verwacht kan worden dat de invloed van een niet-westerse achtergrond kleiner zal worden naarmate studenten langer op Nederlandse scholen hebben gezeten.
- Hun leesgedrag: hoe vaak lezen de studenten respectievelijk boeken, tijdschriften en nieuwsberichten op internet? Uitgangspunt was dat studenten door veel te lezen kennis verwerven die hun kansen op succes op toetsen die basiskennis aardrijkskunde, geschiedenis en natuur & techniek meten, zullen vergroten.
- Het opleidingsniveau van hun beide ouders. Er is veel onderzoek dat effecten laat zien van het opleidingsniveau van ouders op de onderwijskansen van hun kinderen (Zumbuehl & Dillingh, 2020).

In hoofdstuk 5 bespreken we de relaties tussen de verzamelde gegevens en de scores op de toelatingstoetsen.

#### **4.4.2 De interviews**

We hebben gesprekken gevoerd met een aantal eerstejaars pabostudenten en met mbo-studenten die overwegen zich aan te melden voor een vervolgopleiding aan de pabo. De gesprekken waren in de eerste plaats bedoeld om te verkennen in hoeverre onze bevindingen en voorlopige conclusies aansloten bij de ervaringen van een aantal studenten. We hadden niet de pretentie om onze conclusies met betrekking tot mogelijke verklaringen voor de achterblijvende toetsresultaten van studenten met een niet-westerse achtergrond empirisch te 'toetsen' aan de uitkomsten van de interviews. Daarvoor hebben we te weinig studenten gesproken.

De gesprekken bestonden steeds uit drie delen. Na een opening waarin de studenten bevestigd werden over hun opvattingen over ons onderzoek en hun eigen ervaringen met en opvattingen over de toetsen, kregen ze een aantal opgaven (bijlage 3) voorgelegd met de vraag om die 'hardop denkend' te beantwoorden. De antwoorden van de studenten boden aanknopingspunten om een aantal aspecten van de opgaven die ons tijdens het onderzoek waren opgevallen, nader met hen te bespreken. Ten slotte hebben we de studenten een paar van onze voorlopige conclusies voorgelegd.

Tijdens de gesprekken met docenten hebben we dezelfde opzet en dezelfde zes opgaven als uitgangspunt gekozen. De eigen ervaringen met de toetsen hadden dan uiteraard geen betrekking op ervaringen die de informanten als kandidaat hadden opgedaan, maar met hun ervaringen als docent.



## De verzamelde gegevens en de toetsresultaten

In dit hoofdstuk worden algemene omschrijvingen gegeven van de verzamelde data. Het betreft onder andere kenmerken van de aspirant-studenten die aan het onderzoek hebben meegewerkt en hun resultaten op de toelatingstoetsen voor de vakken aardrijkskunde, geschiedenis en natuur & techniek. Deze toetsresultaten geven een eerste indruk: welke verschillen vinden we tussen de groepen aspirant-studenten met verschillende migratieachtergronden? Zijn er toelatingstoetsen waar ze meer van elkaar verschillen en zijn er andere variabelen die de gevonden verschillen kunnen helpen verklaren? In het vervolg van dit hoofdstuk spreken we van 'kandidaten' in plaats van 'aspirant-studenten' om de leesbaarheid te vergroten.

Na dit hoofdstuk volgen in Hoofdstuk 6 de resultaten op itemniveau.

In reguliere jaren zijn er zes verschillende afnameperiodes, 'toetsvensters', waarbij kandidaten toetsen kunnen maken voor alle drie de vakken waarvoor aanvullende toelatingseisen gelden. Elk jaar worden de eerste vier periodes, in april, juni, juli en augustus, gebruikt door meer dan 85% van de kandidaten. De andere twee periodes vallen in november en januari. In 2020 zijn er andere afnamemomenten geweest. Het toetsvenster van april viel af door de maatregelen rond corona en ook de andere periodes zijn anders ingedeeld. Uiteindelijk waren er tien verschillende afnameperiodes in 2020, van juni 2020 tot en met januari 2021. Meer dan 95% van de toetsafnames heeft plaatsgevonden tijdens de eerste zeven afnameperiodes. In 2020 zijn er in totaal 16.501 toetsafnames geweest, dat zijn er iets meer dan de 16.415 toetsafnames in 2019.

De data voor dit onderzoek naar niet-westerse bias zijn verzameld in de toetsvensters in september en oktober 2020. Dat waren de vijfde en zesde afnameperiode van 2020. Alle betrokken kandidaten hebben in de week voor de toetsafnames een mail ontvangen met informatie over het onderzoek (Bijlage 1). De kandidaten die hebben meegewerkt aan het onderzoek, hebben direct na de afname van de toets actief ingestemd met deelname.

In de afnameperiode van het onderzoek zijn er in totaal 4.029 toetsafnames geweest<sup>2</sup>. Dit betrof 1.686 toetsen in september en 2.349 in oktober. Er moet opgemerkt worden dat er een onderscheid is tussen het aantal toetsafnames en het aantal kandidaten. Kandidaten kunnen binnen dezelfde afnameperiode voor meer dan één vak een toets gemaakt hebben. Ook kunnen kandidaten een toets in een latere afnameperiode herkansen. Per vak en binnen een afnameperiode is iedere toetsafname wel door een unieke kandidaat gemaakt. De maatregelen die nodig zijn om de privacy van de deelnemende personen te beschermen, maken het niet mogelijk om deze herkansende kandidaten te herkennen. Het aantal kandidaten dat in oktober 2020 een vak uit september heeft herkanst, is waarschijnlijk heel klein. Daarom worden de per toetsvenster en per vakgebied afgelegde 'toetsafnames' in deze analyse behandeld als afzonderlijke kandidaten.

<sup>2</sup> Er zijn 4.035 afnamerecords, maar bij 6 van deze records zijn alle opgaven overgeslagen en is er ook geen vragenlijst ingevuld. Deze zes records worden niet meegeteld als gemaakte toetsen.

## 5.1 Betrouwbaarheid

Voordat we dieper ingaan op de kenmerken van de kandidaten die aan dit onderzoek hebben meegewerkt, staan we in deze paragraaf kort stil bij de betrouwbaarheid van de toetsen die zij hebben afgelegd.

In de afnameperiodes september en oktober zijn er per vakgebied vier toetsversies afgenomen. De verschillende toetsversies krijgen allemaal een uniek nummer, in deze periodes waren dat de nummers 60 tot en met 63. De betrouwbaarheden van deze toetsversies zijn gegeven in tabel 5.1. De betrouwbaarheidsschatting is bepaald met Guttmans lambda-2, die te verkiezen is boven Cronbachs alfa (Sijtsma, 2009a, 2009b)<sup>3</sup>.

Tabel 5.1: Betrouwbaarheid van de toetsversies in september en oktober 2020

	AK	GS	NT
Versie 1 (60)	0,87	0,86	0,74
Versie 2 (61)	0,86	0,85	0,76
Versie 3 (62)	0,83	0,85	0,79
Versie 4 (63)	0,83	0,87	0,78

## 5.2 Respons en kenmerken van de onderzochte groep

In dit onderzoek is er geen sprake van een steekproef uit een welomschreven steekproefkader en er is ook geen sprake van randomisering. De onderzochte groep is volledig tot stand gekomen door zelfselectie, namelijk de kandidaten die zich opgegeven hebben voor de toetsen, en binnen die groep degenen die hebben aangegeven mee te willen werken aan het onderzoek en die de vragenlijst hebben ingevuld. Binnen deze groep kandidaten heeft geen selectie plaatsgevonden.

We hebben onderzocht of de kandidaten die in de onderzoeksperiode de toetsen hebben gemaakt, verschillen van de kandidaten in andere afnameperiodes. Dit is per vakgebied bekeken. Ook kunnen we nagaan of de kandidaten die mee hebben gewerkt aan het onderzoek, verschillen van de kandidaten die dat niet wilden. We kijken daarvoor naar de kenmerken die bij elke toetsafname van iedere kandidaat beschikbaar zijn: geslacht en vooropleiding. Deze kenmerken worden bij de inschrijving voor de toetsen standaard bij iedere kandidaat opgevraagd. We kijken ook naar het percentage geslaagden in een afnameperiode.

Uit voorgaande jaren is bekend dat de kandidaten die in de eerdere afnameperiodes van een jaar de toetsen maken, hogere slagingspercentages laten zien, dan de kandidaten die later in dat jaar de toetsen maken. Ook in 2020 zien we dat verschil: de slagingspercentages in de eerste afnameperiodes zijn hoger dan in de latere periodes. Het verschil is zelfs iets groter dan in 2019 het geval was. Het verschil wordt vooral veroorzaakt door hogere slagingspercentages in de eerste periode. In 2020 ligt het slagingspercentage in de eerste periode op 75,4% van de toetsen<sup>4</sup> en dat slagingspercentage is hoger

<sup>3</sup> Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107-12; Sijtsma, K. (2009b). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog*, 44, 561-567; [https://pure.uvt.nl/ws/portalfiles/portal/1134380/Over\\_misverstanden\\_rond\\_cronbachs.pdf](https://pure.uvt.nl/ws/portalfiles/portal/1134380/Over_misverstanden_rond_cronbachs.pdf)

<sup>4</sup> Er zijn kleine verschillen over de verschillende vaardigheden. De resultaten bij Natuur & Techniek liggen dichtbij de gemiddelde slagingspercentages, bij Aardrijkskunde zijn ze iets hoger (3 á 4%), en bij Geschiedenis iets lager (-3 á -4%). Daarbij zijn er weinig verschillen over de jaren, en volgen alle drie vaardigheden dezelfde trends. Om die reden worden nu alleen de gemiddelden over alle drie vaardigheden besproken.



dan in 2019. Toen was dat percentage 71,4%. De vijfde en de zesde periode, waarin de dataverzameling voor het onderhavige onderzoek plaatsvond, beschouwen we als 'latere' periodes. Zoals gebruikelijk bij latere afnameperiodes ligt het slagingspercentages met 71% weer iets onder het gemiddelde slagingspercentage van dat jaar. De vijfde en zesde periode verschillen onderling ook van elkaar. De latere periode toont weer een lager slagingspercentage: 74% in september, en 68% in oktober.

Een toetsresultaat is meegenomen in het verdere onderzoek als er aan twee voorwaarden is voldaan. Ten eerste moet de kandidaat expliciet toestemming gegeven hebben om zijn/haar antwoorden op de vragenlijst mee te nemen in het onderzoek en ten tweede moeten de antwoorden op deze vragen ook zijn ingevuld. In totaal voldoen 2.518 (62%) van de kandidaten aan beide voorwaarden. Er waren 26 kandidaten die de toestemmingsvraag hadden overgeslagen, maar wel de vragenlijstvragen hadden ingevuld. Deze kandidaten zijn niet meegenomen in het onderzoek. Het komt ook voor dat een kandidaat toestemming geeft, maar een of meer vragen overslaat. De antwoorden die wel zijn ingevuld, zijn gebruikt binnen dit onderzoek.

Als we kijken naar de slagingspercentages dan zien we dat de kandidaten die meewerken aan het onderzoek over het algemeen iets hogere slagingspercentages laten zien dan degenen die niet meewerken. De verschillen zijn in september (76% tegenover 69%) groter dan in oktober (69% tegenover 67%). Bij Natuur & Techniek (7% verschil) zien we een groter verschil dan bij de andere twee vaardigheden (2% à 3% verschil). Bij een uitsplitsing naar afnamemoment en vak is alleen bij Natuur & Techniek in september het verschil significant (67% versus 78%;  $p < .05$ )<sup>5</sup>.

De groepen van kandidaten die wel en geen toestemming verleenden, verschilden onderling op slechts één van de achtergrondvariabelen die standaard worden bevroegd bij de inschrijving voor de toetsen. Het enige significante verschil betreft de mannen die bij het vak Geschiedenis in september vaker dan vrouwen toestemming gaven om de gegevens te gebruiken (80% versus 69%). Verder waren de twee groepen vergelijkbaar wat betreft leeftijd, wel of geen dyslexie, geslacht of vooropleiding.

De variabelen geslacht en vooropleiding in periode 5 en 6 in 2020 zijn ook vergeleken met die in 2019. Dan is te zien dat de verdeling in beide jaren niet zeer veel verschilt. De man-vrouw verdeling is in beide jaren gelijk. In 2020 zijn er iets meer kandidaten van de havo afkomstig, en iets minder van het mbo of andere opleidingen. In de vijfde en zesde afnameperiode zijn relatief iets meer mannen dan bij de andere momenten, en lijkt de verdeling naar vooropleiding meer op de verdeling van de vooropleiding in 2019.

Al met al kan de eindconclusie zijn dat de deelnemers aan het onderzoek niet substantieel verschillen van de deelnemers in afgelopen twee jaren. Er is geen reden om aan te nemen dat deze groep uitzonderlijk is.

## 5.3 De variabelen van de vragenlijst

De vragen in de vragenlijst hebben betrekking op een aantal kenmerken van de kandidaten:

- het geboorteland van de kandidaat zelf en het geboorteland van beide ouders (afzonderlijk bevroegd);
- de dagelijkse omgangstaal, met ouders respectievelijk met vrienden;
- het aantal jaar basisonderwijs dat de kandidaat in Nederland heeft genoten;
- het leesgedrag van de kandidaat;
- de betrokkenheid van de kandidaat bij een bepaald gebied en
- het opleidingsniveau van de ouders (moeder en vader afzonderlijk).

De kandidaten waren van te voren bekend met deze vragen. De gegevens die hierna worden gepresenteerd betreffen in alle gevallen zelfrapportage die niet door middel van externe informatie gecontroleerd is.

---

<sup>5</sup> Wat betreft de reden voor de kleine verschillen is het gissen. Een hypothese kan zijn dat kandidaten die het gevoel hadden slecht gepresteerd te hebben op de toets ook geen zin meer hadden om verder met de toets bezig te zijn en daarom de vragenlijst niet meer in wilden vullen.

De vragen aangaande het geboorteland zijn gebruikt om met behulp van de definitie van het CBS (zie par. 4.4.1 en Bijlage 4) de opdeling te maken tussen de onderzoeksgroepen: kandidaten van wie beide ouders in Nederland zijn geboren (geen migratieachtergrond), kandidaten met een westerse migratieachtergrond (westers) of kandidaten met een niet-westerse migratieachtergrond (niet-westers). Kinderen van in Nederland geboren ouders die zelf kind zijn van migranten, worden hier (en door CBS) als Nederlands beschouwd. De verdeling van achtergrondvariabelen over de drie vakken is gegeven in tabel 5.2. Bij vier observaties is niet bekend in welke van deze drie categorieën de kandidaat valt.

Tabel 5.2: Aantal observaties (N) en percentages (%) van de migratieachtergronden per vak

Vak	Aardrijkskunde (AK)		Geschiedenis (GS)		Natuur & Techniek (NT)		Totaal	
	N	%	N	%	N	%	N	%
Geen	703	86%	668	84%	776	86%	2147	85%
Westers	28	3%	33	4%	38	4%	99	4%
Niet-westers	90	11%	93	12%	85	9%	268	11%
Totaal	821	100%	794	100%	899	100%	2514	100%

In totaal zijn er 268 toetsen afgenomen bij kandidaten met een niet-westerse migratieachtergrond (d.w.z. 11% van het totaal aantal in september en oktober afgelegde toetsen) en 99 (4%) toetsen bij kandidaten met een westerse migratieachtergrond. Het totaal aantal kandidaten verdeelt zich over de drie vakken.

Voor ieder vak zijn er in de twee periodes vier verschillende toetsversies beschikbaar. Dat betekent dat de aantallen kandidaten per toetsversie lager zijn. Deze aantallen zijn gegeven in tabel 5.3.

Tabel 5.3: Aantal observaties per toetsversie per vak voor de verschillende migratieachtergronden

Vak	Aardrijkskunde (toetsversies)				Geschiedenis (toetsversies)				Natuur & Techniek (toetsversies)			
	AK60	AK61	AK62	AK63	GS60	GS61	GS62	GS63	NT60	NT61	NT62	NT63
Geen	170	180	179	174	175	189	150	154	198	190	182	206
Westers	5	6	10	7	8	8	11	6	8	11	12	7
Niet-westers	23	28	21	18	30	27	16	20	23	24	21	17
Totaal	198	214	210	199	213	224	177	180	229	225	215	230

Zoals eerder beschreven, zijn de resultaten op de toetsversies onderling vergelijkbaar. De toetsversies zijn per vak niet alleen inhoudelijk vergelijkbaar, maar ook wat betreft de psychometrische eigenschappen. Met behulp van item-respons-theorie (IRT) is ervoor gezorgd dat de vier toetsversies vrijwel even moeilijk en betrouwbaar zijn. Om echter ook voor zeer kleine verschillen te kunnen corrigeren en er zeker van te zijn dat iedereen met een gelijke vaardigheid ook een gelijk oordeel krijgt, wordt er gewerkt met 'vaardigheidsschattingen' in plaats van toetsscores. In de rest van dit hoofdstuk worden de toetsresultaten steeds weergegeven in termen van deze geschatte vaardigheden.

Naast de antwoorden op de vragen naar het geboorteland zijn ook de antwoorden op de andere vragen in de vragenlijst nader bekeken. Deze gegevens worden in paragraaf 5.6 besproken.

## 5.4 Slagingspercentages en verdelingen

In paragraaf 5.2 werd gerapporteerd dat het gemiddelde slagingspercentage in de onderzoeksperiode op 71% lag. Als we kijken naar de slagingspercentages per migratieachtergrond en per vak, dan zien we de resultaten zoals beschreven in tabel 5.4.

Tabel 5.4: Slagingspercentages per migratieachtergrond

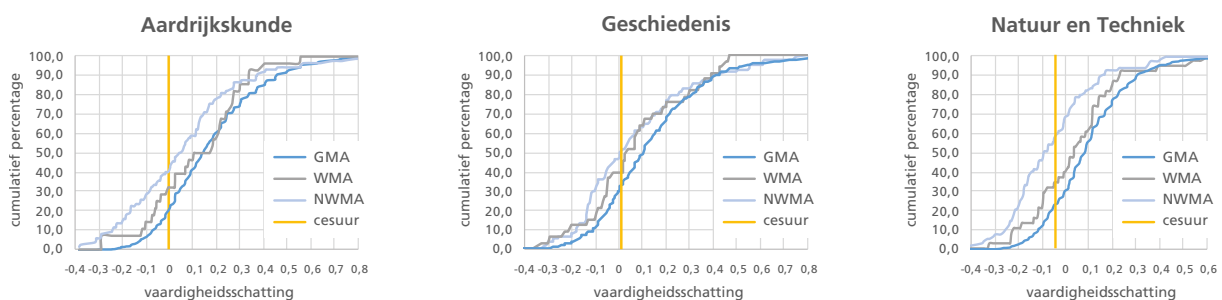
	Geen	Westers	Niet-westers
Geschiedenis	68%	61%	51%
Aardrijkskunde	80%	68%	60%
Natuur & Techniek	77%	66%	42%

De verschillen in slagingspercentages tussen kandidaten van wie beide ouders in Nederland geboren zijn en kandidaten met een westerse migratieachtergrond zijn vergelijkbaar voor Aardrijkskunde en Natuur & Techniek (11-12%). De verschillen tussen deze twee groepen zijn voor Geschiedenis iets kleiner (7%).

De verschillen tussen kandidaten van wie beide ouders in Nederland geboren zijn en kandidaten met een niet-westerse migratieachtergrond zijn groter. Voor Geschiedenis en Aardrijkskunde zijn ze vergelijkbaar (17-20%) en voor Natuur & Techniek nog groter (35%). Dit patroon zien we ook terug bij de verschillen tussen kandidaten met een westerse migratieachtergrond en kandidaten met een niet-westerse migratieachtergrond: de slagingspercentages liggen dan wel dichter bij elkaar (AK/GS: 8-10%, en N&T: 23%).

In Figuur 5.1 zijn de cumulatieve verdelingen gegeven van de vaardigheidsschattingen voor de drie vakken en voor de drie groepen met verschillende migratieachtergronden: geen (= GMA – donkerblauw), niet-westers (= NWMA – lichtblauw), en westers (= WMA - grijs). In deze figuur zijn ook de cesuurlijnen gegeven op de vaardigheidsschaal. Kandidaten met een vaardigheid rechts van de cesuurlijn, zijn geslaagd en links van de lijn zijn de kandidaten gezakt.

Figuur 5.1: Vaardigheidsschattingen voor drie groepen kandidaten



De cumulatieve percentages geven de percentages kandidaten aan met een lagere of gelijke vaardigheidsschatting per schaalscore op de vaardigheidsschaal. In figuur 5.1 is duidelijk zichtbaar dat voor vrijwel elke schaalscore geldt dat het cumulatieve percentage kandidaten met een niet-westerse achtergrond hoger is dan het cumulatieve percentage kandidaten zonder migratieachtergrond. De verschillen tussen de vaardigheidsschattingen voor de onderscheiden groepen kandidaten in tabel 5.4 gelden dus niet alleen op het cesuurpunt, maar over alle schaalscores op de vaardigheidsschaal.

## 5.5 Hoofdeffecten van migratieachtergrond

De resultaten van de analyses worden als eerste gegeven voor de belangrijkste onafhankelijke variabele van dit onderzoek: de migratieachtergrond.

Tabel 5.5: Vaardigheidsscores per achtergrond per vak

Vak	Aardrijkskunde		Geschiedenis		Natuur & Techniek	
	Gemiddelde	St.afw	Gemiddelde	St.afw	Gemiddelde	St.afw
Geen	0,175	0,217	0,143	0,249	0,095	0,176
Westers	0,116	0,203	0,072	0,215	0,047	0,197
Niet-westers	0,060	0,258	0,067	0,248	-0,059	0,176
Totaal	0,160	0,224	0,131	0,249	0,078	0,182

In tabel 5.5 zien we dat de groep kandidaten met een niet-westerse migratieachtergrond bij alle vakken de laagste gemiddelde vaardigheidsscore haalt en de groep kandidaten met twee in Nederland geboren ouders (hier aangeduid als 'Geen (migratieachtergrond)') de hoogste.

Op basis van de gegevens in tabel 5.5 is het lastig om te bepalen of de gepresenteerde verschillen nu groot of klein zijn. Om dat te kunnen, kunnen de verschillen beter gepresenteerd worden als effectgroottes. Effectgroottes<sup>6</sup> onder<sup>7</sup> de 0,20 worden vaak als 'niet betekenisvol' gezien, verschillen tussen 0,20 en 0,50 als 'klein', verschillen tussen 0,50 en 0,80 als 'matig', en verschillen groter dan 0,80 als 'groot'. De standaardafwijkingen van de verschillende groepen verschillen niet significant van elkaar zodat deze effectgroottes ook goed te interpreteren zijn. De effectgroottes worden gegeven in tabel 5.6. Verschillen met effectgroottes groter dan 0,31 zijn significant ( $p < 0,01$ ).

Tabel 5.6: Effectgroottes van de verschillen tussen de drie groepen

Verschillen	Aardrijkskunde	Geschiedenis	Natuur & Techniek
Geen vs niet-westers	0,52*	0,31*	0,88*
Geen vs westers	0,26	0,29	0,27
Westers vs niet-westers	0,26	0,02	0,61*

\* significant verschil ( $p < 0,01$ )

Alle verschillen tussen de groepen kandidaten zonder migratieachtergrond en kandidaten met een niet-westerse migratieachtergrond zijn significant ( $p < 0,01$ ). Het verschil is klein bij Geschiedenis, matig bij Aardrijkskunde en groot bij Natuur & Techniek. Kijken we naar de verschillen tussen de groepen kandidaten zonder migratieachtergrond en kandidaten met een westerse migratieachtergrond, dan zien we dat de effectgroottes amper variëren over de vaardigheden. De effectgroottes van de verschillen zijn op één na ook niet significant. Dat is voor een groot deel te verklaren doordat het aantal observaties in de groep kandidaten met een westerse migratieachtergrond klein is. Als we kijken naar de verschillen tussen

<sup>6</sup> Effectgroottes worden verkregen door het verschil te delen door de (gepoolde) standaardafwijking.

<sup>7</sup> We kijken nu alleen naar de positieve waarden. Deze stelregel geldt in principe voor absolute waarden. Dat betekent dat het gaat over een absoluut verschil kleiner dan 0,20, 0,50 en 0,80.

de groep kandidaten met een westerse migratieachtergrond en een niet-westerse migratieachtergrond dan zien we alleen bij Natuur & Techniek een significant verschil.

Om de verschillen beter te kunnen interpreteren en te onderzoeken, zijn de verschillen ook bekeken als gecorrigeerd wordt voor diverse achtergrondvariabelen. Daarvoor kijken we naar het effect van de variabelen die bij inschrijving altijd opgevraagd worden, en de variabelen die zijn bevraagd in de aanvullende vragenlijst. In de volgende paragraaf 5.6 worden de hoofdeffecten van deze achtergrondvariabelen beschreven. In paragraaf 5.7 bespreken we de samenhang tussen migratieachtergrond en de andere achtergrondvariabelen .

## 5.6 Hoofdeffecten van andere achtergrondvariabelen

### De vooropleiding van de kandidaten

De rapportages die Cito jaarlijks levert aan de Vereniging Hogescholen, laten zien dat de kandidaten met een mbo-opleiding onderwijsassistent of een andere mbo-opleiding in de sector Zorg en Welzijn ieder jaar relatief laag scoren op de toelatingstoetsen. Hun scores zijn significant lager dan de scores van kandidaten met een havodiploma of een mbo 4-diploma uit een andere sector, bijvoorbeeld mbo groen of techniek. Dit verschil is ook zichtbaar bij de kandidaten die aan dit onderzoek hebben meegewerkt. In bijlage 5 staan de hoofdeffecten in een tabel weergegeven met de mate van significantie. De gegevens maken duidelijk dat vooral de onderwijsassistenten lager scoren dan kandidaten met een ander mbo-diploma of een havodiploma. Dat geldt ook voor de kandidaten met een niet-westerse migratieachtergrond. De aantallen kandidaten met een niet-westerse migratieachtergrond zijn te klein om significante effecten aan te tonen.

Tabel 5.4 presenteert de slagingspercentages van de kandidaten met verschillende migratieachtergronden. Daarbij werd zichtbaar dat het slagingspercentage onder kandidaten met een niet-westerse achtergrond bij alle drie de vakken lager is dan onder de kandidaten met een andere achtergrond. Het verschil is steeds het grootst tussen de kandidaten met een niet-westerse migratieachtergrond en de kandidaten van wie beide ouders in Nederland zijn geboren.

In tabel 5.7 worden de slagingspercentages van de kandidaten met verschillende migratieachtergronden opnieuw vergeleken, maar nu wordt ook rekening gehouden met de vooropleidingen van de onderscheiden groepen. Er worden drie vooropleidingen onderscheiden: 'havo', 'mbo 4-onderwijsassistent' en 'alle andere mbo-opleidingen'. Tabel 5.7 maakt duidelijk dat de uitsplitsing in drie categorieën per cel lage aantallen kandidaten met een niet-westerse migratieachtergrond oplevert. Daarom kunnen de op deze manier onderscheiden subgroepen niet meer beschouwd worden als representatief voor de volledige populatie van kandidaten met een niet-westerse migratieachtergrond en de betreffende vooropleiding. Veel van de effecten van een niet-westerse migratieachtergrond zijn daarom niet meer significant. Een verder onderscheid binnen de categorieën 'havo' en 'mbo anders' zou helemaal geen betekenisvolle gegevens opleveren.

Tabel 5.7: Aantal kandidaten en slagingspercentages per vooropleiding en migratieachtergrond

opleiding	migratieachtergrond	Aantal kandidaten			% geslaagd		
		GS	AK	NT	GS	AK	NT
Havo	Geen	139	220	270	78%	86%	84%
	Niet-westers	13	19	18	69%	79%	72%
Mbo Anders	Geen	141	138	138	66%	84%	83%
	Niet-westers	24	25	23	58%	68%	48%
Mbo Onderwijsassistent	Geen	258	247	261	64%	69%	67%
	Niet-westers	36	35	38	39%	46%	24%

In tabel 5.4 zagen we voor de gehele populatie bij de toetsen Geschiedenis en Aardrijkskunde een verschil in slagingspercentages van 17% respectievelijk 20% tussen de kandidaten zonder migratieachtergrond en de kandidaten met een niet-westerse migratieachtergrond, en bij Natuur & Techniek een verschil van 35%. Na de uitsplitsing naar vooropleiding in tabel 5.7 zijn de verschillen in slagingspercentages tussen de kandidaten met verschillende migratieachtergronden bij de havisten lager dan in tabel 5.4, namelijk 9%, 7% en 12%. Bij de kandidaten met de vooropleiding 'onderwijsassistent' zijn de verschillen tussen de onderscheiden etnische groepen groter: 25%, 23% en 43%. Bij kandidaten met een ander mbo-diploma is het verschil bij Geschiedenis vergelijkbaar met dat bij de havisten, namelijk 8%, en bij de andere twee toetsen vergelijkbaar met de gehele populatie: 16% en 35%. In bijlage 6 tot en met 8 worden voor de drie vakken de gemiddelde scores gegeven voor al deze groepen, evenals voor de groepen die niet in deze tabel staan. Daarin zijn ook de aantallen kandidaten per groep gegeven.

### Overige achtergrondvariabelen

Om te onderzoeken of kenmerken als 'taalkeuze', 'eigen ervaring in het basisonderwijs', 'leesgedrag', 'betrokkenheid met een bepaalde regio' of 'opleidingsniveau van de ouders' van invloed zijn op de vaardigheidsscores die kandidaten halen bij de toelatingstoetsen, is de kandidaten aan het einde van de toets gevraagd om een aantal vragen naar die kenmerken te beantwoorden.

Tabel 5.8 presenteert de waargenomen effecten van deze kenmerken, voor zover die significant zijn gebleken. Er wordt in deze tabel geen onderscheid gemaakt tussen kandidaten met verschillende migratieachtergronden. Bij alle drie de vakken zien we een significant verschil in score tussen mannen en vrouwen en tussen kandidaten die aangeven dat ze met hun ouders thuis vooral Nederlands spreken en kandidaten die zeggen thuis (vooral) andere talen te spreken. We zien bij Aardrijkskunde en Geschiedenis dat een aantal specifieke persoonlijke factoren correleert met de behaalde toetsscore: wanneer een kandidaat zijn of haar hele basisschool in Nederland doorliep, als hij of zij zich meer betrokken voelt bij Nederland of een regio in Nederland, wanneer een kandidaat dagelijks een boek of een krant leest en als het opleidingsniveau van de vader hbo of universitair is, dan is de kans op een hogere score groter. Bij de overige kenmerken, bijvoorbeeld bij het opleidingsniveau van de moeder, zagen we in dit onderzoek geen significante effecten.

Tabel 5.8: Hoofdeffect en significantie van kenmerken, uitgesplitst naar vak

Kenmerk			AK		GS		NT	
			effect	sign*	effect	sign	effect	sign
Geslacht	Man	Vrouw	0,42	***	0,29	***	0,31	***
Taal met ouders	Vooraf Nederlands	Niet vooraf Nederlands	0,69	***	0,30	***	0,77	***
Taal met bekenden	Vooraf Nederlands	Niet vooraf Nederlands	0,38	*	0,25	*	0,08	
Ervaring in basis-onderwijs	Heel BO (1-8)	Niet heel BO (1-8)	0,65	***	0,39	*	0,40	
Betrokkenheid bij eigen plaats/regio	Vooraf betrokken bij eigen regio	Niet vooraf betrokken bij eigen regio	-0,20	***	-0,16	*	0,06	
Boeken lezen	Dagelijks	Niet dagelijks	0,47	**	0,68	**	0,30	
Krant lezen	Dagelijks	Niet dagelijks	0,46	**	0,20		0,05	
Opleidingsniveau vader	havo, mbo 4 of vwo	hbo of universiteit	-0,18	*	-0,35	***	-0,12	
Opleidingsniveau vader	vmbo, mavo, mbo 1, 2 of 3	hbo of universiteit	-0,26	**	-0,28	**	-0,13	

significantiëniveau alfa ( $\alpha$ ): \*\*\* =  $\alpha < ,001$ ; \*\* =  $,001 \leq \alpha < ,01$ ; \* =  $,01 \leq \alpha < ,05$

## 5.7 Samenhang tussen migratieachtergrond en andere achtergrondvariabelen

Er zijn verschillen waargenomen tussen de toetscores van kandidaten met verschillende migratieachtergronden. De kandidaten met twee in Nederland geboren ouders halen gemiddeld de hoogste scores. De kandidaten met een niet-westerse migratieachtergrond scoren gemiddeld het laagst. De vraag die volgt, is of deze verschillen blijven bestaan als er voor het effect van de andere variabelen gecorrigeerd wordt. Het zou kunnen zijn dat relatief veel kandidaten uit de groep met een niet-westerse migratieachtergrond tot een groep behoren die vanwege andere kenmerken relatief laag presteert. In de vorige paragraaf is dit al onderzocht voor de achtergrondvariabele vooropleiding. Daarbij werd duidelijk dat voor alle groepen die op basis van hun vooropleiding worden onderscheiden geldt dat kandidaten met een niet-westerse achtergrond gemiddeld lager scoren dan kandidaten waarvan beide ouders in Nederland zijn geboren.

We hebben onderzocht of de verschillen tussen de groepen kandidaten met verschillende migratieachtergronden ook blijven bestaan als we naar de andere achtergrondvariabelen kijken zoals gesproken taal, eigen ervaring in het basisonderwijs, leeservaring, betrokkenheid met een bepaalde regio of de opleiding van de ouders. Voor alle niveaus (waarden) van de variabelen zijn de gemiddelde vaardigheden bepaald voor de drie groepen die op basis van hun migratieachtergrond worden onderscheiden. De uitkomsten zijn gegeven in bijlage 6 (voor Geschiedenis), in bijlage 7 (voor Aardrijkskunde) en in bijlage 8 (voor Natuur & Techniek). In de tabellen zijn naast de gemiddelden en de standaardafwijkingen ook de aantallen en percentages kandidaten gegeven. De percentages tonen de verdeling van de kandidaten over de onderscheiden groepen.

Voor een aantal groepen is het lastig om de verdelingseigenschappen goed te bepalen vanwege de lage aantallen observaties. Dat is regelmatig het geval bij de groep kandidaten met een westerse migratieachtergrond omdat deze groep op zich al niet heel groot is. Als deze groep kandidaten verder opgedeeld wordt in subgroepen binnen een bepaalde categorie dan speelt dat natuurlijk nog meer. Dit geldt bijvoorbeeld ook voor de groep kandidaten met een niet-westerse migratieachtergrond met een dyslexieverklaring<sup>8</sup>.

Aangezien de belangrijkste vragen rond dit onderzoek vooral het contrast tussen de kandidaten zonder en die met een niet-westerse migratieachtergrond betreffen, concentreren we ons in deze paragraaf op gegevens daarover. Het overzicht van de hoofdeffecten staat in tabel 5.9. Een gemiddeld effect is het effect dat gevonden wordt als over alle niveaus (waarden) binnen de variabele bekeken wordt wat de effectgrootte is tussen de twee groepen kandidaten. Er wordt hierbij gewogen voor de populariteit van de categorieën binnen de variabelen. Een voorbeeld van de manier waarop dit is berekend voor Aardrijkskunde en de variabele opleiding staat in bijlage 9.

Tabel 5.9: Effectgrootte geen migratieachtergrond – niet-westerse migratieachtergrond per achtergrondvariabele per vaardigheid

Geen – Niet-westers	Gemiddeld effect			Verskil met totaal effect		
	AK	GS	NT	AK	GS	NT
<b>TOTAAL</b>	0,52	0,31	0,87			
<b>Afnameperiode</b>	0,53	0,31	0,87	0,01	0,00	-0,01
<b>Dyslexie</b>	0,51	0,29	0,84	-0,01	-0,01	-0,03
<b>Geslacht</b>	0,45	0,25	0,84	-0,07	-0,05	-0,04
<b>Leeftijd</b>	0,74	0,42	0,88	0,22	0,12	0,01
<b>Opleiding</b>	0,51	0,31	0,80	-0,01	0,01	-0,07
<b>Taal met Ouders</b>	0,23	0,34	0,75	-0,29	0,04	-0,12
<b>Taal met Bekenden</b>	0,47	0,31	0,86	-0,05	0,01	-0,01
<b>Hoe lang BO</b>	0,48	0,28	0,85	-0,04	-0,03	-0,02
<b>Lezen boek</b>	0,49	0,29	0,79	-0,03	-0,02	-0,09
<b>Lezen krant</b>	0,56	0,32	0,93	0,04	0,02	0,06
<b>Lezen nieuw internet</b>	0,55	0,44	0,92	0,04	0,13	0,04
<b>Betrokkenheid</b>	0,51	0,24	0,87	-0,01	-0,06	0,00
<b>Opleiding moeder</b>	0,56	0,41	0,90	0,04	0,10	0,03
<b>Opleiding vader</b>	0,66	0,39	0,93	0,14	0,08	0,05

\* Zie de bijlagen voor de verschillende niveaus binnen de variabelen

<sup>8</sup> Bij de groep kandidaten zonder migratieachtergrond heeft over de drie vakken heen 7,8% een dyslexieverklaring, bij de groep kandidaten met een westerse migratieachtergrond is dat 5,1% en bij de kandidaten met een niet-westerse migratieachtergrond 2,6%. Of dat betekent dat dit de mate van dyslexie weerspiegelt, of dat het betekent dat de kandidaten zonder migratieachtergrond beter de weg in het landschap van de dyslexieverklaringen kent, is natuurlijk een heel andere, zij het zeker relevante discussie.



Al met al laten deze resultaten zien dat het corrigeren over het algemeen geen grote impact heeft. In een aantal gevallen zien we dat het effect kleiner wordt, zoals na correctie voor de taal die met de ouders gesproken wordt, met name bij Aardrijkskunde. Het omgekeerde effect zien we bij leeftijd en opleiding ouders. De verschillen zijn over het algemeen beperkt.

Tot slot zijn er meerdere lineaire regressieanalyses uitgevoerd bij alle drie de toelatingstoetsen. In die modellen is gewerkt met alle mogelijke achtergrondvariabelen, en is bekeken welke variabelen nog effect hebben wanneer voor de overige variabelen gecorrigeerd wordt. In deze analyses worden alleen de kandidaten meegenomen bij wie voor alle variabelen waarden bekend zijn. Dat zijn bij Geschiedenis 749 toetsresultaten, bij Aardrijkskunde zijn dat er 785, en bij Natuur & Techniek 851. In alle gevallen blijkt migratieachtergrond een effect te hebben op de toetsresultaten. De variabelen die daarnaast bij alle vaardigheden effect hebben, zijn leeftijd, geslacht en vooropleiding. Bij alle drie de getoetste vakken presteren de oudere kandidaten iets beter dan de jongere, de mannen iets beter dan de vrouwen en kandidaten met een havo-opleiding beter dan kandidaten met een mbo-zorg & welzijnopleiding tot onderwijsassistent. Bij de vakken Geschiedenis en Aardrijkskunde heeft ook ervaring met het basisonderwijs impact: kandidaten die acht jaar in Nederland basisonderwijs hebben gevolgd, halen hogere scores. Bij Geschiedenis heeft ook leesgedrag (vaak boeken lezen) een positief effect, evenals de opleiding van de vader. Bij Aardrijkskunde presteren de kandidaten die thuis vooral Nederlands spreken wat beter. De resultaten van deze analyses sluiten hiermee aan bij de gepresenteerde resultaten in paragraaf 5.6.

## 5.8 Conclusies op het niveau van de toets

De gemiddelde prestaties van kandidaten zonder migratieachtergrond, met een westerse migratieachtergrond en met een niet-westerse migratieachtergrond, verschillen bij alle drie de toelatingstoetsen: Geschiedenis, Aardrijkskunde en Natuur & Techniek. Deze verschillen blijven zichtbaar als er wordt gecorrigeerd voor de effecten van de (andere) achtergrondkenmerken van de kandidaten.

Het verschil tussen kandidaten zonder migratieachtergrond en kandidaten met een westerse migratieachtergrond is vrij constant over de verschillende vakken. Bij alle drie de vakken kan het verschil getypeerd worden als een 'klein negatief effect'.

Als we de scores van kandidaten zonder migratieachtergrond vergelijken met de kandidaten met een niet-westerse migratieachtergrond, dan vinden we grotere verschillen. De effectgrootte van het verschil is klein bij het vak Geschiedenis. Een matig verschil wordt gevonden bij het vak Aardrijkskunde. Een groot verschil, in termen van effectgrootte, wordt gevonden bij het vak Natuur & Techniek.

In het volgende hoofdstuk wordt gekeken naar de mogelijke invloed van kenmerken van afzonderlijke opgaven op de gevonden verschillen.





# Onderzoek naar DIF

## 6.1 Inleiding

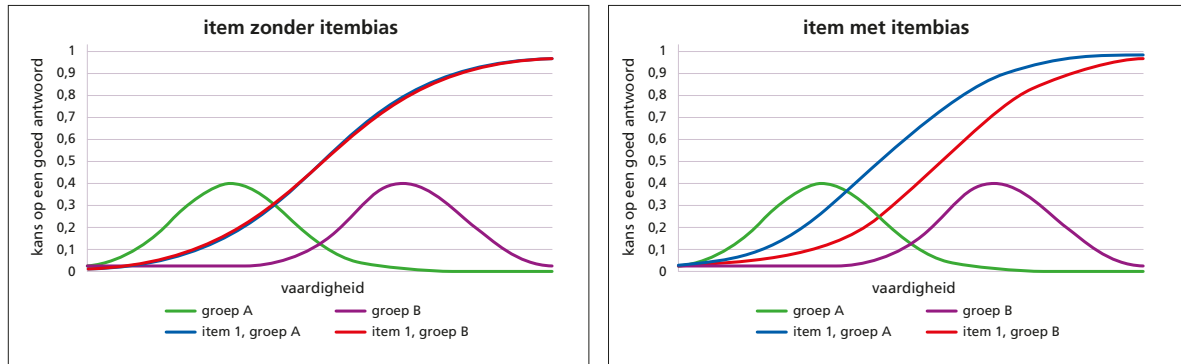
In hoofdstuk 3 is al een korte introductie van differential item functioning, of DIF, gegeven. DIF-onderzoek betreft altijd een specifiek kenmerk, zoals gender of migratieachtergrond. In onderzoek naar DIF worden de personen die de test gemaakt hebben op basis van dat specifieke kenmerk verdeeld in groepen. Ongeacht tot welke groep personen behoren geldt: wanneer twee personen exact hetzelfde niveau van kennis of vaardigheid hebben, zouden ze ook exact dezelfde kans moeten hebben om een item dat die kennis of vaardigheid meet, goed te maken. Verschillen personen in hun kennisniveau of vaardigheid, dan verschillen de kansen om het item goed te maken: de vaardigere persoon heeft een grotere kans om het item goed te maken dan de minder vaardige persoon. In het vervolg van dit stuk hebben we het over 'vaardigheid', dit is een algemene term waaronder ook het kennisniveau van een kandidaat valt.

In figuur 6.1 zijn verdelingen gegeven van groep A en groep B, waarbij groep A minder vaardig is dan groep B. De x-as in deze figuren toont het verloop van het vaardigheidsniveau. De y-as geeft voor de groepen A en B de vaardigheidsverdeling weer en voor de items 1 en 2 de kans op een goed antwoord. In de linkerfiguur is item 1 zonder bias weergegeven. De relatie tussen de vaardigheid en de kans om item 1 goed te maken, is voor beide groepen gelijk. Alleen de vaardigheid bepaalt de kans om dit item goed te beantwoorden. Dit voorbeeld laat ook zien dat het ontbreken van itembias niet betekent dat evenveel personen in groep A als in groep B de opgaven goed maken. Het ontbreken van itembias betekent alléén dat de prestatie van een kandidaat alleen van zijn/haar vaardigheid afhangt. De proportie personen uit groep A dat dit item goed zal maken zal lager liggen dan de proportie van groep B die het goede antwoord geeft doordat de leden van groep B de gemeten vaardigheid gemiddeld beter beheersen.

In het geval van het rechteritem 2 mét bias zien we dat niet alleen de vaardigheid er toe doet, maar ook welke groep het item maakt. De relatie tussen de vaardigheid en de kans om item 2 goed te maken, verschilt voor de twee groepen. Twee personen met exact dezelfde vaardigheid hebben dan niet meer een gelijke kans om dit item goed te maken. Een persoon uit groep A, met exact dezelfde vaardigheid als een persoon uit groep B, heeft ondanks de gelijke vaardigheid (ongeacht of die gelijke vaardigheid nu groot of klein is) toch een grotere kans om deze opgave goed te maken. Bij deze vorm van DIF spreken we van 'uniforme' itembias.

Als we kijken naar de proporties goede antwoorden in groep A en groep B, dan zullen die voor het rechteritem minder van elkaar verschillen dan bij het linker item. Als de bias van dit item omgekeerd is (ondanks gelijke vaardigheid toch gemakkelijker voor B dan voor A, oftewel de rode lijn ligt hoger dan de blauwe lijn), dan zal de proportie goede antwoorden in groep B veel groter zijn dan in groep A: groep B is vaardiger én het item is 'partijdig' in het voordeel van groep B.

Figuur 6.1 Illustratie van itembias



Om itembias te onderzoeken worden kandidaten niet alleen verdeeld op basis van het onderzochte kenmerk, maar ook gegroepeerd in groepen met een gelijke vaardigheid. Deze vaardigheid wordt bepaald met behulp van de score op de toets. Het meeste onderzoek naar itembias wordt uitgevoerd zonder extern vaardigheidscriterium: de toetsscore of de vaardigheidsscore weerspiegelt dan de vaardigheid. Binnen de groepen van gelijke vaardigheid wordt vervolgens gekeken naar de subgroepen waar het biasonderzoek zich op richt. In ons onderzoek betreft dit de etnische achtergrond, waarbij we nu alleen het onderscheid tussen de kandidaten zonder migratieachtergrond en de kandidaten met een niet-westerse migratieachtergrond, onderzoeken.

In hoofdstuk 5 zagen we dat de vaardigheid van de groep zonder migratieachtergrond hoger lag dan die van de groep kandidaten met een westerse migratieachtergrond en die van de groep kandidaten met een niet-westerse migratieachtergrond. Als een groep vaardiger is dan een andere groep, dan houdt dat in grote lijnen in dat alle opgaven voor de vaardigere groep makkelijker zijn. Dit geldt als de opgaven op dezelfde manier functioneren en er dus geen sprake is van itembias.

Er zijn ook methoden die niet met de totaalscore werken maar alleen kijken naar de verhouding in itemmoeilijkheid per itempaar. Daarbij wordt gebruik gemaakt van IRT, waarin voor de items een moeilijkheidsparameter geschat kan worden. Als er geen itembias is, dan geldt voor alle itemparen dat de afstand tussen die parameter bij een itempaar gelijk is voor de verschillende groepen. Als de afstand verschilt, is dat een opvallend itempaar. Items die deel uitmaken van veel itemparen waarbij verschillende afstanden gevonden worden, zijn opvallende items.

Tot slot moet opgemerkt worden dat DIF of relatief veel schendingen bij itempaarevaluatie een statistische observatie is. Verder inhoudelijk onderzoek van de opgaven zelf moet duidelijk maken of er werkelijk sprake is van bevoordeling van een groep boven een andere groep. Als dat het geval is, wordt er gesproken over itembias.

## 6.2 Data voor itembiasonderzoek

De data zoals gebruikt in het onderzoek, staan beschreven in hoofdstuk 5. Elke versie van de toelatingstoets bestaat uit 60 meerkeuzevragen (items). Per vak zijn 174 items statistisch onderzocht, wat betekent dat in totaal 522 items onderzocht zijn.

Bij het onderzoek naar bias wordt gekeken of er items zijn waarbij andere factoren dan de te meten vaardigheid een rol spelen bij het goed maken van een item. Het itembiasonderzoek richt zich hier op onderzoek naar bias voor de volgende twee groepen: kandidaten zonder migratieachtergrond en kandidaten met een niet-westerse migratieachtergrond. De kandidaten met een westerse migratieachtergrond zijn niet in dit onderzoek meegenomen omdat het aantal waarnemingen bij die groep te klein is. In tabel 5.1 hadden we al gezien dat voor alle vakken er minder dan 40 toetsafnames waren die gemaakt zijn door kandidaten met een westerse migratieachtergrond. Met vier toetsversies per vak houdt dat in dat er regelmatig minder dan 10 observaties uit die groep per versie gevonden worden. In tabel 6.1 worden de aantallen observaties per item per vak gegeven voor de drie groepen.

Tabel 6.1: Aantal observaties per item per vak per type migratieachtergrond

Aantal observaties per item	Geen migratieachtergrond			Westerse migratieachtergrond			Niet-westerse migratieachtergrond <sup>9</sup>		
	AK	GS	NT	AK	GS	NT	AK	GS	NT
Minimum	172	151	185	5	7	8	18	17	17
Maximum	362	368	407	18	19	23	52	58	48
Mediaan	181	177	199	7	8	11	24	28	24
Gemiddeld	229	218	254	9	11	13	30	31	28

## 6.3 Methode en resultaten van itembiasonderzoek

Er is in dit onderzoek gebruik gemaakt van een drietal verschillende methoden voor DIF-onderzoek. Twee methoden zijn klassieke DIF-onderzoeksmethoden: de Mantel Haenszel-methode (MH) en de Logistische Regressiemethode (LR). Deze twee methoden geven per opgave aan of er statistisch gezien sprake is van DIF. Bij de derde methode wordt gebruik gemaakt van een Rasch-model, dat wil zeggen een IRT model. Deze methode richt zich op de evaluatie van de itemparen<sup>10</sup>. Deze 'itempaar-methode' geeft opvallende itemparen aan, hetgeen betekent dat deze methode niet zozeer DIF-items aangeeft. Wel kan gesteld worden dat een item opvallend is als het in veel opvallende itemparen voorkomt. Er is echter geen statistische grens die aangeeft welk percentage daarvan opvallend moet zijn om te stellen dat een item opvallend is. We hanteren hier de grens van 10%, maar kijken ook naar items waarbij een lager percentage opvallend is. De opvallende items komen in aanmerking om nader bekeken te worden.

- **Mantel Haenszel (MH)**

De DIF-analyse volgens de Mantel Haenszel methode heeft in geen van de onderzochte toetsvarianten DIF-items geïdentificeerd.

- **Logistische regressie (LR)**

De logistische regressieanalyse heeft in totaal 17 (3%) DIF-items geïdentificeerd. Het betreft 3 items bij Geschiedenis, 7 items bij Aardrijkskunde en 7 items bij Natuur & Techniek. Dat betekent dat bij Aardrijkskunde en bij Natuur & Techniek 4% van de items potentieel partijdig is, en bij Geschiedenis minder dan 2%.

- **Differential item pair functioning (itempaar-methode)**

Wanneer we als grens hanteren, dat een opgave nader onderzocht moet worden als een opgave onderdeel is van meer dan tien opvallende itemparen dan zien we dat dat bij Geschiedenis geldt voor zes opgaven. Bij Aardrijkskunde is dat 1 item en bij Natuur & Techniek zijn dat er ook 6 opgaven. Als we de bandbreedte uitbreiden en alle items bekijken die onderdeel zijn van minstens 5 opvallende itemparen, dan zijn er uiteraard meer opgaven per vak die opvallen. Dan zijn er bij Geschiedenis 12 items geïdentificeerd om nader te onderzoeken. Voor Aardrijkskunde zijn dat er ook 12 en voor Natuur & Techniek zijn er 20 items geïdentificeerd.

<sup>9</sup> Er is ook specifiek onderzoek gedaan naar de groep kandidaten met een niet-westerse migratieachtergrond, waarbij beide ouders een niet-westerse achtergrond hadden. Dan blijft 70 tot 75% van de aantallen over, wat betekent dat er voor die groep gemiddeld ongeveer 21 observaties per item zijn. Daar zijn analyses mee uitgevoerd, maar door de lage aantallen observaties leverden die analyses nog minder significante resultaten op. We rapporteren daarom alleen de resultaten voor de groep zoals deze oorspronkelijk is gedefinieerd als niet-westerse migratieachtergrond: kandidaten met minstens één ouder die in een niet-westers land is geboren.

<sup>10</sup> Bechger, T., & Maris, G. (2015). A Statistical Test for Differential Item Pair Functioning. *Psychometrika*, volume 80, pp 317–340. Internet publicatie 2014: <https://link.springer.com/article/10.1007/s11336-014-9408-y>

Van deze drie verschillende methoden is het duidelijk dat de MH-methode het minst gevoelig is. Of de LR of de itempaar methode gevoeliger is, hangt af van waar de grens getrokken wordt om een opgave als opvallend aan te merken. Er zijn in totaal 17 opgaven die door middel van de LR-methode opgepikt worden als DIF-item. Als de grens van 5 opvallende itemparen gebruikt wordt, komen er in totaal 44 opvallende items naar voren bij de itempaar-methode. 16 van de 17 LR-DIF-items horen ook tot deze 44 items. Ligt de grens hoger namelijk bij meer dan 10 opvallende itemparen, dan zijn er dus 13 opvallende opgaven. Van deze 13 opgaven vallen er 11 ook op met de LR-methode.

Bij het vinden van de aantallen DIF-items is Natuur & Techniek het meest opvallende, en Geschiedenis het minst opvallende vak. Dat is ook nog steeds het geval als de bredere selectie genomen wordt op basis van de itemparen, al zijn de aantallen voor Geschiedenis en Aardrijkskunde dan gelijk.

De lage aantallen opvallende opgaven hebben ook te maken met lage aantallen observaties per item voor de groep kandidaten met een niet-westerse migratieachtergrond. Afwijkingen komen daardoor niet snel als statistisch significant naar voren. Het is daarom ook zinvol om naar de p-waarden te kijken. Het evalueren van de p-waarden helpt bij het interpreteren van de DIF. Bij bovenstaande methoden van DIF is niet aangegeven welke groep nu mogelijk benadeeld wordt, en wie mogelijk bevoordeeld.

De p-waarde komt overeen met het percentage goede antwoorden dat op een opgave gegeven wordt. Hoe lager de p-waarde, hoe moeilijker de opgave. De moeilijkheid kan verschillen over groepen als de vaardigheid in de groepen verschilt. In tabel 6.2 zien we dat de gemiddelde p-waarde voor de drie vakken verschilt voor de kandidaten zonder migratieachtergrond en de kandidaten met een niet-westerse migratieachtergrond. Zoals te verwachten op basis van de hogere gemiddelde toetsscore en daarmee de vaardigheid, zijn de gemiddelde p-waarden van de groep kandidaten zonder migratieachtergrond hoger dan die van de groep kandidaten met een niet-westerse migratieachtergrond.

Tabel 6.2: P-waarden in de groep kandidaten zonder en met een niet-westerse migratieachtergrond

	Gemiddelde P-waarde			Reikwijdte verschil in p-waarden	
	Geen	Niet-westers	verschil	90% van de verschillen tussen	
AK	69,4	60,6	8,9	-9,2	26,9
GS	59,0	54,1	4,9	-12,6	22,4
NT	65,1	55,0	10,1	-10,3	30,5

Dit hoeft niet te betekenen dat er sprake is van itembias. Wanneer dit verschil ongeveer gelijk is over alle items, dan is er geen itembias. Wanneer het verschil opvallend anders is bij bepaalde items, dan kan dat een indicatie van itembias zijn. DIF-onderzoek geeft statistisch aan wanneer een verschil opvallend anders is, maar je kan er ook relatief naar kijken: welke 5% van de items bevoordeelt de ene groep en welke 5% de andere groep? Daarvoor wordt het verschil tussen de p-waarden van de twee groepen genomen en wordt gekeken welke opgaven veel beter worden gedaan door de groep kandidaten zonder migratieachtergrond, en welke opgaven beter worden gedaan door de groep kandidaten met een niet-westerse achtergrond. Items waarbij het verschil buiten de 90-procentsrange vallen, zijn opvallende items wat betreft de p-waarden.

### 6.3.1 Opvallende items bij Geschiedenis

In tabel 6.3 zijn vijf opvallende geschiedenisitems gegeven. Van deze items is item 85837 het meest opvallend. Deze opgave valt in alle categorieën (LR, 15% van de itemparen, afwijkend p-waarde verschil). Bij dit item zien we dat het item beter gedaan wordt door de kandidaten met een niet-westerse migratieachtergrond (71% van de kandidaten uit die groep maakt deze opgave goed). Kandidaten zonder migratieachtergrond maken deze opgave minder goed (44% goede antwoorden). Aan de andere kant van zien we item 85471 die weer beter door de kandidaten zonder migratieachtergrond gedaan wordt (72% versus 43%).

Tabel 6.3: Opvallende geschiedenisitems

Item	Aantal	DIF indicatie		Aantal observaties		P-waarden			
		label	versies	LOG	% ip	GMA	NWMA	GMA	NWMA
85146	1	LR	5	151	17	0,65	0,94	-0,29	-3,20
85471	1	_	6	177	30	0,72	0,43	0,28	2,21
85512	2	LR	6	333	50	0,60	0,74	-0,14	-1,77
85837	1	LR	15	191	28	0,44	0,71	-0,27	-3,04
85942	2	_	6	342	45	0,73	0,47	0,26	2,00

Naast deze vijf items zijn alle items die naar voren kwamen uit de verschillende methoden inhoudelijk bekeken.

### 6.3.2 Opvallende items bij Aardrijkskunde

In tabel 6.4 zijn acht opvallende items bij Aardrijkskunde weergegeven. Van deze opgaven is 86753 de meest opvallende. Bijna 2 op de 5 itemparen waar dit item deel van uitmaakt, vallen op als afwijkend. Ook de LR pikt dit op. Deze opgave wordt door zeer veel kandidaten zonder migratieachtergrond goed gemaakt (88%), maar door minder dan de helft (48%) van de kandidaten met een niet-westerse migratieachtergrond. Aan de andere kant is opgave 86317 opvallend. Hier is ruim 1 op de 5 itemparen opvallend, en wordt het item door 77% van de kandidaten met een niet-westerse migratieachtergrond goed gemaakt en door 51% van de kandidaten zonder migratieachtergrond. Ook dit item wordt met LR opgepikt als opvallend.

Tabel 6.4: Opvallende Aardrijkskunde-items

Item	Aantal	DIF indicatie		Aantal observaties		P-waarden			
		label	versies	LOG	% ip	GMA	NWMA	GMA	NWMA
86317	1	LR	21	181	22	0,51	0,77	-0,26	-3,26
86663	2	LR	10	355	46	0,71	0,41	0,30	1,98
86743	1	_	10	172	24	0,86	0,54	0,32	2,15
86753	2	LR	39	346	42	0,88	0,48	0,40	2,93
86842	2	LR	32	362	50	0,57	0,76	-0,19	-2,52
86926	2	_	8	346	42	0,81	0,52	0,28	1,82
86964	1	_	4	181	22	0,56	0,73	-0,17	-2,37
87005	1	_	<3	174	18	0,66	0,83	-0,18	-2,47

Naast deze acht items zijn alle items die naar voren kwamen uit de verschillende methoden inhoudelijk bekeken.

### 6.3.3 Opvallende items bij Natuur & Techniek

Zeven opvallende items bij Natuur & Techniek staan in tabel 6.5. Het meest opvallende item, 82384, is bij 1 op de 3 itemparen waar het deel van uitmaakt als opvallend gevonden, het items is opgepikt door LR-analyse en heeft een afwijkend p-waarde verschil. Deze opgaven wordt door ruim 80% van de kandidaten zonder migratieachtergrond goed gemaakt, maar door slechts 35% van de kandidaten met een niet-westerse migratieachtergrond. Aan de andere kant zien we item 82752. Dit item wordt door 65% van de kandidaten met een niet-westerse migratieachtergrond goed gemaakt, en maar door 41% van de kandidaten zonder migratieachtergrond.

Tabel 6.5: Opvallende Natuur & Techniek-items

Item label	Aantal versies	DIF Indicatie		Aantal observaties		P-waarden			
		LOG	% ip	GMA	NWMA	GMA	NWMA	Verskil	Afwijking
82033	2	LR	25	393	38	0,93	0,68	0,24	1,14
82344	1	LR	6	208	17	0,72	0,41	0,30	1,62
82384	2	LR	34	392	48	0,81	0,35	0,46	2,89
82752	1	LR	10	208	17	0,41	0,65	-0,23	-2,69
82829	2	LR	3	378	45	0,84	0,89	-0,05	-1,18
82910	2	LR	18	393	38	0,67	0,79	-0,12	-1,78
82948	2	LR	4	407	41	0,60	0,63	-0,03	-1,07

Naast deze zeven items zijn alle items die naar voren kwamen uit de verschillende methoden inhoudelijk bekeken.

## 6.4 Verklaringen voor DIF: kijkwijzer

Nadat met behulp van statistische analyses de opgaven zijn opgespoord die verschillend functioneren voor studenten met twee in Nederland geboren ouders en studenten waarvan minstens één ouder in een niet-westers land geboren is, gaan we op zoek naar mogelijke verklaringen voor het afwijkende gedrag van de gevonden 'DIF-opgaven'. We doen dat in drie stappen.

De eerste twee stappen zijn ontleend aan de aanbevelingen van de Educational Testing Service (ETS, 2004). We gaan op zoek naar antwoorden op de volgende twee vragen.

#### 1. Toetst de DIF-opgave kennis en vaardigheden die behoren tot de toelatingstoetsen voor de pabo?

Een opgave die kennis en/of vaardigheden toetst die niet behoort/behoren tot de toelatingseisen voor de pabo, is per definitie niet valide. Of zo'n opgave daardoor extra nadelig is voor kandidaten met een niet-westerse achtergrond, staat daarmee niet vast.

Om vast te stellen wat de opgaven moeten meten, gaan we uit van de inhoud van de Handreikingen aan de kandidaten. Volgens informatie van Cito zijn de Handreikingen ook het criterium aan de hand waarvan de inhoud van een concept-opgave wordt getoetst voordat de toetsen wordt vastgesteld (bijlage 10).

Bij de analyse van de inhoud van de opgaven, hierna 'de toetsdoelen', maken we onderscheid tussen:

a. de kennis die door de opgave wordt getoetst en

b. de taak die door de opgave wordt getoetst.

We gaan hierbij uit van het onderscheid dat Cito bij de constructie maakt tussen opgaven die vragen om 'beschrijven' dan wel 'toepassen'.



De tweede vraag heeft betrekking op de kennis en vaardigheden die kandidaten nodig hebben om de opgaven met DIF goed te beantwoorden.

2. *Is er sprake van construct-irrelevante factoren die nadelig uitpakken voor kandidaten met een niet-westerse achtergrond?*

Om mogelijke 'construct-irrelevante factoren' op het spoor te komen, maken we onderscheid tussen mogelijke:

- a. *'culturele' biasbronnen*: in hoeverre is de context waarin een opgave wordt gesitueerd, 'cultureel bepaald'?
- b. *'talige' biasbronnen*: in hoeverre doet de opgave een beroep op beheersing van aspecten van leesvaardigheid die niveau 3F van het Referentiekader Taal overschrijden? We hanteren beheersingsniveau 3F als 'norm' omdat dat niveau geldt als het eindniveau voor havo- en mbo 4-opleidingen.
- c. biasbronnen die samenhangen met gebruikte *afbeeldingen en illustraties*.

Omdat het in de beperkte context van dit onderzoek niet haalbaar is de uitkomsten van onze analyses empirisch te toetsen, hebben we de uitkomsten van onze inhoudelijke analyses van de opgaven, voorgelegd aan een aantal 'ervaringsdeskundigen': eerstejaars pabostudenten en mbo 4-studenten die zich oriënteren op doorstroming naar de pabo. Daarnaast hebben we een aantal andere deskundigen bevestigd: roc-docenten en vmbo-docenten. Op opbrengsten van deze gesprekken moeten licht werpen op de derde vraag:

3. *In hoeverre sluiten de uitkomsten van onze inhoudelijke analyse van de opgaven met DIF aan bij de ervaringen en opvattingen van een aantal experts, d.w.z. ervaringsdeskundige studenten en docenten?*

De maatregelen die de verdere verspreiding van het coronavirus moesten helpen beperken, hebben het moeilijk gemaakt om in contact te komen met informanten die bereid zouden zijn om een gesprek met ons te voeren. Het aantal ervaringsdeskundigen dat we hebben gesproken, is daardoor beperkt gebleven. De gegevens die hierna worden gepresenteerd over de uitkomsten van de gevoerde gesprekken, moeten dan ook met enige terughoudendheid worden geïnterpreteerd. Hoewel het onmogelijk is om er vergaande conclusies uit te trekken, kunnen de opbrengsten van interviews wel helpen bij de interpretatie van de uitkomsten van de inhoudelijke analyse van de opgaven.

In verband met de noodzakelijke geheimhouding van de opgaven, blijft wat volgt beperkt tot 'globale' gegevens over de geanalyseerde opgaven. Alleen de zes opgaven die tijdens de interviews zijn besproken, worden in dit rapport gepresenteerd in de vorm waarin ze aan de kandidaten worden voorgelegd (zie bijlage 3). Van die zes opgaven is één opgave eerder als 'voorbeeldopgave' gepubliceerd via [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl).

Hierna worden de voornaamste uitkomsten van de inhoudelijke analyse van de opgaven besproken. De aandacht gaat daarbij in de eerste plaats naar de opgaven die als 'opvallend' naar voren komen uit de statistische analyses die in het eerste deel van dit hoofdstuk zijn beschreven. Om de informatie over de 20 meest 'opvallende' opgaven te kunnen duiden, hebben we ten behoeve van de inhoudelijke analyse van deze opgaven ook inzage gekregen in een groter aantal (45) opgaven die in een of twee van de uitgevoerde DIF-analyses als mogelijk 'anders dan andere items' naar voren kwamen. Op verzoek van FLOT heeft Cito daarnaast nog inzage gegeven in 16 opgaven die willekeurig gekozen zijn uit de opgaven die in één van de uitgevoerde analyses naar voren kwamen als opgaven die opvallend zouden functioneren.

## 6.5 Analyse van opgaven Geschiedenis

Er zijn in totaal 25 opgaven van de toelatingstoets Geschiedenis bekeken. In deze paragraaf staan de 5 opgaven (85146, 85471, 85512, 85837 en 85942) centraal die uit de statistische analyses van Cito naar voren zijn gekomen (zie par. 6.3.1). De overige 20 opgaven zijn gebruikt om te evalueren in hoeverre de 5 'opvallende' opgaven kenmerken hebben die maken dat ze in de analyses anders functioneren dan andere opgaven. Van die 20 opgaven zijn er 14 uit één of meer van de analyses naar voren gekomen als 'DIF-items'.

Voor bijna alle geanalyseerde opgaven geldt dat de – vetgedrukte - toetsvraag met drie of vier antwoordmogelijkheden wordt voorafgegaan door een context die bestaat uit een afbeelding en een korte tekst. De afbeeldingen hebben verschillende functies. Bij sommige vragen moet de afbeelding worden gebruikt om tot het juiste antwoord te komen, bij andere opgaven is de afbeelding alleen illustratief.

Tabel 6.5.1 laat zien hoe de 25 opgaven die inhoudelijk zijn geanalyseerd, zijn verdeeld over de verschillende tijdvakken. Hierbij valt op dat tijdvak 7 geen opvallende items oplevert en tijdvak 9 de meeste.

Tabel 6.5.1: Verdeling van de 25 geanalyseerde items over de tijdvakken

tijdvaknummer	aantal geanalyseerde items
1	3
2	2
3	1
4	2
5	3
6	4
7	0
8	1
9	6
10	3

Cito classificeert 13 van de inhoudelijk geanalyseerde 25 opgaven als 'beschrijven' en 12 als 'toepassen'.

### 6.5.1 Inhoudelijke kenmerken van de opgaven Geschiedenis

Van de vijf opgaven die uit de analyses als 'DIF-items' naar voren komen, heeft er 1 betrekking op tijdvak 1, 1 op tijdvak 3, 1 op tijdvak 9 en hebben er 2 betrekking op tijdvak 10. Cito classificeert 4 van deze opgaven als 'beschrijven' en 1 als 'toepassen'.

Een aantal zaken valt op als we de 5 opgaven waarbij Cito constateert dat er sprake is van DIF (zie tabel 6.3) inhoudelijk onder de loep nemen. Opgave 85837 is weergegeven in bijlage 3.

- Bij 3 opgaven (85146, 85512 en 85837) is de DIF in het nadeel van kandidaten met een Nederlandse achtergrond en bij 2 (85471 en 85942) opgaven is de DIF in het nadeel van kandidaten met een niet-westerse achtergrond. Drie van deze items worden dus gemiddeld genomen beter gemaakt door niet-westerse kandidaten.
- Slechts 2 van deze 5 opgaven (85512 en 85146) vragen direct naar de leerstof zoals die in de Handreiking wordt beschreven. Deze 2 opgaven worden door Cito ook allebei geclassificeerd als 'beschrijven'. Voor de andere 3 opgaven geldt dat de kandidaat informatie uit de gegeven context – de tekst en/of de afbeelding - moet gebruiken om de te beantwoorden vraag juist te interpreteren. Bij 1 van die opgaven (85471) wordt een begrip uit de Handreiking gebruikt, zodat die koppeling voor de hand liggend is, maar bij de 2 andere items (85837 en 85942) moet de kandidaat kritischer doordenken. Bij die laatste twee opgaven, moet de leerling de geboden afbeelding analyseren in relatie met de begeleidende tekst en deze koppelen aan de geleerde kennis. Bij één van deze opdrachten (85942) verwoordt de Handreiking de kennis op een wijze die het juiste antwoord onlogisch kan maken. Bij de andere opgave is het juiste antwoordalternatief op grond van de handleiding wel logisch, maar niet vanzelfsprekend.

Deze opgave (85837, bijlage 3) vraagt de kandidaat een op een schoolplaat afgebeelde kerk van Willibrord geografisch te plaatsen. De Handreiking beschrijft (in een zinsnede die tussen haakjes staat) dat Willibrord in de omgeving van Utrecht werkte, maar hoe groot die omgeving precies begrepen moet worden, wordt niet uitgelegd. Een antwoordalternatief als 's-Hertogenbosch kan, zeker als de aspirant-student wel een reproductiekennis, maar geen gevoel voor afstanden in die historische periode heeft, een logisch alternatief zijn.

- Conceptuele kennis, in de zin van meta-conceptuele kennis, speelt in 4 van deze opgaven geen rol. De student moet vooral inhoudelijke kennis gebruiken en er wordt nergens gevraagd om na te denken over zaken als oorzaak-gevolgrelaties of multi-perspectiviteit. In 1 opgave moet een student de geboden afbeelding wel kunnen zien als een gevolg van een in de kennisbasis beschreven ontwikkeling. Deze vraag staat wel als 'beschrijven' gecategoriseerd.

Al met al komt er geen duidelijk patroon uit de analyse van de opgaven met DIF naar voren. De bekeken items vragen met name naar reproductie of naar een toepassing van kennis die in de Handreiking beschreven staat.

### 6.5.2 Itembias of moeilijkheidsgraad?

Van de 20 geanalyseerde opgaven die minder (14) of geen (6) DIF vertonen, zijn er 3 opgaven waarin de opgave andere termen gebruikt dan de Handreiking. Bij één (85439) van die opgaven is het juiste antwoord wel op basis van de aangeboden kennis te geven, bij een andere (84042) gaat het om een andere geografische term waardoor de vraag moeilijker wordt. Deze laatste opgave bevraagt dus een uitgebreidere topografische kennis dan in de beschrijving van de inhoud bij dit specifieke onderdeel is verwerkt. De derde opgave (85502) gebruikt een term die niet voorkomt in de Handreiking. Dit begrip is een substantief historisch begrip en mag niet als gebruikelijke woordenschat verondersteld worden. Daarbij geeft de laatste opgave door middel van een afbeelding informatie die niet in de Handreiking wordt beschreven.

Voor 4 opgaven kan worden gesteld dat de bewoordingen van de Handreiking niet aansluiten bij de vraag. Voor 1 van die opgaven (84030) geldt dat de Handreiking het begrip wel benoemt, maar dit niet uitlegt, terwijl de opgave zelf van de kandidaat verwacht om op basis van een beschrijving aan te geven dat deze term daarnaar verwijst. Bij een andere opgave (85864) wordt een element uit de vraag niet besproken in de Handreiking en bij de laatste twee (85246 en 85982) wordt een oorzakelijk verband in de Handreiking niet expliciet gegeven, terwijl deze in de opdracht wel direct wordt gevraagd.

Van de hiervoor besproken 7 opdrachten werden er 6 gemiddeld beter gemaakt door kandidaten met twee in Nederland geboren ouders dan door kandidaten met een niet-westerse achtergrond. Deze zeven opdrachten zijn verspreid over beide categorieën (4 toepassen, 3 beschrijven), waardoor er geen duidelijke relatie gelegd kan worden met het taxonomische niveau. Slechts bij 2 opgaven kan een culturele voorkennis een rol hebben gespeeld. Vanwege de noodzakelijke geheimhouding kan hier helaas niet verder op worden ingegaan. Opvallend is verder dat commerciële methoden de kennis van de meeste 'toepassingsvragen' als reproductiekennis aanbieden. De 'toepassingsvragen' hebben daardoor vaak meer het karakter van 'beschrijven'.

Qua taal is de context bij de meeste vragen naar onze mening duidelijk. De illustraties leveren soms mogelijk meer problemen op. In de meeste gevallen is de afbeelding letterlijk bedoeld als illustratie bij de vraag en biedt het plaatje geen informatie die de vraag gemakkelijker of moeilijker maakt. In een enkel geval voegt het plaatje - zeker in combinatie met het bijschrift - echter informatie toe die het goed beantwoorden van vraag wel kan bemoeilijken. Zo toont een opgave (85941, bijlage 3) die betrekking heeft op de slachtoffers in het laatste oorlogsjaar, een afbeelding van bevrijdingsfeesten. Die festiviteiten contrasteren scherp met de opgave die vraagt naar een oorzaak van het grote aantal burgerslachtoffers in het laatste oorlogsjaar. Daar komt bij dat de vliegtuigen die te zien zijn op de afbeelding, één van de antwoordalternatieven (bombardementen van de steden) aantrekkelijk maakt.

### 6.5.3 Wat zeggen de experts?

Er zijn twee opgaven voorgelegd aan (aspirant-)pabostudenten en andere experts: de eerder genoemde items 85837 en 85941 (bijlage 3) en item 82802. De kandidaten is gevraagd om deze opgaven 'hardop denkend' te maken. De meeste kandidaten, ook de studenten met twee in Nederland geboren ouders die al geslaagd zijn voor de toelatingstoetsen, zijn niet in staat deze vragen enkel op basis van parate kennis direct (reproductief) te beantwoorden. Zij proberen daarom eerst een aantal antwoordalternatieven weg te strepen.

Ter illustratie bespreken we item 85837 over de locatie van de belangrijkste kerk die Willibrord liet bouwen. Uit de statistische analyse is gebleken dat de niet-westerse kandidaten deze opgave vaker goed beantwoorden dan de kandidaten met twee in Nederland geboren ouders. Er zijn vier antwoordmogelijkheden, vier steden in het huidige Nederland. Tijdens de interviews blijkt dat de studenten met een niet-westerse achtergrond vaker overwegen welke stad het meest centraal gelegen was dan wel het belangrijkste was, vanuit de gedachte dat iemand daar waarschijnlijk een kerk gaat bouwen. Geen van deze kandidaten verwijst naar de herkomst van Willibrord. Met die aanpak komt men dan vaak op het juiste antwoord. Studenten met twee in Nederland geboren ouders, verwijzen daarentegen vaak naar het gegeven dat Willibrord van buiten Nederland kwam en kiezen vanuit die voorkennis voor steden die aan de rand van Nederland liggen. Daarmee komen ze dan uit bij een onjuist antwoord. Niet-westerse studenten nemen in hun antwoord de herkomst van Willibrord minder expliciet mee. Sommigen geven aan dat ze nog nooit van Willibrord gehoord te hebben. Maar hun antwoord op de vraag is correct. Minder voorkennis lijkt bij deze opgave goed uit te pakken.

Het tweede item betreft de oorzaak van een grote sterfte in de bevolking van de steden in West-Nederland aan het einde van de Tweede Wereldoorlog. Bij het voorlezen van deze vraag valt op dat Nederlandse pabostudenten allemaal spontaan de bijbehorende term noemen: 'de hongervinter'. Deze term staat niet in de opgave en lijkt dus een onderdeel van hun referentiekader te zijn. Het begrip verwijst hen meteen naar het juiste antwoord. Dat betekent overigens niet automatisch dat de studenten ook het goede antwoord kiezen. Niet iedere Nederlandse student positioneert de hongervinter tijdens de oorlog, zodat de bekendheid met de hongervinter hen niet altijd naar het juiste antwoord leidt. Sommige kandidaten plaatsen de hongervinter, wellicht onder invloed van de afbeelding, in de eerste naoorlogse jaren. Bij het 'wegstrepen' van foute antwoorden, blijkt vaker hoe kandidaten via onjuiste kennis tot een goed antwoord komen. Zo komen sommige studenten bij deze opgave op grond van onjuiste gegevens – Nederlandse steden zouden alleen aan het begin van de oorlog gebombardeerd en de Jodenvervolging zou ook bij de eerste oorlogsjaren horen – uit bij het goede antwoordalternatief.

De niet-westerse (aspirant-)studenten die we spraken, laten zien hoe hun eigen referentiekader de keuze van een antwoord kan beïnvloeden. Zo laat een van hen bij het hardop denken weten dat zij het begrip 'slachtoffers' uit de vraag, in combinatie met de prent van bevrijdingsfestiviteiten waarop militaire voertuigen zijn te zien, verbindt met oorlogshandelingen. Die associatie, in combinatie met de (juiste) kennis dat er in de Tweede Wereldoorlog bombardementen zijn geweest, leidt haar naar een fout antwoordalternatief. Vooral bij het 'wegstrepen van de foute antwoorden', een strategie die alle geïnterviewden veelvuldig toepassen, lijkt naar voren te komen dat het voor de beantwoording van de vragen relevante referentiekader van niet-westerse aspirant-studenten minder uitgebreid is dan dat van westerse aspirant-studenten. Zoals gezegd pakt dat kleinere referentiekader soms ook gunstig uit.

Met betrekking tot de voorbereiding op de toets benadrukken de eerstejaars studenten dat ze vooral veel gememoriseerd hebben. Echter, zowel de kandidaten die vertellen dat ze lessen gevolgd hebben, als de geraadpleegde docenten geven aan dat deze memorisatie succesvoller wordt als een docent de structuur of het narratief van de te leren stof, heeft toegelicht. Wie voorbereidende cursussen had gevolgd, had het gevoel beter voorbereid te zijn. Opvallend is daarbij dat alle kandidaten desgevraagd aangeven dat de kennis die ze tijdens de voorbereiding op de toetsen hebben verworven, intussen zover is weggezakt dat ze niet verwachten de toets opnieuw te halen als deze overgedaan zou moeten worden.

#### 6.5.4 Conclusie

De meeste opgaven, zowel die met DIF als de andere opgaven, bevragen historische feitenkennis of substantieve kennis, kennis over unieke historische fenomenen. Het feit dat alle eerstejaars pabostudenten die we spraken, aangeven dat 'stampen' de voornaamste succesfactor was geweest bij de voorbereiding op deze toelatingstoetsen, sluit hier bij aan.

Sommige opgaven kunnen niet direct naar de Handreiking worden geleid. Bij de beantwoording van deze opgaven lijken de studenten vooral te gaan 'associëren': ze gaan af op hun 'algemene kennis' omdat de noodzakelijke historische kennis ontbreekt. Omdat deze algemene kennis wel beïnvloed lijkt te zijn door de culturele achtergrond, beïnvloedt deze achtergrond de wijze waarop antwoordalternatieven worden weggestreept. Hoewel dat soms in het voordeel van niet-westerse kandidaten werkt, ontstaat het beeld dat een westers cultureel referentiekader studenten helpt bij het effectiever wegstrepen van foute antwoorden. Een echt historische manier van denken lijkt voor alle studenten die we spraken, te hoog gegrepen.

Uit dit beperkte onderzoek ontstaat de indruk dat er niet zozeer sprake is van een verschil in taalbeheersing tussen de onderscheiden groepen studenten, maar wel in de mate waarin de studenten uit de verschillende groepen relevante voorkennis kunnen inzetten bij het beantwoorden van de vragen. Zo maakt een deel van de opgaven gebruik van begrippen die niet expliciet worden uitgelegd in de Handreiking en die mogelijk cultureel bepaald zijn. Bij andere opgaven lijkt het onderwerp op zichzelf onderdeel te zijn van een cultureel narratief dat westerse kandidaten thuis eerder meekrijgen dan niet-westerse.

De items tonen op zichzelf dan wel nauwelijks bias, maar wanneer de voorbereiding op de toets en de beantwoording ervan worden meegenomen, wordt een mogelijke bias wel duidelijk. Een vak van geschiedenis is sterk verbonden met de culturele achtergrond waarin het functioneert. Deze culturele achtergrond komt in nagenoeg alle opgaven terug. Voor kandidaten die die achtergrond thuis niet hebben meegekregen, is de door alle studenten veelvuldig gehanteerde oplossingsstrategie van het wegstrepen van antwoordalternatieven, minder effectief dan voor andere kandidaten.

## 6.6 Analyse van opgaven Aardrijkskunde

In totaal zijn er in september en oktober 2020 174 verschillende opgaven aan studenten voorgelegd. Gemiddeld scoren niet-westerse studenten 9% lager op deze vragen dan studenten met twee in Nederland geboren ouders. Van de 25 opgaven die in het kader van dit onderzoek inhoudelijk zijn geanalyseerd, zijn er volgens Cito 12 'beschrijvend' van karakter en 13 gericht op 'toepassing'.

De bestudeerde opgaven passen qua vorm in het model van meerkeuzeitems zoals die bekend zijn uit andere toetsen en examens die door Cito zijn ontwikkeld. Bij de doelgroep van de toelatingstoetsen kan dit model ook goed bekend zijn door de beschikbare voorbeeldexamens. Een opgave bestaat doorgaans uit een inleidende tekst die vaak wordt begeleid door een afbeelding of kaartmateriaal, een vetgedrukte vraag en drie of vier antwoordmogelijkheden. Vaak zijn de inleidende teksten en afbeeldingen nodig om studenten tot het goede antwoord te leiden, maar in voorkomende gevallen is een bron alleen illustratief. In een enkel geval is een afbeelding misleidend. Zo wordt er een vraag gesteld over de rivier de Rijn waarbij een afbeelding wordt weergegeven van een woest kolkende rivier (item 86039). Het is zeer de vraag of deze afbeelding studenten helpt om tot het goede antwoord te komen. De afbeelding lijkt eerder te leiden naar een van de afleiders, die betrekking heeft op de ongeschiktheid van de Rijn voor de scheepvaart. De betreffende vraag is inderdaad relatief vaak fout beantwoord, maar vertoont geen DIF.

### 6.6.1 Inhoudelijke kenmerken van de opgaven Aardrijkskunde

Er zijn 8 items (tabel 6.4) waar een opvallende DIF-score (86753, 86964, 86842, 86317, 87005, 86926, 86663) aanleiding geeft om dieper in te gaan op de inhoud en de kenmerken van de opgave. Vier van deze opgaven vragen volgens Cito om 'toepassing'. De andere 4 vragen zijn volgens Cito beschrijvend van karakter. Verder zijn 4 vragen beter gemaakt door studenten met twee in Nederland geboren ouders en zijn 4 vragen beter gemaakt door studenten met een niet-westerse achtergrond. Uiteraard zijn de vragen met DIF geen representatief sample van het totaal opgaven. Van alle 174 Aardrijkskunde-items die tijdens de twee onderzochte toetsvensters aan kandidaten zijn voorgelegd, zijn er slechts 34 die beter zijn gemaakt door studenten met een niet-westerse achtergrond.

In de vier items met DIF die door Nederlandse studenten beter zijn gemaakt, is het verschil gemiddeld 32,6% ten opzichte van studenten met een niet-westerse migratieachtergrond. Omgekeerd worden items die beter worden gemaakt door kandidaten met een niet-westerse achtergrond, gemiddeld slechts 19,9% beter gescoord.

Vragen met DIF zijn niet gebonden aan één specifiek kernthema uit de Handreiking. Opvallend is dat de vraag die de hoogste DIF-score heeft gehaald (en daarmee dus het item dat zich het meest afwijkend gedraagt) nauwelijks past binnen de Handreiking. Het betreffende item, 86753, gaat in op de aardbevingen in Groningen. Het goede antwoord luidt 'door het delven van grondstoffen uit de bodem van Groningen'. Het is opmerkelijk dat deze opgave door de constructeurs wordt verbonden aan het thema 'bestaansmiddelen'. Mogelijk zou deze opgave inhoudelijk beter passen bij thema 7, 'energiebronnen'. In de Handreiking worden aardbevingen echter vooral in relatie tot 'platentektoniek' besproken. Waarschijnlijk zullen juist de studenten die de Handreiking goed bestudeerd hebben, het antwoord in de verkeerde richting zoeken en uitkomen bij de afleider die verwijst naar bewegingen van aardplaten in de Groningse bodem. Dat effect kan versterkt zijn doordat voorbeeldvraag 8 uit de Handreiking Aardrijkskunde bijna identiek is aan deze vraag. In opgave 86753 is er bovendien sprake van verwarrend taalgebruik. Het 'delven van grondstoffen' waar in de vraag naar verwezen wordt, wordt door veel sprekers van het Nederlands niet automatisch geassocieerd met wat doorgaans wordt aangeduid als 'aardgaswinning'. Het ziet ernaar uit dat vooral de kandidaten met de niet-westerse achtergrond 'last' hebben van de discutabele kenmerken van deze opgave. Slechts 44% van hen kiest het goede antwoord. Van studenten met twee Nederlandse ouders kiest 88% het juiste antwoord. Waarschijnlijk zijn zij zodanig goed bekend met de gevolgen van de aardgaswinning in Groningen dat ze zich zelfs door de verraderlijke afleider die verwijst naar aardplaten en de discutabele formulering 'het delven van grondstoffen' in plaats van het gangbare 'winnen van aardgas', niet op het verkeerde been laten zetten.

De vier vragen met DIF die beter gemaakt worden door studenten met een niet-westerse achtergrond, zijn alle vier te relateren aan een referentiekader dat niet heel specifiek 'Nederlands' is. Het betreft vragen over gastarbeiders (item 86964), de verspreiding van talen in Afrika (item 86842), klimaatgebieden in Afrika (item 86317) en werkzaamheden in de informele sector (item 87005). Het zou echter te ver gaan om hier te spreken van een niet-westers referentiekader.

### 6.6.2 Itembias of moeilijkheidsgraad

Drie van de vier vragen die beter zijn gemaakt door studenten met een niet-westerse achtergrond zijn volgens Cito gesteld op toepassingsniveau. Slechts één vraag vraagt om een beschrijving. In de vragen met DIF die door Nederlandse studenten beter zijn gemaakt, zijn beschrijvende vragen juist in de meerderheid. Het is op dit punt niet mogelijk om te achterhalen of dit voor de totale itembank geldt. De indeling van de vragen naar toepassing en beschrijving lijkt vrij willekeurig. Zo is de vraag over talen in Afrika gelabeld als 'toepassing' terwijl het voor veel studenten een simpele reproductievraag zal zijn. Een toepassingsvraag over het stroomgebied van de Mississippi (86852) is feitelijk een verkapte topografievraag en bevat nauwelijks toepassings-elementen. Omgekeerd is een vraag over platentektoniek (86926, bijlage 3) gelabeld als 'beschrijvend' terwijl er óók argumenten zijn om dit als toepassingsvraag te omschrijven.

### 6.6.3 Wat zeggen de experts?

Twee vragen zijn met studenten besproken. Beide vragen werden door studenten met een niet-westerse achtergrond aanmerkelijk slechter gemaakt dan door de studenten met twee in Nederland geboren ouders. Opvallend is dat bij beide items de gegeven context als problematisch wordt ervaren door de geïnterviewde docenten.

De eerste besproken vraag gaat over platentektoniek in Zuidoost-Azië (item 86926, bijlage 3). Waar Nederlandse studenten deze vraag in 81% van de gevallen goed beantwoorden, geldt dit voor studenten met een niet-westerse achtergrond voor slechts 52%. De vraag toetst volgens de constructeurs 'beschrijven'. Het item wordt begeleid door een weinig informatieve kaart van Zuidoost-Azië waarop de belangrijkste plaatgrenzen zichtbaar zijn. In deze vraag wordt ervan uitgegaan dat een student herkent om welk gebied het gaat en welke plaatbewegingen daar plaatsvinden. Dat vereist geografische inzichten waar de keuzemogelijkheden echter nauwelijks een beroep op doen. Er wordt hier naar gecompliceerde kennis gevraagd buiten een algemeen referentiekader. Hoewel de kaart door een geïnformeerde lezer ongetwijfeld juist geïnterpreteerd kan worden, mag worden aangenomen dat hij voor veel aspirant-studenten minder duidelijk is. Hij voldoet in elk geval niet aan de eisen die volgens de Handreiking Aardrijkskunde aan een goede kaart moeten worden gesteld: *"Om een kaart te kunnen lezen heeft elke goede kaart een legenda, schaal, windroos en titel (blz. 10 van de Handreiking)"*. Alle genoemde onderdelen van een goede kaart ontbreken in deze opgave. Erger nog, de plaatbewegingen in de regio worden niet weergegeven in de kaart terwijl de plaatbeweging feitelijk de kern van de vraag betreft. De geïnterviewde eerstejaars pabostudenten én de mbo'ers wisten desondanks in de meeste gevallen het gebied wel te herkennen, maar ze konden geen van allen uit de voeten met de weergegeven plaatgrenzen. De context die de kaart had willen bieden, werd door studenten dan ook niet herkend.

Wat in dit geval helpt om tot het goede antwoord te komen, is een globaal wereldbeeld en kennis van platentektoniek. De aspirant-studenten die het goede antwoord wisten, gaven aan dat zij dit wisten doordat zij er een documentaire over hadden gezien of een bericht uit het journaal. Juist platentektoniek is een onderdeel dat op het vmbo nauwelijks wordt behandeld in de Aardrijkskundelessen. Niet vreemd dus dat studenten met een mbo-achtergrond hier grote moeite mee hebben. Aspirant-studenten met een niet-westerse achtergrond zijn relatief sterk vertegenwoordigd in de groep die via het mbo bij de toelatingstoetsen voor de pabo belandt. Platentektoniek speelt slechts een zeer beperkte rol in het vmbo-programma. In het meest gunstige geval hebben deze studenten 3 of 4 lessen in de eerste klas gehad over platentektoniek en vulkanisme. Niet vreemd dus dat deze aspirant-studenten hier relatief vaak onderuit gaan.

Bij item 86663 (bijlage 3) lijkt een op zichzelf relevante tekst te zorgen voor problemen bij de geïnterviewde studenten. In de betreffende vraag dienen kandidaten te bepalen uit welke grondsoort de turf die in het verleden als brandstof werd gebruikt, gewonnen werd. De vraag bevat een korte inleidende tekst waarin wordt uitgelegd dat turf lange tijd als brandstof diende maar dat het daarvoor eerst gedroogd moest worden. De tekst wordt begeleid door een zwart-wit foto van een gebied waar turf wordt gestoken en een rustende turfstecker.

Bij de bespreking van deze opgave met kandidaten komen verschillende problemen naar voren. Vrijwel alle studenten die we naar deze vraag lieten kijken, hadden grote moeite met het beantwoorden ervan. Dat geldt ook voor de studenten die de toelatingstoets Aardrijkskunde gehaald hadden en de havostudenten die die toets niet hadden hoeven te maken omdat aardrijkskunde in hun examenpakket had gezeten. De problemen lijken ook niet specifiek voor studenten met een niet-westerse achtergrond. Het woord 'turf' is nagenoeg bij alle studenten onbekend. Het komt in de Handreiking ook niet voor. Omdat er in de Handreiking wel gesproken over brandstof uit veen, past de vraag deels nog steeds binnen de grenzen van de beschreven leerstof. Uit onze interviews blijkt dat studenten de opgave over de hen onbekende 'turf' proberen te beantwoorden door in de afbeelding en in hun eigen voorkennis op zoek gaan naar informatie over grondsoorten. Zand en klei vallen dan bij de meeste studenten snel af doordat ze twijfelen aan de brandbaarheid van deze grondsoorten. De meeste studenten komen dan uit op de keuze tussen löss en veen. Een beperkte kennis van de verspreiding van veen en löss in Nederland en de onbekendheid met turf zorgen er dan voor dat de meeste studenten dan alsnog de fout ingaan. Löss blijkt hier als grootste onbekende het aantrekkelijkste antwoordalternatief.

## 6.6.4 Conclusie

Al met al vinden we in de geanalyseerde opgaven nauwelijks kenmerken die wijzen op niet-westerse bias. De gesprekken met de studenten doen weliswaar vermoeden dat persoonlijke ervaringen en 'everyday knowledge' die studenten al verworven hebben voorafgaand aan de toets, onevenredig bijdragen aan het wel of niet slagen voor de toets, maar we vinden niet veel gegevens die wijzen op culturele itembias. Misschien is het geen toeval dat de paar vragen die opvallend goed gemaakt worden door studenten met een niet-westerse achtergrond, te relateren zijn aan een niet-westers referentiekader maar 4 opgaven – van de 174 – zijn er veel te weinig om hard te maken dat er sprake zou zijn van bias. Ook die ene opgave over aardbevingen in Groningen die opvallend veel makkelijker lijkt voor Nederlandse studenten, kan niet dienen als voldoende onderbouwing.

Een tweede relevante bevinding moet zijn dat de contexten die bij de opgaven worden geschetst, in veel gevallen belangrijker blijken dan de itemconstructeurs voor ogen zal hebben gestaan. Een afbeelding die slechts ter illustratie wordt toegevoegd, wordt door de studenten zeer serieus benaderd alsof daar het antwoord in te vinden is. Grotere zorgvuldigheid in het gebruik van tekst en afbeelding kan verwarring voorkomen zodat een zuiverder toetsing mogelijk is. Daarbij past uiteraard ook dat kaarten voldoen aan de eisen die in de Handreiking zelf worden vermeld.

## 6.7 Analyse van opgaven Natuur & Techniek

Er zijn in totaal 31 opgaven van de toelatingstoets Natuur & Techniek geanalyseerd, waarvan 17 opgaven biologiekennis en 14 opgaven techniekennis toetsen. Cito classificeert 10 van deze opgaven als 'beschrijven' en 21 als 'toepassen'.

De opgaven bieden doorgaans weer een context die bestaat uit een afbeelding en een korte tekst. De vraag is vetgedrukt en er zijn meestal vier antwoordmogelijkheden. De afbeeldingen hebben verschillende functies. Bij sommige vragen moet de afbeelding geanalyseerd worden, bij andere vormt die vooral een illustratie. Soms dienen de afbeeldingen als antwoordmogelijkheden.

### 6.7.1 Inhoudelijke kenmerken van de opgaven Natuur & Techniek

Er zijn 7 opgaven (zie tabel 6.5) waarbij opvallende DIF is geconstateerd, waarvan 3 opgaven biologiekennis en 4 opgaven techniekennis toetsen. Cito classificeert 2 van deze opgaven als 'beschrijven' en 5 als 'toepassen'. Er zijn 3 opgaven waarbij kennis van specifieke terminologie essentieel is om de vraag goed te kunnen beantwoorden. Er zijn 3 opgaven die het toepassen van conceptuele kennis toetsen. Bij 3 opgaven is de DIF in het voordeel van kandidaten met een Nederlandse achtergrond en bij 4 opgaven is de DIF in het voordeel van kandidaten met een niet-westerse achtergrond.

Van de 7 opgaven met opvallende DIF zijn er 2 niet te herleiden naar de Handreiking. Deze opgaven gaan over aquaplaning en een pijnboom. Beide opgaven worden door Cito geclassificeerd als 'toepassen', maar de Handreiking biedt onvoldoende kennis om deze opgaven via toepassing van die kennis te kunnen beantwoorden. Beide opgaven zullen echter niet moeilijk zijn om te beantwoorden voor kandidaten die de beschreven context herkennen. De opgave over aquaplaning, een context die men bij rijles leert kennen, wordt nadrukkelijk beter gemaakt door kandidaten met een Nederlandse achtergrond. De opgave over de pijnboom, een context die typisch is voor het Middellandse Zee- of ander seizoensmatig droog gebied, wordt nadrukkelijk beter gemaakt door kandidaten met een niet-westerse achtergrond.

Van de 7 opgaven met DIF is er dan nog 1 waarbij het herleiden naar de Handreiking moeilijk is. Het gaat hier om het toepassen van conceptuele biologiekennis. De opgave wordt vergelijkbaar gemaakt door beide groepen kandidaten, maar aangezien de kandidaten met een niet-westerse achtergrond de toets over het algemeen slechter maken, komt deze opgave naar boven als vraag met DIF.



Van de overige 4 opgaven met DIF, die wel goed kunnen worden herleid naar de Handreiking, classificeert Cito er 2 als 'beschrijven' en 2 als 'toepassen'. De 2 opgaven die 'beschrijven' toetsen, vereisen het begrijpen van terminologie, te weten 'mechanische', 'enzymatische' en 'bacteriële spijsvertering' (DIF in voordeel van kandidaten met een niet-westerse achtergrond) en 'stollen', 'condenseren', 'verdampen' en 'vervluchten' (DIF in het voordeel van kandidaten met twee in Nederland geboren ouders). De 2 opgaven die 'toepassen' toetsen vereisen het kunnen toepassen van conceptuele kennis, te weten de tamelijk abstracte concepten 'dichtheid' (DIF in voordeel van kandidaten met een niet-westerse achtergrond) en reflectie van verschillende kleuren licht (DIF in voordeel van kandidaten met twee in Nederland geboren ouders).

Er komt geen duidelijk patroon uit de analyse van de opgaven met DIF naar voren. Alleen de opgaven over aquaplaning en de pijnboom zouden mogelijk een culturele biasbron kunnen bevatten. Ook hier zijn echter verschillende interpretaties mogelijk. Kandidaten met een Nederlandse achtergrond starten mogelijk eerder met rijlessen dan kandidaten met een niet-westerse achtergrond. Dit zou kunnen samenhangen met culturele kenmerken van de onderscheiden groepen - een rijbewijs is in de ene groep misschien minder belangrijk dan in de andere -, maar het verschil zou ook kunnen samenhangen met sociaaleconomische verschillen. De kosten van het halen van een rijbewijs zijn immers hoog. Voor de andere opgave geldt misschien dat een pijnboom herkenbaarder is voor kandidaten met een migratieachtergrond in gebieden met seizoensmatig droog klimaat, zoals Marokko en Turkije. Aan de andere kant zou ook kunnen worden aangenomen dat juist kinderen van in Nederland geboren ouders wat vaker op vakantie zijn geweest in streken waar zo'n bomen voorkomen. De opgaven met DIF zijn ook niet bijzonder talig en er is geen sprake van afbeeldingen die tot bias zouden kunnen leiden.

### 6.7.2 Itembias of moeilijkheidsgraad?

Van de 24 geanalyseerde opgaven die geen opvallende DIF vertonen, zijn er 7 waarbij kennis van specifieke terminologie essentieel is om de vraag goed te kunnen beantwoorden. Daarnaast zijn er 11 opgaven die het toepassen van conceptuele kennis toetsen, en deze worden veelal ook als 'toepassen' geclassificeerd door Cito. Er zijn 8 opgaven niet te herleiden tot de Handreiking en bij 4 opgaven is dat moeilijk.

Net als bij de opgaven die naar voren kwamen uit de DIF-analyses, treden ook in de andere 24 opgaven die we hebben bekeken, de grootste verschillen in score op bij vragen die Cito classificeert als 'toepassen'. Het gaat opnieuw om vragen waarbij het herkennen van de context een cruciaal verschil maakt, of waarbij abstracte concepten moeten worden toegepast in een tamelijk abstracte context. Méér dan bij de opgaven met opvallende DIF valt op dat grote verschillen in scores optreden bij opgaven die niet of moeilijk te herleiden zijn naar de Handreiking, en waarbij dus herkenning van de context essentieel is.

Tot slot valt op dat er bij de 6 opgaven waar kandidaten met een Nederlandse achtergrond het hoogst op scoorden, steeds een groot verschil in score met kandidaten met een niet-westerse achtergrond zichtbaar is. Ook hier kunnen de context die in de opgave wordt geschetst en het referentiekader van de kandidaten een rol spelen, zoals kennis van Neanderthalers en EHBO.

### 6.7.3 Wat zeggen de experts?

Er zijn twee opgaven voorgelegd aan kandidaten en andere experts. Item 82802 (bijlage 3) vraagt uit vier afbeeldingen (wind bij een zeilboot, water bij een molen, plantaardige olie in een auto, ijzer dat smelt door brandende steenkool) de *niet-duurzame* energiebron te kiezen. Het is een opgave waarbij kennis van de getoonde contexten essentieel is. Vraag 22 uit voorbeeldtoets 2 (bijlage 3) vraagt de verandering van een schaduw gedurende een dag te bepalen. Ook hier moet gekozen worden uit vier afbeeldingen. Het is een opgave die de kandidaat dwingt een logische redenering op te zetten of te pendelen naar eigen ervaringen.

Hardop denken en andere reacties van kandidaten en andere experts laten zien dat alle kandidaten, ook

die met een Nederlandse achtergrond, de neiging hebben vragen die ze niet direct (reproductief) weten te beantwoorden te benaderen met een associatieve manier van denken.

- Bij item 82802 over niet-duurzame energiebronnen geven voornamelijk kandidaten met een niet-westerse migratieachtergrond aan afgeleid te worden door de auto en het begrip olie, terwijl een kandidaat van Nederlandse afkomst juist op grond van het woord 'plantaardige' dit antwoordalternatief wegstreept. Daarbij is opvallend dat de geïnterviewde kandidaten het begrip 'niet-duurzaam' omschrijven als 'slecht voor het milieu', terwijl het in de opgave (foutief) gebruikt wordt als 'minst klimaatneutraal'. Die subtiele fout in de opgave lijkt een geamplificeerd effect te hebben bij kandidaten met een niet-westerse migratieachtergrond. Ze associëren de auto en plastic flessen plantaardige olie meer met 'slecht voor het milieu'.
- Bij vraag 22 uit voorbeeldtoets 2 over schaduw blijken alle kandidaten, ook die met een Nederlandse achtergrond, niet in staat een goede redenering op te zetten, maar ook niet om een vertaalslag naar de eigen ervaring te maken. Er worden – ook door de pabostudenten die al hebben aangetoond aan de nadere vooropleidingseisen te voldoen - overwegingen genoemd als "hoe hoger de zon staat, hoe langer de schaduw", en "de zon komt op in het Noorden en gaat onder in het Zuiden". Een verklaring voor de gesignaleerde DIF hebben we niet gevonden.

#### 6.7.4 Conclusie

Ook de toetsen Natuur & Techniek bevatten maar heel weinig opgaven die in de statistische analyses opvallen door signalen die wijzen op DIF. Ook de inhoudelijke analyse van de opgaven die wél opvallend gedrag vertonen, levert nauwelijks informatie op die wijst op culturele of talige biasbronnen.

De analyses en de gesprekken met studenten en docenten leveren echter een aantal gegevens op die mogelijk kunnen bijdragen aan een verklaring voor twee opvallende bevindingen die in tabel 5.5 en tabel 5.6 in het vorige hoofdstuk werden gepresenteerd:

- kandidaten met twee Nederlandse ouders én kandidaten met een niet-westerse achtergrond hebben in de twee onderzochte toetsvensters de laagste vaardigheidsscores behaald bij het de toelatingstoets Natuur & Techniek en
- van de drie toelatingstoetsen laat de toelatingstoets Natuur & Techniek het grootste verschil zien tussen de gemiddelde vaardigheidsscores van kandidaten met twee in Nederlands geboren ouders en kandidaten met een niet-westerse achtergrond.

Uit alle gesprekken die we met pabo-studenten en mbo 4-studenten hebben gevoerd, komt naar voren dat de 'natuurwetenschappelijke manier van denken' die in de toets Natuur & Techniek wordt gevraagd, voor veel studenten te hoog is gegrepen. Studenten die bij de andere twee toelatingstoetsen goede resultaten kunnen halen door de leerstof zoals sommigen dat zelf aangaven, 'in hun hoofd te stampen', lijken bij Natuur & Techniek aan te lopen tegen opgaven die méér vragen dan reproductie van kennis die ze hebben opgeslagen. 'Toepassingsvragen' leiden dan niet meer tot toepassing van verworven kennis, maar komen neer op het al dan niet herkennen van de context die in de opgave wordt geschetst. Onze bevinding dat studenten relatief veel moeite hebben met N&T-opgaven die niet of moeilijk kunnen worden herleid naar de Handreiking, zou die hypothese kunnen ondersteunen. We zien ook dat kandidaten als ze het goede antwoord op een opgave niet paraat hebben, of daar niet helemaal zeker over zijn, gaan redeneren en associëren om op basis van hun 'kennis van de wereld' het beste antwoord te vinden. Kandidaten met een niet-westerse achtergrond blijven daarbij opvallend vaak hangen in een soort van 'theoretische' redenering, die ergens uit de bocht vliegt. Kandidaten met twee Nederlandse ouders lijken vaker bewust of onbewust wel in staat om een betekenisvolle relatie te leggen tussen een opgave en dat wat zij wél weten en/of wat ze zelf ervaren hebben. Deze observatie leidt tot de, uiteraard voorzichtige, hypothese dat verschillen tussen de scores van de twee verschillende groepen studenten niet in de eerste plaats veroorzaakt worden door verschillen in grootte van voorkennis of taalbeheersing, maar vooral door de mate waarin die kennis geïntegreerd is, in onderlinge samenhang en met eigen ervaringen. Onze gesprekken doen vermoeden dat de contexten die in de opgaven van de toelatingstoets Natuur & Techniek worden geschetst, voor de kandidaten met een niet-westerse achtergrond wat 'vreemder' zijn dan voor de studenten met twee in Nederland geboren ouders. Het lijkt alsof studenten met een niet-westerse achtergrond hun opgedane kennis vooral hebben 'opgestapeld', terwijl die kennis bij de andere groep studenten wat meer 'verweven' is met wat ze al wisten. In het volgende hoofdstuk gaan we daar verder op in.

## 6.8 Samenvatting en conclusies

In dit hoofdstuk staat de eerste onderzoeksvraag centraal: bevatten de toelatingstoetsen voor de pabo opgaven die studenten met een niet-westerse achtergrond ten onrechte benadelen? De verzamelde gegevens laten zien dat voor alle drie de toelatingstoetsen geldt:

- dat ze weinig opgaven bevatten die DIF-vertonen en
- dat een inhoudelijke analyse van die DIF-opgaven en een aantal aanvullende opgaven én gesprekken met kandidaten en docenten nauwelijks aanwijzingen geven die bevestigen dat westerse biasbronnen de gevallen van 'different item functioning' die wél worden vastgesteld, zouden kunnen verklaren.

De conclusie van dit deel van het onderzoek kan dan ook luiden dat we géén gegevens hebben gevonden die erop wijzen dat de achterblijvende scores op de toelatingstoetsen van kandidaten met een niet-westerse achtergrond verklaard kunnen worden doordat de items van de toelatingstoetsen westerse bias bevatten. Zelfs als we bij de inhoudelijke analyse van de opgaven die 'statistisch' DIF vertonen, bronnen voor een culturele bias over het hoofd zouden hebben gezien, zouden we erop kunnen vertrouwen dat de complete opgavenbanken van Cito zo weinig cultureel gebiaste opgaven bevatten dat de afzonderlijke toelatingstoetsen die uit deze banken worden samengesteld, nauwelijks of geen gebiaste opgaven bevatten. Daarmee zou het significante verschil in scores van de in dit project onderscheiden groepen aspirant-studenten, niet afdoende verklaard kunnen worden. In het volgende hoofdstuk gaan we in op de vraag of we andere toetsgerelateerde verklaringen kunnen vinden voor de achterblijvende toetsresultaten van kandidaten met een niet-westerse achtergrond.

Het feit dat we nauwelijks of geen bronnen voor culturele bias hebben aangetroffen, betekent niet dat we helemaal geen verbeterpunten hebben gesignaleerd. De inhoudelijke analyse van de 81 opgaven (waarvan in totaal 65 met meer of minder evidentie voor aanwezige DIF, en 16 zonder signalen die wijzen op DIF) die in dit onderzoek konden worden bekeken, heeft geleid tot een aantal bevindingen die waarschijnlijk kunnen bijdragen aan verbeteringen die voor alle kandidaten, ongeacht het geboorteland van hun ouders, gunstig kunnen uitpakken. We hebben de bevindingen met betrekking tot mogelijke verbeterpunten verwerkt in een aanvullende checklist voor de toetsconstructeurs die is opgenomen in bijlage 11 bij dit rapport. Bij het beschouwen van die checklist is het van belang om onder ogen te houden dat de aandachtspunten voortkomen uit een diepgaande analyse van 65 opgaven die op één of andere manier zijn opgevallen tijdens de DIF-analyses. Voor het grootste deel van de opgaven die in september en oktober aan aspirant-studenten zijn voorgelegd, geldt dat ze niet zijn geanalyseerd omdat ze in geen enkele fase van dit project zijn opgevallen als opgaven die mogelijk afwijkend gedrag vertonen.



# Mogelijke biasbronnen in de toetsdoelen en het voorbereidende materiaal



## 7.1 Inleiding

In het vorige hoofdstuk werd duidelijk dat de achterblijvende scores van de niet-westerse kandidaten die hebben deelgenomen aan de toetsvensters van september en oktober 2020, niet bevredigend verklaard kunnen worden door de aanwezigheid van een westerse bias in de toetsopgaven. De statistische analyses hebben opgeleverd dat slechts een zeer klein deel van de itembanken waaruit de toelatingstoetsen worden samengesteld DIF bevat, en dat we die gevallen van DIF slechts in zeer weinig gevallen in verband kunnen brengen met een mogelijke westerse biasbron. In dit hoofdstuk verschuiven we de aandacht van de afzonderlijke toetsopgaven naar de toetsen als geheel. Twee onderzoeksvragen staan centraal:

1. *Bevat de inhoud van de pabo-toelatingstoetsen Aardrijkskunde, Geschiedenis en Natuur & Techniek zoals die beschreven wordt in de Handreikingen, kenmerken die aspirant-pabostudenten met een niet-westerse achtergrond, ten onrechte kunnen benadelen?*
2. *Heeft de manier waarop studenten zich moeten voorbereiden op de toelatingstoetsen en hebben de hulpmiddelen die ze daarvoor krijgen aangereikt, kenmerken die aspirant-pabostudenten met een niet-westerse achtergrond, onbedoeld en ten onrechte kunnen benadelen?*

We starten in de volgende twee paragrafen met een beschouwing over kennis (7.2) en de manier waarop mensen kennis verwerven (7.3). In paragraaf 7.4 gaan we dieper in op de rol die taal daarbij speelt. In paragraaf 7.5 brengen we de inhoud van het eerste deel van dit hoofdstuk in verband met resultaten van onderzoek naar biasbronnen en presenteren we een 'kijkkader' waarmee we in de volgende paragrafen de in de Handreikingen beschreven doelen van de drie afzonderlijke toetsen – Aardrijkskunde, Geschiedenis en Natuur & Techniek – én de manier waarop studenten zich op de toets voorbereiden, onder de loep nemen. In paragraaf 7.9 bespreken we kort de opbrengsten van de gesprekken die we met studenten en andere experts hebben gevoerd. We sluiten het hoofdstuk af met een korte samenvatting en conclusies.

## 7.2 Powerful knowledge en referentiekaders

De leerdoelen die horen bij de toelatingstoetsen, zijn gebaseerd op een advies van de SLO aan de minister over de vakinhoudelijke kennis die bij pabo-instroom getoetst moet worden (Diephuis e.a. 2013). Bij de analyse van de inhoud van de toetsen, hierna 'de leerdoelen', gaan we in dit onderzoek uit van wat Diephuis e.a. (2013, p. 9) daar in hun advies over vastleggen: *"De beoogde vakinhouden en leerdoelen vormen de noodzakelijke voorkennis waarmee een toekomstig pabostudent de opleiding succesvol kan afronden zonder dat in de eerste leerjaren tijd aan het aanleren van elementaire voorkennis besteed hoeft te worden. (...) Het gaat om kennis en vaardigheden waarvan verwacht kan worden dat de student deze na het slagen voor een toelatingstoets met succes kan inzetten bij aanvang van de pabo-opleiding. Het gaat om reële en haalbare inhouden en beheersingsniveaus"*.

De vorm van kennis waar Diephuis e.a. het hier over hebben, wordt wel aangeduid als 'powerful knowledge' (Young en Muller, 2013). Powerful knowledge is de beste (vak)kennis die het onderwijs leerlingen en studenten te bieden heeft en die door de meeste mensen niet enkel kan worden opgedaan door eigen ervaringen. Het is kennis die mensen in staat stelt om de wereld om hen heen te begrijpen en om zelf nieuwe kennis te verwerven. Het is 'krachtige kennis' omdat die ons in staat stelt om beargumenteerde stellingen in te nemen in maatschappelijke discussies en om van daaruit te bedenken wat wij kunnen doen om aan een betere wereld bij te dragen.

Theorieën omtrent powerful knowledge geven aan dat het om 'systematische kennis' gaat. Het is kennis die is opgeslagen in de vorm van een werkbaar referentiekader (*framework of knowledge*), waarbinnen de losse 'brokjes' informatie onder andere via vakconcepten in een coherent geheel zijn samengebracht. Krachtige kennis moet dan ook onderscheiden worden van 'informatie', of 'losse feitenkennis'. Losse feiten worden niet alleen moeilijker onthouden; ze helpen mensen ook niet om daadwerkelijk op een doordachte manier over hun leefwereld na te denken en zich verder te ontwikkelen.

## 7.3 Het verwerven van powerful knowledge

Uit onderzoek komt naar voren dat het verwerven van powerful knowledge en het (opnieuw) omzetten van losse feitenkennis naar krachtige kennis door twee dingen wordt vergemakkelijkt.

Allereerst is dat het aanbrengen van een structuur (Beneker & Van Gaans, 2018). Het systematische karakter van powerful knowledge, zeker voor de zogenaamde zaakvakken, krijgt het beste vorm in een narratief (Howson & Shemilt 2011; Shemilt 2000). Zo'n verhaalstructuur maakt immers duidelijk hoe de verschillende feiten samenhangen en maakt daarmee het geheel aan informatie beter leerbaar. Daar komt bij dat structuur helpt om een onderscheid te maken tussen hoofd- en bijzaken. Wie de rode draad weet, is eerder in staat te bepalen welke feiten tot de kern, oftewel de rode draad, behoren en welke feiten eerder voorbeelden zijn van die rode draad. In een verhaal kunnen verschillende feitelijke uitspraken (eenheden van informatie) aan de hand van oorzaak-gevolg verbanden aan elkaar worden gekoppeld, waardoor er een groter geheel ontstaat. Waar die koppeling ontbreekt, wordt kennis 'los zand'. Juist die samenhang in de om reeds verworven kennis helpt mensen om nieuwe kennis sneller te verwerken en die kennis te gebruiken om nieuwe uitdagingen in de eigen leefwereld beter te begrijpen.

Daarnaast blijkt ook een uitgebreid referentiekader van 'alledaagse kennis' een positief effect te hebben op het verwerven van powerful knowledge. Het gaat daarbij zowel om heel algemene kennis, zoals kennis van de loop van de belangrijkste rivieren of van de bewegingen van de zon over de hele dag, als om een (vak)specifieke begrippenkennis, bijvoorbeeld weten wat turf is of wat nozems zijn. Er is hierbij sprake van wat wel wordt beschreven als een 'Matteuseffect' (Van Steensel & Segers 2017): hoe meer je al weet, hoe gemakkelijker je nieuwe dingen leert. Powerful knowledge betekent weliswaar veel méér dan *common sense* of alledaagse kennis die alleen op de eigen ervaringen en niet op wetenschappelijke inzichten is gebaseerd, maar een rijke leeromgeving en verband tussen nieuwe kennis en die eigen, dagelijkse ervaringen helpen wel degelijk om de opgedane kennis daadwerkelijk te begrijpen én om die te onderhouden. Wanneer opgedane kennis niet wordt gebruikt, dan zal het systematische karakter ervan geleidelijk aan weer verdwijnen, waardoor men uiteindelijk een losse en vaak incomplete verzameling van herinnerde feiten overhoudt.

## 7.4 Taalvaardigheid en leren

Mensen hebben taal nodig om kennis te verwerven en te delen. Met het verwerven van nieuwe kennis en inzichten, breiden mensen ook hun taalvaardigheid uit. Taalverwerving is dan ook bij uitstek een proces van 'leven lang leren': wanneer mensen eenmaal het systeem van een taal hebben verworven - de klanken, een basiswoordenschat, de morfologie, de grammatica -, dan zullen ze hun kennis, woordenschat en vaardigheden in het gebruik van dat taalsysteem nog hun leven lang blijven uitbreiden. Nieuwe ervaringen leiden tot nieuwe kennis en daarmee ook tot de taal die nodig is om over die nieuwe kennis na te denken, te praten, te schrijven en te lezen.

De relatie tussen het verwerven van nieuwe kennis en het leren van taal, is duidelijk zichtbaar in beschrijvingen van taalvaardigheidsniveaus. In Nederland zijn er twee verschillende 'referentiekaders voor taalvaardigheid' in gebruik: het Referentiekader voor Moderne Vreemde Talen (Council of Europe 2001) en het Referentiekader Taal (Expertgroep 2008, 2009). Beide 'meetlatten voor taalvaardigheid', die in het Nederlandse onderwijs veelal door elkaar gebruikt worden, beschrijven een aantal taalvaardigheidsniveaus die zich van elkaar onderscheiden door de mate van beheersing van een aantal 'vormelijke' kenmerken van het betrokken taalsysteem én door de aard en omvang van de inhoud waarover en de contexten waarin mensen via die taal kunnen communiceren. Naarmate hun taalvaardigheidsniveau stijgt, kunnen taalgebruikers over steeds meer thema's effectiever en efficiënter communiceren.

In beide referentiekaders wordt een niveau onderscheiden, in het Europees Referentiekader aangeduid als 'B2', in het Referentiekader Taal als '3F', waarop een taalgebruiker verondersteld wordt de beschikking te hebben over redelijke basiswoordenschat en een goede beheersing van het taalsysteem. In het Raamwerk Nederlands, dat in opdracht van het Ministerie van OCW door een projectgroep van CINOP in samenwerking met andere expertisebureaus werd ontwikkeld, wordt leesvaardigheid op dat niveau wordt beschreven als *"Kan een grote variatie aan teksten zelfstandig lezen binnen eigen opleiding/werk of in het dagelijks leven. Begrijpt teksten over specialistische onderwerpen als hij voldoende kennis heeft van het desbetreffende onderwerp"* (Bohnen e.a., 2007, p. 58).

Niveau 3F/B2 is het streefniveau van havo en mbo-4. Van de studenten die zich voorbereiden op de toelatingstoetsen, zowel van moedertaalsprekers van het Nederlands als van studenten voor wie het Nederlands een tweede taal is, mag verwacht worden dan ze beschikken over leesvaardigheid in het Nederlands op dat niveau.

Een verkenning van de Handreikingen en de andere materialen die de aspirant-studenten ter beschikking hebben om zich op de drie toetsen voor te bereiden - de handboeken, de online-inleidingen - maken duidelijk dat het taalgebruik in die teksten géén formele kenmerken heeft die niet passen bij de beschrijvingen van niveau 3F/B2. Het taalgebruik is niet overdreven formeel, de zinnen zijn niet héél erg ingewikkeld, we vinden weinig voorbeelden van figuurlijk taalgebruik of idioom.

Toch zien we hier een 'talige' biasbron voor sommige studenten, die nauw samenhangt met de inhoud van de Handreikingen. De aangereikte teksten over de inhoud van de toetsen behandelen op een zeer beknopte manier héél veel nieuwe informatie. Ze bevatten nauwelijks voorbeelden of andere toevoegingen, zoals bijvoorbeeld illustraties, herhalingen, herformuleringen, verwijzingen naar bekende contexten, die de teksten toegankelijker maken. Het zelfstandig lezen en begrijpen van dergelijke teksten is géén vaardigheid die zonder meer verwacht kan worden van studenten die het Nederlands beheersen op niveau 3F/B2. Hier is eerder 4F/C1 een vereiste. Dat niveau onderscheidt zich niet of nauwelijks door vormkenmerken van 3F/B2, maar vooral door de beschrijving van de inhoud en de contexten waarover en waarin taalgebruikers zelfstandig kunnen communiceren. Beheersers van niveau 4F/C1 worden geacht zelfstandig te kunnen communiceren over méér en abstractere inhoud, dan beheersers van niveau 3F/B2. Op niveau 4F/C1 worden taalgebruikers dan ook geacht zich zelfstandig, via taal en zonder ondersteuning, nieuwe kennis eigen te kunnen maken.

## 7.5 Een kijkkader voor de inhoud van de toelatingstoetsen

Uitgaande van de hiervoor behandelde overwegingen met betrekking tot het 'wezen' van de kennis en vaardigheden die in de toetsdoelen van toelatingstoetsen zijn beschreven en de manier waarop dat soort kennis en vaardigheden verworven worden, hebben we de 'perspectieven' onderscheiden van waaruit we de in de Handreikingen beschreven inhoud, én de manier waarop die verworven moeten worden, zullen analyseren. Doel van de analyse is het opsporen van kenmerken van die doelen en van de manier waarop kandidaten zich daarop voorbereiden, die kunnen helpen verklaren dat studenten met een niet-westerse achtergrond gemiddeld minder succesvol zijn bij het afleggen van de toelatingstoetsen dan studenten met twee in Nederland geboren ouders. We sluiten hier weer aan bij de uitkomsten van eerder verricht onderzoek naar culturele bias in toetsen (Kerkhoff & Vallen, 1984; Uiterwijk, 1994; Van Schilt, 2004) én aanwijzingen voor toetsconstructeurs die uit dergelijk onderzoek voortkomen (Bügel & Sanders, 1998).

We onderscheiden de volgende twee onderling samenhangende 'zichtlijnen'.

#### a. Het narratief/de structuur

In hoeverre is er een 'narratief' of andere vorm van structurering herkenbaar in de beschrijving van de leerdoelen en zo ja, is dat narratief voor studenten met een niet-westerse achtergrond even herkenbaar als voor studenten met twee in Nederland geboren ouders? We gaan er hier vanuit dat een herkenbaar narratief de studenten zal helpen om de leerstof te structureren en te memoriseren. Structuur zal tijdens het afleggen van de toets óók helpen om de gevraagde kennis op het goede moment uit het geheugen op te diepen.

#### b. Voorkennis

In hoeverre veronderstelt de beschrijving van de leerdoelen voorkennis – dat wil zeggen kennis die niet expliciet in de leerdoelen is vastgelegd -, en om welke voorkennis gaat het dan? Kan die voorkennis in gelijke mate verondersteld worden aanwezig te zijn bij de op basis van hun etnische achtergrond onderscheiden groepen studenten? Het uitgangspunt is hier dat reeds aanwezige kennis tijdens het bestuderen van de stof zal helpen om de nieuwe kennis betekenis te geven én om tijdens het examen de vragen én de antwoordmogelijkheden goed te interpreteren.

Langs deze twee zichtlijnen bespreken we hierna de inhoud van de toelatingseisen voor de drie verschillende vakken. We staan daarbij niet meer apart stil bij de rol die 'taalvaardigheid' als biasbron kan spelen. Zoals in de voorafgaande paragraaf is beschreven, gaan we ervan uit dat de aspirant-studenten Nederlands beheersen op niveau 3F/B2. Voor elk van de drie Handreikingen bij de toelatingstoetsen geldt dat het zelfstandig bestuderen ervan, zonder ondersteuning, meer voorkennis en daarmee een hoger leesvaardigheidsniveau vereist dan op niveau 3F verondersteld mag worden.

## 7.6 Bronnen voor toetsbias in de toelatingstoets Geschiedenis

### Geschiedenis: het narratief

De *Handreiking Geschiedenis* van het SLO (SLO, 2014) benoemt twee belangrijke domeinen: vakgebonden vaardigheden en kennis van de tien tijdvakken. De benoemde vakgebonden vaardigheden zijn achtereenvolgend:

- omgaan met historische bronnen;
- hanteren van de tijdbalk;
- werken met historische periodes;
- het leggen van relaties tussen oorzaken en gevolgen en
- leggen van relaties tussen continuïteit en verandering.

De student moet deze vaardigheden op een basaal niveau beheersen. Met betrekking tot de tijdvakken moeten aspirant-studenten zevenendertig grotere fenomenen (kenmerkende aspecten: ontwikkelingen en verschijnselen) beheersen op een begripsniveau. De Handreiking biedt geen narratief van het verleden, maar wordt gepresenteerd als een verzameling losse feiten, verspreid over verschillende tijdvakken en kenmerkende aspecten, dat wil zeggen als thema's die uniek zijn voor dat tijdvak. De unieke thema's worden niet in door groter verhaal, een vertelling over de tijdvakken heen, met elkaar verbonden.

Vanuit het perspectief van een historicus staat de beschreven stof niet helemaal los van een narratief. Hij of zij zal de voor de aspirant-student impliciet blijvende structuur van een achterliggend hoofdzakelijk westers narratief zeker herkennen. Een aspirant-student die zich voorbereidt op de toelatingstoetsen, zal de samenhang in de te bestuderen stof, echter zelf moeten zoeken. Wie dit narratief als referentiekader van jongs af aan meegekregen heeft, heeft waarschijnlijk wat mentale kapstukken om de aangeboden kennis te integreren. Voor wie een ander narratief aangereikt heeft gekregen, is de kennis ongetwijfeld moeilijker te memoriseren.



## Geschiedenis: voorkennis

Het impliciet aanwezige narratief heeft ertoe geleid dat de leerstof die in de Handreiking wordt beschreven, vooral betrekking heeft op fenomenen die betekenisvol zijn binnen een westers narratief. Wat opvalt is dat er in de beschreven leerstof ook wat negatieve elementen – zwarte pagina's, zoals ons optreden in Indonesië – worden genoemd. Er zijn ook wat niet-westerse elementen zichtbaar, maar die krijgen aanzienlijk minder aandacht dan het westerse 'verhaal'. Het is en blijft in de allereerste plaats de geschiedenis van een op christelijke cultuur gebaseerd, geseculariseerd en democratisch Nederland, dat zich vooral verbonden weet met (West-)Europa. Dat dit Europa en daarmee Nederland zich nooit ontwikkelde onafhankelijk van de rest van de wereld, wordt niet of nauwelijks vermeld, net zomin als de bijdragen van andere culturen en staten op onze geschiedenis. Er wordt ook heel weinig gebruik gemaakt van kansen om bij de beschrijving van de 'Nederlandse geschiedenis' in elk geval te verwijzen naar niet-westerse ontwikkelingen. We geven hier een paar voorbeelden die de 'westerse' inhoud van de kennisbasis illustreren.

- Bij de verspreiding van het christendom worden meerdere namen genoemd, bij dat van de Islam alleen Mohammed (en alleen als tijdsaanduiding: 'Na de dood van Mohammed...').
- De kruistochten worden enkel kort besproken onder het fenomeen 'de opkomst van handel en het ontstaan van steden'. Daarmee worden ze narratief gekoppeld aan de economische ontwikkeling van het westen, een koppeling die in niet-westerse narratieven waarschijnlijk ontbreekt. De komst van nieuwe groepen Nederlanders naar Nederland is onderdeel van het laatstgenoemde fenomeen en neemt slechts enkele zinnen in beslag.
- De studenten moeten het ontstaan van de landbouw kunnen uitleggen en kunnen plaatsen in het Midden-Oosten. De grote rijken die hieruit ontstonden en die een hoge mate van ontwikkeling kenden, vormen echter géén onderdeel van de leerstof.
- Bij het ontstaan van het christendom stelt de tekst dat de boodschap van Jezus is opgenomen in de Bijbel, hetgeen historisch al bevestigd kan worden, en binnen bijvoorbeeld een islamitische narratief minder wordt geaccepteerd.

## Geschiedenis: samenvatting en conclusies

Onze beschouwing van de inhoud van de toelatingstoets geschiedenis leidt tot de conclusie dat die kenmerken heeft die kunnen helpen verklaren dat niet-westerse kandidaten minder succesvol zijn. Het gebrek aan een expliciet narratief zal het verwerken en memoriseren waarschijnlijk voor alle studenten moeilijker maken dan nodig is. We mogen echter verwachten dat het narratief dat impliciet wél aanwezig is, voor kinderen van twee in Nederland geboren ouders vaker herkenbaar zal zijn dan voor kinderen van ouders die in een niet-westers land zijn geboren. De student met in Nederland geboren ouders die van jongs af aan iets van dit westerse narratief heeft meegekregen, heeft waarschijnlijk méér mentale 'kapstokken' en méér voor de toelatingstoetsen relevante voorkennis dan een student met niet-westerse wortels.

# 7.7 Bronnen voor toetsbias in de toelatingstoets Aardrijkskunde

## Aardrijkskunde: het narratief

De toetsdoelen lopen bij Aardrijkskunde uiteen van fysisch-geografische vraagstukken zoals de verspreiding van vulkanen en landschappen van Nederland tot en met sociaalgeografische vraagstukken zoals de verspreiding van talen in Afrika. De structuur, of het 'narratief', zal net zoals bij geschiedenis het geval was, voor veel studenten moeilijk zijn om te herkennen. Hoewel zinnen als '*de mens leeft met en van de natuur*' mogelijk richting geven voor de te ontdekken structuur in de leerdoelen, zal de beschreven toetsinhoud voor veel studenten overkomen als een verzameling van losse feiten en feitjes.

De verscheidenheid aan onderwerpen wordt duidelijk in de acht gekozen kernthema's. Die omhelzen thema's zoals '*aarde, klimaat en landschappen*', '*grenzen en identiteit*' en '*water*'. Geografische werkwijzen kunnen handvatten geven om de te bestuderen stof in een context te plaatsen.

Door consequent jezelf vragen te stellen over de te leren stof, kan er samenhang ontstaan tussen de verschillende feiten. Zo kan de verspreiding van aardbevingen en vulkanen duidelijk worden door een relatie te leggen met de te bestuderen stof over platentektoniek en de verschillende plaatbewegingen. De 'powerful knowledge' die nodig is voor een dergelijke manier van omgaan met de beschreven leerstof, hoort inderdaad ook bij de toelatingseisen voor de pabo. Het vergt echter oefening, training en onderwijs om je deze verschillende geografische werkwijzen aan te leren. Als beheersing van die vakvaardigheden óók een voorwaarde is om je met succes op de toetsen voor te bereiden, dan is er in onze ogen sprake van bias. Aspirant-studenten die de vaardigheden niet beheersen op het moment dat ze starten met de voorbereiding op de toetsen, zullen benadeeld zijn ten opzichte van de studenten die al kennis hebben gemaakt met deze vaardigheden en/of die toegang hebben tot ondersteuning.

### **Aardrijkskunde: voorkennis**

Een groot deel van de beschreven toetsinhouden is ontleend aan de reguliere programma's voor po en vo. Binnen de beschreven thema's is daarbij op een aantal plaatsen dan ook een sterke focus op Nederland. Zo gaat er binnen het onderdeel "landschappen" veel aandacht uit naar de zes dominante landschappen in Nederland. De bevolkingsspreiding van Nederland wordt bekend verondersteld en bij natuurrampen worden overstromingen als dreiging genoemd voor Nederland. Bij het thema 'water' is bovendien veel functionele topografie verwerkt door te verwijzen naar de belangrijke waterlichamen in en rond Nederland.

Deze keuze voor een oriëntatie op Nederland ligt voor de hand omdat een belangrijk deel van de wereldoriëntatie die de toekomstige leerkrachten zullen gaan onderwijzen, ook over Nederland zal gaan. Studenten die meer van Nederland hebben gezien, zullen hierdoor bij de voorbereiding op de toetsen in het voordeel zijn.

### **Aardrijkskunde: samenvatting en conclusies**

De inhoud van de toelatingseisen Aardrijkskunde is in de eerste plaats gericht op Nederland. Aspirant-studenten die méér van Nederland hebben gezien, zullen daardoor meer voorkennis hebben die hen helpt de vereiste kennis en vakvaardigheden te verwerven.

De meeste studenten zullen in de Handreiking bij de toelatingstoets Aardrijkskunde weinig houvast vinden om de te beheersen stof te structureren. De vaardigheden die zij hiervoor nodig hebben, horen bij de powerful knowledge die onderdeel is van de toelatingseisen. Aspirant-studenten die vóórdat ze zich gaan voorbereiden op de toelatingstoetsen al kennis hebben gemaakt met manieren om aardrijkskundige kennis te structureren, zullen in het voordeel zijn ten opzichte van kandidaten voor wie dit perspectief op de wereld om hen heen, nieuw is. Wie via het mbo naar de pabo gaat en maximaal drie jaar onderwijs aardrijkskunde in de onderbouw van het vmbo heeft gehad, zal bij de voorbereiding op de toelatingstoetsen Aardrijkskunde weinig handvatten hebben om een structuur aan te verbinden.

## **7.8 Bronnen voor toetsbias in de toelatingstoets Natuur & Techniek**

In deze paragraaf worden kenmerken van de Handreiking Natuur & Techniek (SLO, 2013c) beschreven, die mogelijk aanleiding geven voor verschillen in scores tussen kandidaten met een Nederlandse achtergrond en een niet-westerse migratieachtergrond. De Handreiking bevat het examenprogramma Natuur & Techniek, de voorbeeldopgaven en de toetsmatrijs.

Het examenprogramma Natuur & Techniek bestaat uit vaardigheden en kernconcepten. Per vaardigheid en kernconcept is een aantal leerdoelen geformuleerd. Deze leerdoelen zijn nader gespecificeerd in cursieve tekst. Meestal is deze specificering niet meer dan een opsomming van begrippen, al dan niet op drie niveaus. Hieronder worden kenmerken van examenprogramma, voorbeeldopgaven en toetsmatrijs nader beschreven met behulp van twee criteria: narratief en voorkennis.

### **Natuur & Techniek: het narratief**

Anders dan bij Geschiedenis lijkt er bij Natuur & Techniek niet direct sprake van verschillende narratieven. Natuur & Techniek wordt beheerst door één specifiek narratief, het natuurwetenschappelijke denken. Die onderzoekende, gedistantieerde, empirische en analytische manier van denken is herkenbaar in het examenprogramma en de voorbeeldopgaven, maar wordt nergens geëxpliciteerd.

Voor aspirant-studenten uit een omgeving waarin meer traditioneel gedacht wordt, kan die natuurwetenschappelijke manier van denken minder herkenbaar zijn. Enerzijds zouden specifieke onderwerpen, zoals voortplanting van de mens en de evolutietheorie, in de taboesfeer kunnen liggen, anderzijds kunnen geboden contexten eerder een reactie van emotionele betrokkenheid oproepen dan van analyse. Bijvoorbeeld, voorbeeldopgave 1 in de Handreiking vraagt naar de locatie van de evenwichtsorganen in het lichaam. Een meer analytisch georiënteerde aspirant-student zal die spontaan, zonder nadere informatie, eerder dicht bij de hersenen (dus oren) plaatsen, dan een meer emotioneel betrokken aspirant-student die mogelijk eerder aan de locatie van het ongemak (dus buik) zal denken. In voorbeeldopgave 9 moet de aspirant-student laten zien dat hij/zij weet dat je in zout water makkelijker drijft. Er wordt daarbij verwezen naar zwemles, wat allerlei emotionele associaties kan oproepen, maar hoe dan ook niets met de gestelde vraag te maken heeft. Zwemles vindt immers nooit plaats in een zoutwaterbad. Ook vragen met een medische context kunnen verwarring oproepen, als aspirant-studenten gewend zijn vanuit vertrouwen of wantrouwen een arts te benaderen, in plaats van vanuit analyse.

### **Natuur & Techniek: voorkennis**

In het examenprogramma zelf wordt weinig contextuele steun gegeven. De leerdoelen bevatten geen context, terwijl de nadere specificering meestal niet meer dan een opsomming van begrippen zonder context is.

In de voorbeeldopgaven wordt wel contextuele steun gegeven. Voor de hoogopgeleide onderzoeker komen de contexten uit het dagelijks leven. Voor de aspirant-studenten is het echter de vraag of die contexten herkenbaar zijn. Er worden een aantal medische contexten aangeboden, zoals een kunstnier, de concentratie insuline in het bloed en de erfelijkheid van de ziekte hemochromatose. Er worden landbouwkundige en industriële contexten aangeboden, zoals de aardappelplant en destillatie van alcoholhoudende dranken. De enige context die past bij de basisschool is die in voorbeeldopgave 9, de zwemles. Daarbij wordt echter een vraag gesteld die in werkelijkheid in die context nooit gesteld zal worden. Aspirant-studenten moeten laten zien dat ze weten dat je beter in zout water blijft drijven dan in diep of warm water, terwijl tijdens een (school)zwemles niet de dichtheid van het water varieert, maar veel meer de didactische talenten van de instructeur en de bouw, de motoriek en het zelfvertrouwen van de leerling.

Al met al is de contextuele steun in het examenprogramma voor aspirant-studenten zeer gering. Dat dit met name aspirant-studenten met een niet-westerse migratieachtergrond benadeelt, is daarmee niet gezegd. Wel kan worden gesteld dat de contextuele steun die geboden wordt, alleen herkenbaar is voor aspirant-studenten met een groot referentiekader. De contexten die in de voorbeeldopgaven wordt gegeven, zijn bovendien vooral modern-westers van aard: hoogstaande medische technologie, aardappels en alcoholische drank. Een moslim die nooit aardappels eet, geen alcohol drinkt en in medische zaken op Allah vertrouwt, zal bijzonder weinig van die contextuele steun herkennen. Hetzelfde geldt voor een kind van laagopgeleide ouders die weinig technische kennis hebben en hun kinderen niet meenemen naar musea als NEMO, Corpus en Naturalis.

## Natuur & Techniek: samenvatting en conclusies

Al met al kan het narratief van natuurwetenschappelijk denken minder toegankelijk zijn voor bepaalde groepen aspirant-studenten. Dit kan vooral problematisch zijn als geboden contexten andere associaties oproepen dan die die nodig zijn voor het goed beantwoorden van de vraag. Daarbij is de aard en omvang van het referentiekader van de kandidaat cruciaal voor diens mogelijkheid om de benodigde inhoud op te kunnen nemen, te kunnen structureren en te kunnen gebruiken.

## 7.9 Bias in de voorbereiding op de toetsen

De website [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl) biedt de aspirant-studenten een schat aan informatie en hulpmiddelen die kunnen ondersteunen bij de voorbereiding op de toelatingstoetsen: Handreikingen met beschrijvingen van de leerdoelen, kaarten, voorbeeldopgaven en een toetsmatrijs, per toets vijf voorbereidingsvideo's met informatie over het doel en de inhoud van de toetsen, uitgebreide besprekingen van voorbeeldopgaven, adviezen voor het maken van een studieplan en geschikt studiemateriaal. Voor studenten die daar behoefte aan hebben, worden er daarnaast nog colleges aangeboden, bieden sommige pabo's nog korte voorbereidende cursussen en kent het mbo een keuzedeel 'voorbereiding pabo'. Toch denken we dat de manier waarop studenten zich op de toetsen moeten voorbereiden, een aantal mogelijke bronnen voor bias bevat. We presenteren hierna onze voornaamste veronderstellingen op dit punt.

### a. Mbo-studenten moeten méér toelatingstoetsen doen dan de meeste havisten.

De meeste aspirant-studenten melden zich in het jaar voordat ze willen doorstromen bij de gewenste opleiding aan. Voor elke toelatingstoets moet een forse hoeveelheid kennis worden geleerd. De voorbereiding op de toelatingstoetsen betekent dan ook voor alle aspirant-studenten een serieuze extra belasting in het eindexamenjaar. Mbo-studenten zijn daarbij in het nadeel ten opzichte van havisten doordat zij alle drie de toelatingstoetsen moeten doen. Havisten hoeven doorgaans slechts één en hoogstens twee toelatingstoetsen te doen doordat ze een of twee van de vereiste vakken al in hun eindexamenpakket hadden. Het is wat dat betreft opmerkelijk dat aspirant-studenten met een mbo 4-diploma uit een andere sector dan Zorg en Welzijn, bij de toelatingstoetsen gemiddeld even succesvol zijn als havisten.

### b. Aspirant-studenten moeten zich vooral door zelfstudie op de toetsen voorbereiden.

We schatten in dat een aantal studenten extra benadeeld zal zijn doordat ze geacht worden de vereiste kennis vooral via zelfstudie te verwerven. We denken daarbij aan studenten met de volgende, onderling samenhangende, kenmerken.

- Aspirant-studenten die de narratieven die in de beschreven kennis impliciet aanwezig zijn niet herkennen, zullen méér benadeeld worden doordat ze vooral aangewezen op zelfstudie, dan aspirant-studenten voor wie de te bestuderen stof meer aanknopingspunten biedt.
- Voor alle studenten die lezen op niveau 3F/B2 geldt dat zij bij zelfstudie in het nadeel zullen zijn ten opzichte van aspirant-studenten die een hoger leesvaardigheidsniveau beheersen.
- Beheersers van 3F/B2 die uitgaan van een ander referentiekader dan de teksten veronderstellen, zullen extra afhankelijk zijn van steun bij het verwerken van de beschreven leerstof. Zij zullen extra benadeeld zijn wanneer ze weinig of geen ondersteuning krijgen bij het verwerven van de leerstof.

We nemen aan dat het mbo-keuzedeel 'voorbereiding Pabo' en de korte ondersteunende cursussen die sommige pabo's aanbieden voor veel aspirant-studenten nuttig zullen zijn, maar we schatten in dat ze niet zullen volstaan om de nadelen van de focus op zelfstudie voor alle aspirant-studenten weg te nemen.

## 7.10 Opvattingen van aspirant-studenten en experts

Onze gesprekken met eerstejaars pabostudenten, aspirant-studenten in het vierde jaar van het mbo, vmbo-docenten en docenten van cursussen die voorbereiden op de toelatingstoetsen, leveren gegevens op die aansluiten bij de veronderstellingen die in dit hoofdstuk worden geformuleerd.

## Structuur en voorkennis

De eerstejaars pabostudenten die de toetsen al gehaald hebben, bevestigen zonder één uitzondering dat ze moeite hebben gehad met het herkennen van structuur en samenhang in de door hen bestudeerde stof. Ze geven aan dat ze de inhoud van de toelatingstoetsen hebben ervaren als een verzameling feiten die als 'los zand' aan elkaar verbonden is.

Natuurlijk hebben we overwogen dat alle leerlingen in de onderbouw van het voortgezet onderwijs een begin van een gemeenschappelijk referentiekader aangereikt hebben gekregen: de inhoud van de toelatingstoets betreft de kerndoelen van het leergebied mens-en-maatschappij. De toetsresultaten en de gesprekken met studenten en docenten die we naar aanleiding van onze bevindingen hebben gevoerd, leiden echter tot vragen over de effectiviteit van het onderwijs in de betreffende vakken gedurende twee (vmbo) of drie (havo en vwo) jaar. De meeste (aspirant-)studenten die we hebben gesproken, vertellen dat ze al enkele jaren geen of nauwelijks onderwijs hebben gehad in de vakken waarin ze toelatingsexamen moeten doen. Ze bevestigen dat er maar weinig is blijven 'hangen' van het onderwijs dat ze in deze vakken gehad hebben. Ze geven aan dat ze heel weinig of niets herkenden in de te bestuderen stof en dat ze het overgrote deel van de geboden informatie als nieuw hebben ervaren. Waarschijnlijk moet dat 'vergeten' verklaard worden doordat de leerlingen de opgedane kennis nadat ze de betrokken schoolvakken hadden laten vallen, niet meer hebben gebruikt. Voor het daadwerkelijk oriënteren op de geschiedenis, de wereld om hen heen en techniek, krijgen jonge mensen óók buiten de school verhalen aangereikt, veelal door mensen die op hen een grotere invloed uitoefenen dan een vakdocent. Bovendien zullen de verschillen tussen de persoonlijke leefwerelden van aspirant-studenten al in de onderbouw van het voortgezet onderwijs van invloed zijn geweest op wat zij daar hebben opgestoken.

De studenten die bij de voorbereiding op de toelatingstoetsen onderwijs hebben gevolgd, bevestigen dat de leerstof daardoor voor hen interessanter werd en gemakkelijker te verwerken. Deze studenten benoemen vooral het aanbrenge van een duidelijke structuur in de als een verzameling los zand waargenomen leerstof, als effectief. Ook de op [www.Goedvoorbereidnaardepabo.nl](http://www.Goedvoorbereidnaardepabo.nl) geadviseerde commerciële methoden, zoals Basiskennis Geschiedenis van Noordhoff Uitgevers, helpen volgens sommige informanten om structuur te ontdekken en kapstukken te bouwen. Opmerkelijk is in dat verband dat één van de eerstejaars studenten (haviste, niet-westerse achtergrond) expliciet aangaf geen enkele methode gebruikt te hebben bij de voorbereiding op de geschiedenisstoets. Dit was het enige toetsonderdeel dat ze had moeten afleggen, maar ze had de geadviseerde methoden niet gebruikt omdat ze die te duur had gevonden om aan te schaffen. Ze had zich voorbereid op de toetsen door de Handreiking en vooral alle voorbeeldexamens 'uit haar kop te knallen'. Ze vertelde de meeste vragen die ze tijdens de toets kreeg voorgelegd, te hebben herkend. Ze is geslaagd.

## Zelfstudie

Gevraagd naar verklaringen voor hun eigen succes bij het afleggen van de toetsen, noemen de eerstejaars pabostudenten allemaal variaties op het antwoord van één van hen: "Het is gewoon een kwestie van stampen". Die nadruk op 'stampen' als leerstrategie, verklaart waarschijnlijk ook wat de eerstejaars pabostudenten vertellen over wat er is blijven hangen van de bestudeerde stof. Als we de eerstejaars studenten aan het eind van het interview vragen welke uitslag ze denken te halen als we hen de toets ter plekke nog een keer zouden voorleggen, dan zeggen ze zonder één uitzondering: "Dan zou ik waarschijnlijk zakken". Een studente voegt als een soort van verontschuldiging toe dat haar stagebegeleidster had aangegeven dat zij vast en zeker ook geen voldoende zou halen als ze de toetsen alsnog zou moeten doen.

Ook de drie mbo-studentes die we spraken, vertelden enthousiast over de inhoud van de lessen die ze gevolgd hadden in het kader van het keuzedeel 'voorbereiding op het hbo'. De stof was interessant en ze hadden met veel plezier gewerkt aan de opdracht om een klein onderzoekje te doen. Ze leken dan ook niet erg tegen de toelatingstoetsen op te zien en hadden het idee dat het keuzedeel een goede voorbereiding was geweest op de toelatingstoetsen. De interviewers kregen de indruk dat deze studenten op het moment van het interview, februari 2021, nog onderschatten wat de toelatingseisen zouden gaan eisen van hun parate kennis over geschiedenis, aardrijkskunde en natuur & techniek. Daarbij viel op dat geen van de drie studenten aangaf dat ze al een van de geadviseerde basismethodes had aangeschaft. Ze waren er ook nog niet zeker van dat ze die zouden gaan aanschaffen: de prijs speelde daarbij een belangrijke rol. Drie van die handboeken kosten samen zo'n 100 euro en dat werd als een hoge prijs ervaren.

Uit de interviews met docenten en kandidaten blijkt ook dat kandidaten inderdaad moeite hebben met het gedistantieerde, analytische natuurwetenschappelijke manier van denken die vooral in de toets Natuur & Techniek, maar óók in de andere toelatingstoetsen wordt gevraagd. Indien kandidaten de context van een opgave herkennen en het antwoord paraat hebben, dan geven zij veelal zonder redenering het goede antwoord op de gestelde vraag. Indien ze het antwoord niet paraat hebben, en dat komt nogal eens voor, dan wordt een mengsel van associatief en logisch denken ingezet om foute antwoorden 'weg te strepen' om bij het beste antwoord uit te komen. Dan zien we veelvuldig hoe antwoordmogelijkheden één voor één worden gewogen en afbeeldingen intuïtief met elkaar worden vergeleken. Meestal wordt dan het antwoord gekozen dat gevoelsmatig het meest logisch óf dat het meest afwijkend is. Die manier van denken leidt vaak tot foute antwoorden.

## 7.11 Samenvatting en conclusies

De inhoudelijke analyse van de Handreikingen bij de toelatingstoetsen en de manier waarop aspirant-studenten zich vooral via zelfstudie op de toetsen moeten voorbereiden, heeft een aantal mogelijke biasbronnen aan het licht gebracht. Daarbij is niet altijd duidelijk dat de mogelijke effecten van die biasbronnen per definitie in het nadeel van aspirant-studenten met een niet-westerse achtergrond zullen uitpakken. We hebben de volgende bronnen voor toetsbias gesignaleerd.

- Mbo-studenten zijn in het nadeel ten opzichte van havisten doordat ze drie toelatingstoetsen moeten afleggen terwijl de havisten doorgaans maar in een of twee vakken een toelatingstoets hoeven te doen.
- Het ontbreken van een duidelijke structuur ('narratief') in de beschrijving van de te beheersen kennis in de Handreikingen zal het voor alle studenten moeilijker maken om de stof te interpreteren, te structureren en te memoriseren. Studenten die de structuur die impliciet aanwezig is niet herkennen omdat ze die, of delen daarvan, niet van huis uit hebben meegekregen, zullen bij de voorbereiding op de toelatingstoetsen in het nadeel zijn ten opzichte van studenten die wat dat betreft meer van huis uit hebben meegekregen.



## Samenvatting en conclusies

In dit hoofdstuk worden de voornaamste resultaten van het uitgevoerde onderzoek samengevat en geëvalueerd. We behandelen de vier onderzoeksvragen die in hoofdstuk 4 zijn geformuleerd, ieder afzonderlijk.

### 8.1 Partijdigheid in de toetsopgaven

In hoofdstuk 6 worden de resultaten van het onderzoek naar de aanwezigheid van itembias gepresenteerd. Dit deel van het onderzoek heeft geen gegevens opgeleverd die aantonen dat aspirant-pabostudenten met een niet-westerse migratieachtergrond, worden benadeeld door een westerse bias in afzonderlijke opgaven van de toelatingstoetsen voor de pabo.

Alle 522 toetsopgaven – 174 per vak - die tijdens de toetsvensters van september en oktober 2021 zijn gebruikt bij de toetsing van Aardrijkskunde, Geschiedenis en Natuur & Techniek, zijn met behulp van drie verschillende statistische procedures geanalyseerd op 'different item functioning'. De analyses leverden heel weinig opgaven op die zich anders lijken te gedragen dan de andere opgaven. In de eerste globale analyses kwamen 65 opgaven naar voren die zich mogelijk anders gedroegen. Daarvan bleven er na een diepgaandere analyse maar 20 over: 5 voor Geschiedenis, 8 voor Aardrijkskunde en 7 voor Natuur & Techniek. Bij elk toetsonderdeel gold dat ongeveer de helft van deze 'DIF-opgaven' door de aspirant-studenten met een niet-westerse achtergrond beter waren gemaakt dan door de aspirant-studenten met twee in Nederland geboren ouders.

Een inhoudelijke analyse van circa 25 opgaven per vakgebied, waaronder alle opgaven die in de eerste analyses DIF leken te vertonen en per toetsonderdeel 5 of 6 opgaven zonder DIF, leverde nauwelijks gegevens op die wijzen op de aanwezigheid van itembias die wordt veroorzaakt doordat bepaalde opgaven minder dan andere opgaven zouden aansluiten bij culturele kenmerken van de onderscheiden groepen studenten. Natuurlijk konden er bij sommige opgaven relaties worden gelegd met mogelijk relevante kenmerken van de culturele achtergrond van aspirant-studenten - in de sfeer van pijnbomen die misschien bekender zijn voor studenten met een niet-westerse migratieachtergrond of 'aquaplaning' als een verschijnsel waar studenten met twee in Nederland geboren ouders mogelijk beter bekend mee zijn omdat ze wellicht eerder beginnen aan rijlessen - maar die relaties moesten dan doorgaans ver worden gezocht en worden bovendien maar mager ondersteund door de verzamelde data. Ook de gesprekken met eerstejaars pabostudenten, mbo 4-studenten en vmbo- en mbo-docenten leverden weinig gegevens op die wijzen op aanwezigheid van culturele itembias.

We hebben óók nauwelijks aanwijzingen gevonden voor een mogelijke 'talige' bias in de opgaven. De circa 75 opgaven die geanalyseerd zijn, voldoen op een heel klein aantal uitzonderingen na, aan alle eisen die gesteld kunnen worden aan de formulering van opgaven voor aspirant-studenten die geacht mogen worden de stof te hebben bestudeerd en Nederlandse teksten te kunnen lezen op niveau 3F

van het Referentiekader Taal. Ook in de interviews met de studenten komt niet naar voren dat zij problemen ervaren met de leesbaarheid van de opgaven. Dat geldt zowel voor de studenten met Nederlands als moedertaal als voor de studenten voor wie het Nederlands een tweede taal is. Waar we tijdens de bespreking van opgaven zijn gestuit op 'taalproblemen', dan hadden die steeds betrekking op onbekendheid met bepaalde concepten: 'löss', 'turf', 'veen'. We hebben onbekendheid met deze woorden in dit onderzoek opgevat als een tekort in de gemeten vakkennis dan wel een aspect van het referentiekader van een kandidaat, en niet als een tekort in taalvaardigheid in het Nederlands. De voorbeelden van mogelijke taalproblemen die werden genoemd door een van de geïnterviewde docenten, hebben we niet als 'biasbron' geïnterpreteerd, omdat we ervan uitgaan dat beheersing van de genoemde voorbeelden verwacht mag worden van de doelgroep van de toelatingstoetsen, die geacht worden Nederlands te beheersen op niveau 3F van het Referentiekader taal. Het ging daarbij om beheersing van een woord als 'sterven' in een context waarin oorlogsslachtoffers aan de orde komen.

Het feit dat het onderzoek geen gegevens heeft opgeleverd die wijzen op de aanwezigheid van bronnen voor een betekenisvolle culturele of talige itembias in de toelatingstoetsen, betekent niet dat het onderzoek helemaal geen verbeterpunten voor de afzonderlijke opgaven heeft opgeleverd. Die punten zijn verwerkt in vier 'aandachtspunten' voor de toetsconstructeurs' die als bijlage 11 aan dit rapport zijn toegevoegd.

## 8.2 Partijdigheid in de inhoud van de toelatingstoetsen

Het onderzoek heeft geleid tot een aantal bevindingen die mogelijk kunnen bijdragen aan een verklaring voor het feit dat aspirant-studenten waarvan minstens één ouder in een niet-westers deel van de wereld is geboren, lagere scores halen dan aspirant-studenten met twee in Nederlandse geboren ouders. We vatten deze bevindingen, die worden gepresenteerd in hoofdstuk 7, hierna kort samen. Daarbij tekenen we nadrukkelijk aan dat wat volgt vooral beschouwd moet worden als een aantal 'hypotheses' die voortkomen uit de analyse van de informatie over de toelatingstoetsen zoals die wordt gedeeld met studenten via [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl) en de opbrengst van gesprekken met een beperkt aantal informanten. Om deze hypothesen te toetsen, zou nader onderzoek nodig zijn.

De eerste bevinding die mogelijk kan bijdragen aan een verklaring voor de achterblijvende toetsresultaten van sommige studenten, heeft betrekking op de omvang van de stof die in de toelatingstoetsen wordt geoperationaliseerd. Bij alle drie de 'vakken' gaat het om een grote hoeveelheid stof die zich uitstrekt over een zeer breed terrein: van 'jagers en verzamelaars' tot en met Annie M.G. Schmidt en actuele technologische ontwikkelingen, van kennis over plaattektoniek en klimaten tot en met de verdeling van rijkdom en armoede en waterbeheer en van de metamorfose van sprinkhanen tot en met het veilige gebruik van spanningsmeters en het realiseren van stevigheid in producten door het gebruik van (onder meer) bogen en driehoeksprofielen.

Studenten die alle drie de toelatingstoetsen moeten doen, staan dan ook voor een enorme opgave. Wat dat betreft bevinden mbo 4-studenten zich in een ongunstige positie ten opzichte van havisten die doorgaans maar één of maximaal twee toelatingstoetsen hoeven te doen. Bijlagen 6, 7 en 8 laten zien dat voor alle drie de onderzochte toelatingstoetsen geldt dat de kandidaten met een niet-westerse migratieachtergrond vaker een mbo-opleiding hebben dan de kandidaten met twee in Nederland geboren ouders. Vooral bij de toets Natuur & Techniek zijn mbo-ers oververtegenwoordigd onder de kandidaten met een niet-westerse achtergrond.

Een tweede kenmerk van de toelatingstoetsen dat zou kunnen bijdragen aan een verklaring voor de achterblijvende toetsresultaten van studenten met een niet-westerse migratieachtergrond, heeft betrekking op de manier waarop de inhoud van de toetsen in de Handreikingen wordt beschreven. Het met begrip lezen van de Handreikingen vereist een niveau van leesvaardigheid dat niveau 3F van het Referentiekader Taal overstijgt. De tekst is zó compact en geeft zo weinig aanknopingspunten bij kennis die bekend mag worden verondersteld bij aspirant-pabostudenten die zich op de toelatingstoetsen moeten voorbereiden, dat beheersingsniveau 3F niet meer volstaat om de Handreiking zonder



ondersteuning zelfstandig door te nemen. De hoge eisen die het lezen van de Handreiking stelt aan de leesvaardigheid van aspirant-studenten, zijn in beginsel voor álle aspirant-studenten met een havo- of mbo 4-diploma problematisch. De leesproblemen zullen echter groter zijn naarmate het referentiekader en de voorkennis van de aspirant-studenten verder verwijderd zijn van de inhoud van de Handreikingen.

Dit onderzoek heeft zich niet uitgestrekt tot een analyse van de leerboeken die via [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl) worden geadviseerd. Een eerste verkenning geeft aan dat ook voor deze leerboeken geldt dat ze in relatief weinig tekst héél veel leerstof beschrijven en dat het zelfstandig lezen een leesvaardigheidsniveau vereist dat hoger is dan 3F. We komen hierna terug op de vraag in hoeverre aspirant-kandidaten een beroep kunnen doen op ondersteuning bij het verwerken van de informatie in de Handreikingen en de aangereikte leerboeken.

## 8.3 Partijdigheid bij de voorbereiding

Met betrekking tot de beantwoording van de vraag naar mogelijke partijdigheid in de voorbereiding op de toelatingstoetsen, heeft het onderzoek geleid tot een aantal bevindingen die nauw samenhangen met het voorafgaande. Opnieuw geven we met nadruk aan dat deze bevindingen vooralsnog niet veel meer zijn dan 'onderbouwde hypothesen' en nader onderzoek vergen.

Bij de voorbereiding op de toelatingstoetsen zijn aspirant-studenten in de allereerste plaats aangewezen op zelfstudie. Zij vinden daarbij weliswaar een keur aan ondersteunend materiaal en mogelijkheden om lessen te volgen via [www.goedvoorbereidnaardepabo.nl](http://www.goedvoorbereidnaardepabo.nl), maar ze zullen hoofdzakelijk op eigen initiatief en op eigen kracht hun weg moeten vinden in al die adviezen, materialen en cursussen. Daar waar ondersteunend onderwijs wordt aangeboden - landelijk, door pabo's en door mbo-opleidingen - gaat het doorgaans om een aantal contacturen dat heel beperkt is in verhouding tot de hoeveelheid stof die behandeld moet worden. Zelfs het keuzedeel 'Vorbereiding Pabo' dat veel mbo-instellingen aanbieden, is in dat opzicht relatief beperkt. Het keuzedeel is bovendien niet uitsluitend gericht op de voorbereiding op de toelatingstoetsen maar ook op het verwerven van de studievaardigheden die de studenten in het hoger beroepsonderwijs nodig zullen hebben.

De analyse van de Handreikingen heeft duidelijk gemaakt dat die zich niet goed lenen voor zelfstudie. Vooral het ontbreken van een 'narratief' of 'structuur' in de beschreven stof, maakt het studenten moeilijk om zonder intensieve begeleiding relaties binnen de beschreven kennis te doordenken en te koppelen aan eigen ervaringen. De te verwerven kennis kan daardoor niet of nauwelijks worden opgeslagen als de 'powerful knowledge' die de studenten op de pabo nodig zullen hebben.

De gevoerde gesprekken met aspirant-studenten maken duidelijk dat veel studenten hun toevlucht nemen tot het 'stampen' van de kennis. Zij maken daarbij vooral dankbaar gebruik van de beschikbare voorbeeldexamens. Deze voorbeeldexamens kunnen studenten die nog niet 'boven de stof' staan en die moeite hebben met de denkvaardigheden die worden vereist, het gevoel geven dat ze werkelijk élk feitje moeten kennen. Dit kan de strategie van 'stampen' versterken. Dat alle studenten die recentelijk voor een of meer van de toetsen zijn geslaagd, bekennen dat ze denken dat ze zonder nieuw stampwerk een volgende keer niet opnieuw zouden slagen voor de toelatingstoetsen, sluit aan bij wat bekend is uit veel onderzoek over leren: kennis die wordt opgeslagen door 'massed practice' en 'stampen', beklijft zelden.

Hoewel bovenstaande hypothesen met betrekking tot mogelijke nadelen van de manier waarop kandidaten zich op de toelatingstoetsen moeten voorbereiden, waarschijnlijk voor élk kandidaat opgaan, zijn er redenen om aan te nemen dat de genoemde nadelen van het accent op zelfstudie sterker zijn naarmate kandidaten over minder relevante voorkennis beschikken als ze aan de voorbereiding beginnen én naarmate ze minder gebruik kunnen maken van steun uit hun omgeving. Waarschijnlijk geldt dat vaker voor aspirant-studenten met een niet-westerse achtergrond dan voor aspirant-studenten met twee in Nederland geboren ouders en een bijbehorend sociaal netwerk.

## 8.4 Bespreekpunten

De omvang van dit onderzoek en de resultaten die daaruit voortkomen, staan niet toe om al te stevige conclusies te trekken omtrent concrete maatregelen die zouden kunnen bijdragen aan meer succeskansen van aspirant-studenten met een niet-westerse achtergrond. In plaats van concrete aanbevelingen, sluiten we dit rapport dan ook af met suggesties voor een aantal discussiepunten die zich tijdens het onderzoek hebben aangediend. Bij de bespreking van deze punten zou de aandacht niet alleen moeten uitgaan naar aspirant-pabostudenten met een niet-westerse migratieachtergrond. Heroverweging van de besluiten die bij de implementatie van de aanvullende toelatingseisen zijn genomen, kan in onze ogen voor alle aspirant-studenten, ongeacht hun etnische achtergrond, relevant zijn.

Alvorens de bespreekpunten die uit dit onderzoek naar voren komen te presenteren, stellen we vast dat elke discussie die naar aanleiding van het onderhavige onderzoek gevoerd wordt, zou moeten starten met overeenstemming omtrent wat het startpunt van élk debat over toetsen en examens zou moeten zijn: de functie van de toetsen en examens die aan de orde zijn. Welke beslissingen moeten met behulp van de te verzamelen toetsresultaten genomen worden? Duidelijkheid over de functie van de toelatingstoetsen is nodig om vast te kunnen stellen welke informatie over de aspirant-studenten daarvoor nodig is en welke toetsvormen en beoordelingsmodellen daarbij passen. Die duidelijkheid is ook nodig om de validiteit te evalueren van de toetsen die worden ontwikkeld om die informatie te verzamelen.

De functie van de toelatingstoetsen voor de pabo ligt vast in artikel 7.25 van in de WHW: aspirant-studenten moeten aantonen dat zij over voldoende kennis beschikken om aan het onderwijs van de opleiding tot leraar basisonderwijs deel te nemen. Deze functie zou het startpunt moeten zijn van het gesprek over de volgende punten.

### 1. *De inhoud van de toelatingseisen.*

Het lijkt ons, meer dan vijf jaar na de invoering van de toelatingseisen, een goed moment om de inhoud van die eisen te herijken. Daarbij zou overwogen kunnen worden om opnieuw te laten onderzoeken in hoeverre de huidige toelatingseisen allemaal relevant zijn voor de opleiding tot leerkracht basisonderwijs en/of voor de 'startbekwaamheid' van pabo-abituriënten voor onderwijs over aardrijkskunde, geschiedenis en natuur en techniek aan basisschoolleerlingen. Het zou goed zijn om daarbij nadrukkelijker dan tot nu toe stil te staan wat de 'powerful knowledge' is die studenten en leerkrachten nodig hebben om zélf te leren en om anderen te stimuleren om te leren. Het gaat niet alleen om (heel veel) feitenkennis, maar ook om de structuur oftewel het 'narratief' dat bij die feiten hoort en de manier waarop die feiten aansluiten bij alledaagse ervaringen. Voorop zou moeten staan dat pabostudenten moeten leren om hun kennis wendbaar toe te passen en om leerlingen op de basisschool tot verwondering te brengen.

### 2. *De beschrijving van de toelatingseisen*

Een tweede bespreekpunt betreft de beschrijving van de toelatingseisen. Op welke manier kunnen de toelatingseisen zó worden beschreven dat de inhoud en relevantie ervan duidelijk worden voor de aspirant-kandidaten? Het gaat hier om tekstenkenmerken die de informatie leesbaar maken voor aspirant-studenten met leesvaardigheid op niveau 3F en – daarmee samenhangend – om de explicitering van de structuur/het 'narratief' in de leerdoelen en de contextualisering ervan in alledaagse ervaringen.

### 3. *De vorm van de toelatingstoetsen*

Ook de operationalisering van de toelatingseisen zou heroverwogen moeten worden. Is een momentopname in de vorm van een meerkeuzetoets inderdaad de meest passende toetsvorm voor de toelatingseisen? Welke terugslag-effecten willen we bereiken dan wel voorkomen? Welke andere toetsvormen zouden (eventueel naast meerkeuzetoetsen) passend en uitvoerbaar kunnen zijn om de beheersing van de 'powerful knowledge' die pabostudenten nodig hebben te meten én om de ontwikkeling daarvan te stimuleren?

### 4. *De voorbereiding op de toetsen*

Het onderzoek waarvan hier verslag is gedaan, heeft ook argumenten opgeleverd om de keuzes die gemaakt zijn omtrent de manier waarop aspirant-kandidaten zich op de toelatingstoetsen moeten voorbereiden, te heroverwegen. In hoeverre is het realistisch om van havo-leerlingen en mbo-studenten te verwachten dat ze zich de grote hoeveelheid kennis die in de drie Handreikingen is beschreven, in hun eindexamenjaar grotendeels via zelfstudie kennis eigen maken?



## Nawoord

Het onderzoeksteam dat deze studie heeft uitgevoerd, is zich vanaf de start van dit onderzoek bewust geweest van de mogelijke reikwijdte van de resultaten van dit onderzoek. Dit is niet het eerste onderzoek dat in Nederland is uitgevoerd naar mogelijke bias in toetsen voor kandidaten met verschillende etnische achtergronden. Vooral in de context van de voormalige Cito Eindtoets Basisonderwijs is er veel aandacht besteed aan onderzoek dat mogelijke benadeling van leerlingen uit uiteenlopende etnische groepen moest helpen voorkomen. In deze studie hebben we dankbaar gebruik gemaakt van de uitkomsten van onderzoek op dit terrein dat veel diepgaander en uitgebreider was dan wat wij in de context van onze opdracht konden waarmaken (Uiterwijk, 1994; Uiterwijk & Vallen, 1997; 2005; Van Schilt-Mol, 2006). Wat ons vanaf het begin echter vooral heeft beziggehouden, zijn de bijzondere kenmerken van de groep die in de onderhavige studie centraal staat: aspirant-pabostudenten met een niet-westerse achtergrond die in het bezit zijn van een havo- of een mbo 4-diploma en die benadeeld lijken te worden bij het afleggen van toetsen die kennis meten op niveau havo-3/vmbo-tl. We hebben ons keer op keer afgevraagd wat de betekenis is van het feit dat we op zoek waren naar verklaringen voor achterblijvende toetsprestaties van aspirant-studenten met een niet-westerse achtergrond die, op een paar uitzonderingen na, allemaal hun hele schoolloopbaan op Nederlandse scholen hebben gezeten. En die daar zo succesvol zijn geweest dat ze - zolang er geen sprake is van bijzondere toelatingseisen - na het behalen van hun diploma toelaatbaar zijn tot het hoger onderwijs.

Deze vraag intrigeert extra in het licht van alle onderzoeksgegevens (OESO, 2018; Onderwijsraad, 2019; Borghans, 2018) die duidelijk maken dat ook in ons land het ideaal van 'gelijke onderwijskansen voor ieder kind', ondanks heel veel inspanningen, nog lang niet is bereikt. Uitgaande van de gegevens over de factoren waarvan bekend is dat ze nog steeds bijdragen aan kansengelijkheid in het onderwijs - in de eerste plaats het opleidingsniveau, de sociaaleconomische status en de geboorteplaats van ouders -, hebben wij ons gerealiseerd dat we in ons onderzoek te maken zouden krijgen met een bijzondere groep niet-westerse studenten. Het gaat om studenten die zich niet hebben laten afschrikken door de aanvullende toelatingseisen voor de pabo en die, zeker in vergelijking met leeftijdsgenoten met een vergelijkbare achtergrond, tot nu toe bovengemiddeld hebben gepresteerd in het onderwijs. Het zijn studenten die erin geslaagd zijn alle toetsen, alle selectiemomenten en alle negatieve verwachtingen waar velen van hen vanaf de instroom in groep 1 van het basisonderwijs mee zijn geconfronteerd, succesvol door te komen. Vaak zullen ze dat hebben gedaan zonder heel veel ondersteuning van thuis en ondanks alle belemmeringen die ons onderwijs voor sommige leerlingen nog steeds lijkt op te werpen. Wat is de betekenis van het feit dat deze selecte groep aspirant-studenten, ook wanneer er wordt gecorrigeerd voor andere achtergrondkenmerken waarvan bekend is dat ze samenhangen met schoolsucces, gemiddeld significant lager scoort op de toelatingstoetsen voor de pabo en vaker behoefte lijkt te hebben aan extra ondersteuning bij de voorbereiding op de toetsen dan leeftijdsgenoten met twee in Nederland geboren ouders? Deze vraag hoort waarschijnlijk beter thuis in de dialoog over kansengelijkheid in het onderwijs dan in onderzoek naar de kwaliteit van de toelatingstoetsen voor de pabo. Dit onderzoek geeft er in elk geval geen antwoord op.





# Literatuurlijst

- Beneker, T. & Van Gaans, G. (2018). Kennis voor de toekomst? In Beneker, T (red.) **Toekomstgericht onderwijs in de maatschappijvakken. een vakdidactisch perspectief vanuit aardrijkskunde, economie, geschiedenis, levensbeschouwing en maatschappijleer**. Amsterdam, LEMM, pp. 30-41.
- Bohenn, E., Jan sen, F., Kuijpers, C., Thijssen, R., Schot, I. & Stockmann, W. (2007). **Raamwerk Nederlands Nederlands in (v)mbo-opleiding, beroep en maatschappij**. 's-Hertogenbosch: CINOP.
- Bolhaar, J., & Van Ruijven, K. (2019). **Het effect van bijzondere nadere vooropleidingseisen op de pabo**. CPB.
- Borghans, L., Diris, R. & Schils, T. (2018). Sociale ongelijkheid in het onderwijs is hardnekkig. <https://didactiefonline.nl/blog/blonz/sociale-ongelijkheid-in-het-onderwijs-is-hardnekkig>
- Broeder, P. , Extra, G., Habraken, M., Van Hout, R., & Keurentjes, H. (1993). **Taalgebruik als indicator van etniciteit: Een studie naar identificatie van allochtone leerlingen**. Tilburg: Tilburg University Press. Ook te downloaden van <https://research.tilburguniversity.edu/en/publications/taalgebruik-als-indicator-van-etniciteit-een-studie-naar-identifi>
- Bügel, K., & Sanders, P. (2006). Toetsbias en itembias. **EXAMENS**, nr.3, p.25.
- Bügel, K. & Sanders, P.F. (1998). **Richtlijnen voor de ontwikkeling van onpartijdige toetsen**. Arnhem: Cito.
- Çankaya, S. (2020). **Mijn ontelbare identiteiten**. Amsterdam: De Bezige Bij.
- Council of Europe (2001). **A Common European Framework of Reference for Languages: learning, teaching, assessment**. Cambridge: Cambridge University Press. Ook te downloaden van [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_en.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf).
- Diephuis, R., Sluijsmans, L., Van der Hoeven, M., Beker, T., Van Graft, T., Oorschot, F., Heldens, H., Van Herwaarden, F., Spek. W., & Stroop, R. (2013). **Naar een instapniveau pabo. Advies inzake bijzondere nadere vooropleidingseisen voor aardrijkskunde, geschiedenis en natuur & techniek**. Enschedé: SLO.
- Driessen, G. (1997). Indicatoren Van Etniciteit in Relatie Tot Predictoren Van Taalvaardigheid in Het Basisonderwijs. **Toegepaste Taalwetenschap in Artikelen**, Vol. 56, pp. 89–103.
- Expertgroep Doorlopende leerlijnen taal en rekenen (2008). **Over de drempels met taal en rekenen. Hoofdrapport van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen**. Enschede: SLO.
- Expertgroep Doorlopende leerlijnen taal en rekenen (2009). **Referentiekader taal en rekenen. De referentieniveaus**. Enschede: SLO.
- Howson, J & Shemilt, D. (2011) Frameworks of knowledge: dilemmas and debates. In Davies, I. (ed) **Debates in History Teaching**, Londen en New York, Routledge, pp. 73-83.

- Kerkhoff, A. & Vallen, T. (1986). Cultural Biases in Second Language Testing of Children. In: Extra, G. & Vallen, T. (eds.), (1986), **Ethnic minorities and Dutch as a second language** (pp. 133-147). Dordrecht: Foris.
- Ministerie van OCenW (2019). **Werken in het onderwijs**. Brief aan de voorzitter van de Tweede Kamer der Staten Generaal. ([link](#))
- OESO (2018). **Equity in education, Breaking down barriers to social mobility**. Paris: ISA, OECD Publishing.
- Onderwijsraad (2019). **Doorgeschoten differentiatie in het onderwijsstelsel. Stand van educatief Nederland**. Den Haag: Onderwijsraad.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How Much Knowledge Is Too Little? When a Lack of Knowledge Becomes a Barrier to Comprehension. **Psychological Science** 30(9), pp. 1344-1351.
- Perfetti, C. & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. **Scientific Studies of Reading**, 18(1), 22-37. Geraadpleegd op internet via: <https://www.tandfonline.com/doi/abs/10.1080/10888438.2013.827687>
- Shemilt, D. (2000), The Caliph's Coin: The Currency of Narrative Frameworks in History Teaching. In Stearns, P.N., Seixas, P. & Wineburg, S. (eds). **Knowing, Teaching and Learning History. National and International Perspectives**. New York & Londen, New York University Press, pp. 83-101.
- Sluijsmans, R. (red.). (2013a). **Toelichting bij de bijzondere nadere vooropleidingseisen pabo. Aardrijkskunde**. Enschedé: SLO.
- Sluijsmans, R. (red.). (2013b). **Toelichting bij de bijzondere nadere vooropleidingseisen pabo. Geschiedenis**. Enschedé: SLO.
- Sluijsmans, R. (red.). (2013c). **Toelichting bij de bijzondere nadere vooropleidingseisen pabo. Natuur & Techniek**. Enschedé: SLO.
- Uiterwijk, H. (1994). **De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen**. Arnhem: Cito (proefschrift Katholieke Universiteit Brabant).
- Uiterwijk, H., & Vallen, T. (1997). Onderzoek naar bias voor allochtone leerlingen in de Cito-Eindtoets Basisonderwijs. In: **Pedagogische Studiën**, 74 (1), p. 21-32.
- Uiterwijk, H. & Vallen, T. (2005). Linguistic sources of item bias for second generation immigrants in Dutch tests. **Language Testing**, 22 (2), pp. 211-234.
- Van Casteren, W., Cuppen, J., Lodewick, J., & Termorshuizen, T. (2019). **Evaluatie vooropleidingseisen pabo**. Nijmegen: ResearchNed.
- Van Schilt-Mol, T. (2006). **Differential Item Functioning en Itembias in de Cito-Eindtoets Basisonderwijs**. (proefschrift Katholieke Universiteit Brabant). Amsterdam: Aksant Academic Publishers.
- Van Steensel, R., & Segers, E. (red.). (2017). **Succesvol lezen in het onderwijs**. Stichting Lezen Reeks, deel 28. Delft: Eburon.
- Von Graevenitz, J., & Visser, M. (2020). **Urban onderwijs. Docenten over gelijke kansen in een veranderende wereld**. Studio Noodzaak.
- Young, M. & Muller, J. (2013). On the powers of powerful knowledge. **Review of Education**, 1(3), pp. 229–250.
- Vuijsje, R. (2020). **Maar waar kom je écht vandaan? 100 verhalen over de Nederlandse identiteit**. Amsterdam, Lebowski publishers.
- Zumbuehl, M & Dillingh, R. (2020). **Ongelijkheid van het jonge kind**. CPB Notitie. CPB.



## Bijlagen

### Bijlage 1 Informatie voor aspirant-studenten en vragenlijst 'kenmerken aspirant-studenten'

Beste aspirant-pabostudent,

Je hebt je aangemeld om deel te nemen aan één of meer toelatingstoetsen voor de pabo. Dat is fijn want het Nederlandse basisonderwijs heeft veel nieuwe en bekwame leerkrachten nodig.

De toelatingstoetsen zijn bedoeld om te selecteren: alleen de kandidaten die voor de toetsen slagen worden tot de pabo toegelaten. Vanwege dat grote belang heeft de Tweede Kamer aan de Minister van Onderwijs gevraagd of elke kandidaat, ongeacht zijn etnische of sociaal-economische achtergrond, een eerlijke kans krijgt om voor deze toetsen te slagen. De Minister heeft aan de Vereniging Hogescholen opdracht gegeven om de vraag van de Tweede Kamer te onderzoeken. De Vereniging Hogescholen heeft Cito en Fontys Lerarenopleiding Tilburg gevraagd om dat onderzoek uit te voeren.

Wat betekent dit voor jou? Als je straks klaar bent met de toelatingstoets, vragen we je of je mee wilt werken aan dit onderzoek en of we jouw toetsresultaten voor dit onderzoek mogen gebruiken. Als je daar 'ja' op zegt, dan zullen we je nog een paar extra (digitale) vragen stellen. Je antwoorden op die vragen gaan we gebruiken bij de analyse van de toetsresultaten. We voegen de extra vragen die we zullen stellen, ter oriëntatie toe aan deze brief.

Het spreekt voor zich dat het straks voor je uitslag of uitslagen helemaal niets uitmaakt of je aan dit aanvullende onderzoek mee wilt werken. Als je meedoet, dan zullen je antwoorden op de extra vragen alléén gebruikt worden voor ons onderzoek naar mogelijke 'partijdigheid' van de toetsen. Daarbij houden we ons natuurlijk aan de algemene verordening gegevensbescherming (AVG).

Mocht je vragen hebben, neem dan contact op: [a.kerkhoff@fontys.nl](mailto:a.kerkhoff@fontys.nl).

We hopen van harte dat je aan dit onderzoek mee wilt werken!

Met vriendelijke groet,  
namens het projectteam,

Anne Kerkhoff, projectleider  
Fontys Lerarenopleiding Tilburg

## Informatie over de aanvullende vragen

Als je meewerkt aan het onderzoek, stellen we je aan het eind van de toelatingstoets een paar vragen over jezelf en je achtergrond. We hebben die informatie nodig om ons onderzoek naar de toetsen te doen. De informatie heeft pertinent geen invloed op je uitslag.

Veel kandidaten moeten meer dan één van de drie toelatingstoetsen (Aardrijkskunde, Geschiedenis en Natuur & Techniek) doen. Zij hoeven straks, als ze willen meewerken, natuurlijk niet elke keer opnieuw de vragen te beantwoorden. Als je eenmaal één keer akkoord bent gegaan en de vragen hebt beantwoord, dan kun je dit als je de volgende keer een toelatingstoets doet, aangeven. Je hoeft de vragen dan niet nog een keer te beantwoorden.

We zullen je de volgende vragen stellen. Het zijn steeds meerkeuzevragen.

*LET OP: je hoeft deze vragen nu niet te beantwoorden. We sturen je deze vragen alleen toe om je te helpen bij je besluit of je straks mee wilt werken. We zullen de vragen pas echt stellen als je aan het eind van de toets hebt aangegeven dat je akkoord gaat met deelname aan het onderzoek. Je zult de vragen dan beantwoorden via de computer waarmee je ook de toetsvragen hebt beantwoord.*

### 1. Geboorteland

In welk land ben je geboren?

In welk land is je moeder geboren?

In welk land is je vader geboren?

*Bij deze vragen kun je steeds kiezen uit een 'rolmenu' met de actuele namen van alle landen van de wereld.*

### 2. Taalgebruik

Welke taal of talen spreek je thuis met je ouders?

- A. vooral Nederlands (of Fries of een Nederlands dialect)
- B. vooral een andere taal dan Nederlands
- C. even vaak Nederlands als een andere taal

Welke taal of talen spreek je in je vrije tijd met vrienden en bekenden?

- A. vooral Nederlands (of Fries of een Nederlands dialect)
- B. vooral een andere taal dan Nederlands
- C. even vaak Nederlands als een andere taal

### 3. Onderwijservaring

Hoelang heb je in Nederland op de basisschool gezeten?

- A. de hele basisschool (groep 1 tot en met 8)
- B. een deel van de basisschool
- C. nooit, ik heb niet in Nederland op de basisschool gezeten

### 4. Leesgedrag

Hoe vaak lees je in je vrije tijd **boeken**?

- A. (bijna) nooit
- B. een paar keer per maand
- C. een paar keer per week
- D. (bijna) elke dag

Hoe vaak lees je in je vrije tijd **kranten of tijdschriften**?

- A. (bijna) nooit
- B. een paar keer per maand
- C. een paar keer per week
- D. (bijna) elke dag

Hoe vaak lees je in je vrije tijd **nieuws en artikelen op internet**?

- A. (bijna) nooit
- B. een paar keer per maand
- C. een paar keer per week
- D. (bijna) elke dag



*5. Betrokkenheid*

Bij welk gebied voel jij je het meest betrokken?

- A. een plaats of regio in Nederland
- B. Nederland als geheel
- C. een land of gebied buiten Nederland

*6. SES*

Wat is (ongeveer) het opleidingsniveau van je moeder?

- A. basisonderwijs
- B. vmbo, mavo, mbo 1, 2 of 3
- C. havo, mbo 4 of vwo
- D. hbo of universiteit
- E. dat weet ik niet

*Wat is (ongeveer) het opleidingsniveau van je vader?*

- A. basisonderwijs
- B. vmbo, mavo, mbo 1, 2 of 3
- C. havo, mbo 4 of vwo
- D. hbo of universiteit
- E. dat weet ik niet

# Bijlage 2 Informatie voor pabo-contactpersonen

Tilburg, september 2020

Beste collega,

Binnenkort zal er weer een groot aantal aspirant-pabostudenten, waarvan velen de opleiding al zijn gestart doordat ze voorlopig zijn toegelaten, deelnemen aan een of meer onderdelen van de Toelatingstoetsen Pabo. De kandidaten zullen aan het einde van de toets gevraagd worden om deel te nemen aan een onderzoek. We willen je in deze brief graag wat nadere informatie geven over dit onderzoek. We hopen dat je daarmee voldoende voorbereid zult zijn op eventuele vragen van aspirant-studenten en studenten.

Het betreft een onderzoek dat Fontys Lerarenopleiding Tilburg en Cito samen uitvoeren in opdracht van de Vereniging Hogescholen. De aanleiding voor dit onderzoek zijn de vragen van de Tweede Kamer naar aanleiding van het aangetoonde drukkende effect van de invoering van de toelatingstoetsen Pabo op de instroom van kandidaten met een niet-westerse achtergrond. De Minister heeft de Tweede Kamer naar aanleiding van de Kamervragen drie toezeggingen gedaan:

- a. summercourses,
- b. digitale materialen om studenten te helpen om zich voor te bereiden en
- c. onderzoek naar de vraag of de toetsen aangepast moeten worden om de onevenredig grote uitval van studenten met een niet-westerse achtergrond te voorkomen.

Het beloofde onderzoek is, via de Vereniging Hogescholen/10 voor de Leraar, opgedragen aan Fontys Lerarenopleiding Tilburg (FLOT) en Cito. FLOT en Cito hebben, in overleg met 10 voor de Leraar, een onderzoeksplan ontwikkeld dat bestaat uit twee delen:

een inhoudelijke analyse, uitgevoerd door medewerkers van FLOT, van het examenprogramma en de opgaven die zicht moet geven op mogelijke culturele en/of talige biasbronnen en

een psychometrische analyse van de toetsdata door Cito.

Ten behoeve van die analyse wordt de kandidaten aan het einde van de toets, nadat ze de 60 toetsvragen hebben beantwoord en vóórdat ze de toets afsluiten, gevraagd om mee te werken aan het onderzoek. Als ze daarmee akkoord gaan, krijgen ze digitaal een beperkt aantal vragen naar persoonsgegevens voorgelegd. Deze dataverzameling zal plaatsvinden in het toetsvenster van september. Alleen als dat nodig is om voldoende kandidaten bereid te vinden om mee te werken, zullen we ook in oktober data verzamelen.

In de bijlage bij deze brief vind je, ter informatie, de brief die kandidaten vóór het toetsvenster waar ze aan gaan deelnemen hebben ontvangen én de extra vragen die gesteld worden aan degenen die bereid zijn om mee te werken.

We hopen je hiermee voorlopig voldoende geïnformeerd te hebben. Uiteraard zullen we je t.z.t. informeren over de uitkomsten van het onderzoek.

Voor nadere informatie kun je terecht bij [a.kerkhoff@fontys.nl](mailto:a.kerkhoff@fontys.nl).

Namens de projectgroep,

Anne Kerkhoff

Fontys Lerarenopleiding Tilburg, lector Talendidactiek

## Bijlage 3 Zes opgaven



86663

### Turfstekers

Tot aan het midden van de vorige eeuw werd er in Nederland veel turf gestoken. Turfstekers staken de turf in rechthoekige blokken uit de grond. Ze droogden de turfblokken en verkochten ze als brandstof voor kachels.

**Uit welke grondstof werd Turf gewonnen?**

- A uit kleigrond
- B uit löss
- C uit veengrond
- D uit zandgrond



86926

## Plaatgrenzen

Op de kaart zie je de grenzen tussen aardplaten in een deel van de wereld.

**Welke gevolgen heeft de ligging van deze platen voor het gebied op de kaart gehad?**

- A. Er groeien veel tropische bossen.
- B. Er woeden regelmatig tropische stormen.
- C. Er zijn hooggebergten met gletsjers.
- D. Er zijn vulkanen.



*Willibrord. Apostel der Friezen*

85837

## Schoolplaat over Willibrord

Willibrord was een belangrijke missionaris die het christendom in onze streken verkondigde. Op de schoolplaat staat de belangrijkste kerk die hij liet bouwen. Die kerk vormde het centrum van zijn activiteiten.

### Waar stond de kerk van Willibrord?

- A. in Leeuwarden
- B. in Maastricht
- C. in 's-Hertogenbosch
- D. in Utrecht



85941

## Slachtoffers in het jaar voor de bevrijding

In mei 1945 vierde Nederland de bevrijding. De bezetting was voorbij. Voor veel mensen was het laatste oorlogsjaar erg zwaar geweest. Vooral in de steden in het westen van Nederland vielen toen veel slachtoffers.

**Waardoor vielen er juist in dat jaar zoveel slachtoffers in de steden in het westen van Nederland?**

- A. De steden werden gebombardeerd door Duitse bommenwerpers.
- B. Er was te weinig voedsel in de steden waardoor mensen stierven van de honger.
- C. Mannen in de steden werden door de Duitsers opgepakt als gijzelaar en geëxecuteerd.
- D. Veel ondergedoken Joodse Nederlanders werden toen in de steden opgepakt en vermoord.



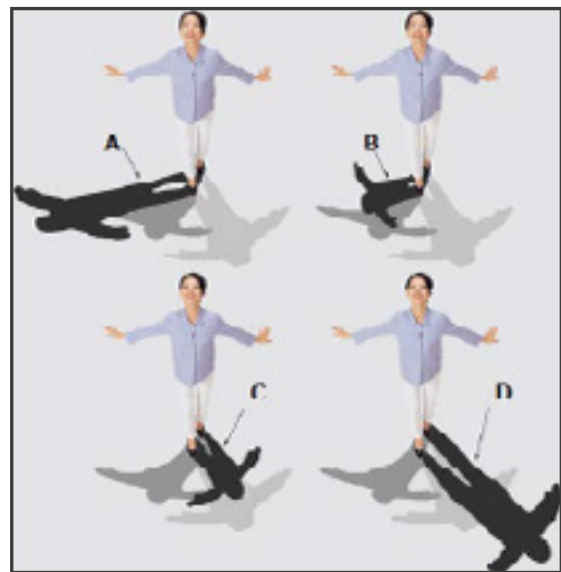
82802

## Energiebronnen

Op de afbeeldingen zie je voorbeelden van het gebruik van verschillende energiebronnen.

**Bij welk voorbeeld wordt gebruik gemaakt van een niet-duurzame energiebron?**

- A. bij de molen die draait door stromend water
- B. bij de auto die rijdt op plantaardige olie
- C. bij ijzer dat smelt door brandende steenkool
- D. bij het zeilschip dat vaart door de wind



Voorbeeldopgave uit Handreiking Aardrijkskunde

## Schaduw op het schoolplein

### Vraag 22 - Schaduw op het schoolplein

Om 10.00 uur en om 14.00 uur tekende juf Yasmin deze schaduw van zichzelf op het schoolplein.

**Hoe kan de getekende schaduw op die plaats er op het eind van de middag uit zien?**

- A. zoals bij A
- B. zoals bij B
- C. zoals bij C
- D. zoals bij D



## Bijlage 4 Etnische achtergrond

Het CBS<sup>11</sup> definieert 'migratieachtergrond' als volgt: "Kenmerk dat weergeeft met welk land een persoon verbonden is op basis van het geboorteland van de ouders of van zichzelf. Een persoon met een eerste generatie migratieachtergrond heeft als migratieachtergrond het land waar hij of zij is geboren. Een persoon met een tweede generatie migratieachtergrond heeft als migratieachtergrond het geboorteland van de moeder, tenzij dat ook Nederland is. In dat geval is de migratieachtergrond bepaald door het geboorteland van de vader.

Bij personen met een tweede generatie migratieachtergrond is er ook een onderscheid tussen personen met één of twee in het buitenland geboren ouders" ([Migratieachtergrond \(cbs.nl\)](#)).

Het CBS definieert een persoon met een niet-westerse migratieachtergrond als: "Persoon met als migratieachtergrond een van de landen in Afrika, Latijns-Amerika en Azië (exclusief Indonesië en Japan) of Turkije. Op grond van hun sociaal-economische en sociaal-culturele positie worden personen met een migratieachtergrond uit Indonesië en Japan tot de westerse migratieachtergrond gerekend. Het gaat vooral om mensen die in het voormalig Nederlands-Indië zijn geboren en werknemers van Japanse bedrijven met hun gezin". [Persoon met een niet-westerse migratieachtergrond \(cbs.nl\)](#).

Bij de classificatie van de kandidaten hebben we de volgende codering gehanteerd.

1 staat voor Nederland

2 staat voor westers

3 staat voor niet-westers

Afghanistan	3	Ecuador	3	Marokko	3	Spanje	2
Albanië	2	Egypte	3	Mongolië	3	Suriname	3
Algerije	3	Ethiopië	3	Namibië	3	Syrië	3
Argentinië	3	Filipijnen	3	Nauru	3	Taiwan	3
Armenië	3	Frankrijk	2	Nederland	1	Thailand	3
Aruba	3	Ghana	3	Nepal	3	Togo	3
Australië	2	Griekenland	2	Nicaragua	3	Tsjechië	2
België	2	Hongarije	2	Nieuw-Zeeland	2	Tunesië	3
Bolivia	3	India	3	Oekraïne	2	Turkije	3
Bosnië en Herzegovina	2	Indonesië	2	Oostenrijk	2	Venezuela	3
Brazilië	3	Irak	3	Pakistan	3	Verenigd Koninkrijk	2
Chili	3	Iran	3	Peru	3	Verenigde Staten	2
China	3	Italië	2	Polen	2	Vietnam	3
Colombia	3	Kaapverdië	3	Roemenië	2	Westelijke Sahara	3
Congo (Democratische Republiek)	3	Kameroen	3	Saint Kitts en Nevis	3	Zuid-Afrika	3
Curaçao	3	Kenia	3	Servië	2	Zweden	2
Denemarken	2	Kosovo	2	Seychellen	3	Zwitserland	2
Dominicaanse Republiek	3	Libanon	3	Sint-Maarten (Nederland)	3		
Duitsland	2	Litouwen	2	Somalië	3		

<sup>11</sup> Het CBS is bezig met een heroverweging van het gebruik van de hoofdingeling 'westers-niet westers'.

## Bijlage 5 Hoofdeffecten en significantie bij opleiding kandidaat

In de volgende tabel zijn de effectgroottes en significantieniveaus weergegeven voor alle kandidaten en uitgesplitst naar geen migratieachtergrond en niet-westerse migratieachtergrond. Omdat er voor Geschiedenis geen kandidaten waren met de vooropleiding Havo Maatschappij konden hier geen effectgroottes worden berekend. Hetzelfde gold voor kandidaten met de vooropleiding Havo Natuur en de toets voor Natuur & Techniek.

Opleiding A	Opleiding B	Alle kandidaten						Geen migratieachtergrond						Niet-westerse migratieachtergrond					
		AK		GS		NT		AK		GS		NT		AK		GS		NT	
		eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.	eff.	sign.
Anders	Havo Maatschappij	-0,12		x		0,07		-0,06		x		0,07		-0,16		x		-0,03	
Anders	Havo Natuur	-0,46	**	-0,16		X		-0,42	**	-0,17		X		-0,35		-0,02		X	
Anders	MBO Anders	-0,10		-0,01		-0,04		-0,06		0,02		-0,08		-0,34		-0,16		-0,03	
Anders	Onderwijsassistent	0,44	***	0,31	**	0,39	**	0,43	***	0,28	*	0,31	*	0,54		0,30		0,79	
Anders	Pedagogisch werk	0,00		0,01		-0,03		-0,11		0,07		-0,23		0,51		-0,42		1,15	
Havo Maatschappij	Havo Natuur	-0,35	**	x		X		-0,36	*	x		X		-0,19		x		X	
Havo Maatschappij	MBO Anders	0,01		x		-0,10		-0,01		x		-0,15		-0,19		x		0,00	
Havo Maatschappij	Onderwijsassistent	0,56	***	x		0,33	***	0,48	***	x		0,24	**	0,69	**	x		0,82	**
Havo Maatschappij	Pedagogisch werk	0,12		x		-0,10		-0,06		x		-0,30		0,67		x		1,18	*
Havo Natuur	MBO Anders	0,36	*	0,15		X		0,35	*	0,19		X		0,01		-0,14		X	
Havo Natuur	Onderwijsassistent	0,90	***	0,47	***	X		0,84	***	0,45	***	X		0,89		0,32		X	
Havo Natuur	Pedagogisch werk	0,46	*	0,17		X		0,30		0,24		X		0,86		-0,40		X	
MBO Anders	Onderwijsassistent	0,55	***	0,31	**	0,43	***	0,49	***	0,26	*	0,39	**	0,88	*	0,46		0,82	*
MBO Anders	Pedagogisch werk	0,10		0,01		0,01		-0,05		0,05		-0,15		0,86		-0,26		1,17	*
Onderwijsassistent	Pedagogisch werk	-0,44	*	-0,30		-0,42	*	-0,54	*	-0,22		-0,54	*	-0,02		-0,72		0,35	

significantieniveau alfa ( $\alpha$ ): \*\*\* =  $\alpha < 0,001$ ; \*\* =  $0,001 \leq \alpha < 0,01$ ; \* =  $0,01 \leq \alpha < 0,05$

# Bijlage 6 Geschiedenis

Bijlage 6 - deel 1 van 2

	GS	Aantallen				Percentages (per kolom)					Gemiddelden				Standaardafwijking			
		NL	WMA	NWMA	Totaal	NL	WMA	NWMA	Totaal	NL	WMA	NWMA	Totaal	NL	WMA	NWMA	pooled	
<b>TOTAAL</b>		668	33	93	794	100		100	100	100	0,14	0,07	0,07	0,13	0,25	0,22	0,25	0,25
<b>MOMENT</b>	5e	314	11	28	353	47		33	30	44	0,17	0,08	0,07	0,16	0,24	0,23	0,22	0,24
	6e	354	22	65	441	53		67	70	56	0,12	0,07	0,07	0,11	0,25	0,21	0,26	0,25
<b>AV Dyslexie</b>	Regulier	608	32	90	730	91		97	97	92	0,15	0,07	0,07	0,13	0,25	0,22	0,25	0,25
	Dyslectisch	60	1	3	64	9		3	3	8	0,12	0,07	-0,07	0,11	0,21		0,05	0,21
<b>AV Geslacht</b>	Man	214	11	16	241	32		33	17	30	0,19	0,11	0,15	0,18	0,26	0,16	0,23	0,26
	Vrouw	454	22	77	553	68		67	83	70	0,12	0,05	0,05	0,11	0,24	0,24	0,25	0,24
<b>AV Leeftijd 5 Eq.Size.Cats</b>	16-18jr	101	2	7	110	15		6	8	14	0,15	0,04	0,03	0,14	0,20	0,03	0,10	0,19
	19jr	153	11	11	175	23		33	12	22	0,07	-0,01	0,07	0,07	0,17	0,22	0,27	0,18
	20-21jr	210	6	24	240	31		18	26	30	0,10	0,07	-0,03	0,09	0,21	0,20	0,14	0,20
	22-27jr	113	7	26	146	17		21	28	18	0,15	0,09	0,06	0,13	0,24	0,15	0,24	0,24
	28-55jr	91	7	25	123	14		21	27	15	0,34	0,19	0,18	0,30	0,37	0,29	0,32	0,35
<b>AV Opleiding 6 code (DET)</b>	Anders	130	12	20	162	19		36	22	20	0,16	0,17	0,08	0,15	0,28	0,19	0,27	0,28
	Havo EM/CM	2	0	1	3	0		0	1	0	0,35	.	-0,17	0,18	0,22			
	Havo NT/NG	137	2	12	151	21		6	13	19	0,20	0,04	0,09	0,19	0,27	0,03	0,15	0,26
	MBO Anders	110	4	17	131	16		12	18	16	0,16	0,21	0,12	0,15	0,26	0,17	0,29	0,26
	Onderwijsassistent	258	13	36	307	39		39	39	39	0,09	-0,06	0,01	0,08	0,21	0,20	0,22	0,21
	Pedagogisch werk	31	2	7	40	5		6	8	5	0,15	0,12	0,19	0,15	0,21	0,25	0,31	0,23
<b>VL v2.1 Taal met Ouders</b>	Vooraf NLS	657	22	39	718	98		67	42	90	0,14	0,09	0,07	0,14	0,25	0,22	0,28	0,25
	Vooraf Anders	0	4	23	27	0		12	25	3	.	0,11	0,17	0,16		0,31	0,27	0,28
	Even vaak NLS/Anders	6	7	31	44	1		21	33	6	0,17	-0,02	-0,02	0,01	0,15	0,15	0,16	0,15
<b>VL v2.2 Taal met Bekenden</b>	Vooraf NLS	636	25	69	730	95		76	74	92	0,15	0,07	0,09	0,14	0,25	0,20	0,27	0,25
	Vooraf Anders	2	2	2	6	0		6	2	1	0,16	0,19	0,24	0,20	0,11	0,37	0,31	0,26
	Even vaak NLS/Anders	24	6	22	52	4		18	24	7	0,14	0,03	-0,02	0,06	0,16	0,26	0,14	0,16
<b>VL v3.1 Hoe lang BO</b>	Heel BO (1-8)	653	27	82	762	98		82	88	96	0,15	0,07	0,08	0,14	0,25	0,21	0,25	0,25
	Deel BO (<8jr)	10	2	2	14	1		6	2	2	0,03	-0,10	0,26	0,05	0,21	0,19	0,01	0,19
	Geen BO (0 jr in NL)	2	4	9	15	0		12	10	2	0,16	0,18	-0,06	0,03	0,11	0,23	0,18	0,19

## Bijlage 6 - deel 2 van 2

		Aantallen				Percentages (per kolom)					Gemiddelden				Standaardafwijking			
		GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	pooled	
VL v4.1 Lezen boek	(bijna) NOOIT	373	15	34	422	56		45	37	53	0,12	0,09	0,08	0,11	0,21	0,17	0,27	0,21
	Paar keer per MAAND	208	12	41	261	31		36	44	33	0,15	0,06	0,06	0,13	0,27	0,24	0,24	0,26
	Paar keer per WEEK	48	5	12	65	7		15	13	8	0,20	0,08	0,03	0,16	0,23	0,31	0,23	0,23
	(bijna) ELKE DAG	33	1	6	40	5		3	6	5	0,34	-0,06	0,09	0,29	0,42		0,31	0,40
VL v4.2 Lezen krant	(bijna) NOOIT	317	13	42	372	47		39	45	47	0,12	0,02	0,06	0,11	0,21	0,16	0,21	0,21
	Paar keer per MAAND	219	10	24	253	33		30	26	32	0,17	0,14	0,06	0,16	0,28	0,24	0,26	0,28
	Paar keer per WEEK	86	8	16	110	13		24	17	14	0,13	0,10	0,08	0,12	0,28	0,23	0,28	0,28
	(bijna) ELKE DAG	40	2	11	53	6		6	12	7	0,21	-0,02	0,09	0,18	0,28	0,44	0,35	0,29
VL v4.3 Lezen nieuw internet	(bijna) NOOIT	39	0	5	44	6		0	5	6	0,15	.	0,15	0,15	0,19		0,13	0,19
	Paar keer per MAAND	97	4	5	106	15		12	5	13	0,13	-0,02	-0,03	0,12	0,21	0,06	0,10	0,20
	Paar keer per WEEK	217	9	21	247	32		27	23	31	0,12	0,11	-0,03	0,11	0,21	0,25	0,18	0,21
	(bijna) ELKE DAG	308	20	62	390	46		61	67	49	0,17	0,07	0,10	0,15	0,29	0,22	0,27	0,28
VL v5.1 Betrokkenheid	Plaats / Regio NL	415	15	41	471	62		45	44	59	0,12	0,08	0,11	0,12	0,22	0,26	0,26	0,23
	Nederland (geheel)	234	12	44	290	35		36	47	37	0,19	0,05	0,05	0,16	0,28	0,19	0,25	0,27
	Land/gebied buiten NL	13	6	8	27	2		18	9	3	0,22	0,09	-0,02	0,12	0,38	0,18	0,16	0,27
VL v6.1 Opleiding MOEDER	basisonderwijs	9	2	17	28	1		6	18	4	0,32	0,27	0,18	0,24	0,49	0,27	0,31	0,37
	vmbo, mavo, mbo 1-3	224	11	22	257	34		33	24	32	0,15	0,05	-0,04	0,13	0,26	0,19	0,13	0,25
	havo, mbo 4 of vwo	231	8	21	260	35		24	23	33	0,13	0,04	0,12	0,13	0,23	0,24	0,28	0,24
	hbo of universiteit	176	10	24	210	26		30	26	26	0,16	0,15	0,05	0,14	0,24	0,16	0,23	0,24
	Weet ik niet	21	2	9	32	3		6	10	4	0,13	-0,28	0,03	0,08	0,19	0,07	0,22	0,20
VL v6.2 Opleiding VADER	Basisonderwijs	15	1	11	27	2		3	12	3	0,26	0,15	0,21	0,23	0,26		0,26	0,26
	vmbo, mavo, mbo 1-3	202	7	28	237	30		21	30	30	0,13	0,06	0,00	0,11	0,23	0,26	0,18	0,23
	havo, mbo 4 of vwo	205	8	26	239	31		24	28	30	0,10	0,16	0,04	0,10	0,23	0,24	0,26	0,23
	hbo of universiteit	206	14	17	237	31		42	18	30	0,20	0,04	0,07	0,18	0,28	0,17	0,28	0,27
	Weet ik niet	32	3	11	46	5		9	12	6	0,13	0,00	0,15	0,12	0,15	0,32	0,27	0,19

# Bijlage 7 Aardrijkskunde

Bijlage 7 - deel 1 van 2

	AK	Aantallen				Percentages (per kolom)				Gemiddelden				Standaardafwijking				
		NL	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	pooled	
<b>TOTAAL</b>		<b>703</b>	<b>28</b>	<b>90</b>	<b>821</b>	<b>100</b>		<b>100</b>	<b>100</b>	<b>100</b>	<b>0,17</b>	<b>0,12</b>	<b>0,06</b>	<b>0,16</b>	<b>0,22</b>	<b>0,20</b>	<b>0,26</b>	<b>0,22</b>
<b>MOMENT</b>	5e	325	6	41	372	46		21	46	45	0,17	0,15	0,02	0,15	0,21	0,23	0,23	0,22
	6e	378	22	49	449	54		79	54	55	0,18	0,11	0,09	0,16	0,22	0,20	0,28	0,23
<b>AV Dyslexie</b>	Regulier	647	26	89	762	92		93	99	93	0,17	0,10	0,06	0,16	0,22	0,21	0,26	0,22
	Dyslectisch	56	2	1	59	8		7	1	7	0,19	0,28	0,08	0,19	0,19	0,09		0,18
<b>AV Geslacht</b>	Man	201	10	16	227	29		36	18	28	0,23	0,13	0,19	0,23	0,21	0,14	0,30	0,22
	Vrouw	502	18	74	594	71		64	82	72	0,15	0,11	0,03	0,13	0,21	0,23	0,24	0,22
<b>AV Leeftijd 5 Eq.Size.Cats</b>	16-18jr	152	3	8	163	22		11	9	20	0,17	0,28	0,03	0,16	0,19	0,11	0,18	0,19
	19jr	153	9	12	174	22		32	13	21	0,11	0,08	0,02	0,10	0,16	0,20	0,15	0,16
	20-21jr	193	8	18	219	27		29	20	27	0,12	0,11	-0,03	0,11	0,19	0,16	0,19	0,19
	22-27jr	101	5	25	131	14		18	28	16	0,16	0,19	-0,03	0,12	0,20	0,26	0,16	0,19
	28-55jr	104	3	27	134	15		11	30	16	0,39	-0,06	0,24	0,35	0,25	0,26	0,34	0,27
<b>AV Opleiding 6 code (DET)</b>	Anders	98	9	11	118	14		32	12	14	0,19	0,11	0,09	0,17	0,21	0,20	0,31	0,22
	Havo EM/CM	129	7	13	149	18		25	14	18	0,20	0,28	0,13	0,20	0,20	0,07	0,15	0,19
	Havo NT/NG	91	2	6	99	13		7	7	12	0,28	0,40	0,18	0,27	0,22	0,22	0,25	0,22
	MBO Anders	101	1	18	120	14		4	20	15	0,20	-0,10	0,17	0,20	0,24		0,33	0,25
	Onderwijsassistent	247	9	35	291	35		32	39	35	0,10	-0,04	-0,04	0,08	0,19	0,11	0,21	0,19
	Pedagogisch werk	37	0	7	44	5		0	8	5	0,21	.	-0,03	0,17	0,25		0,22	0,25
<b>VL v2.1 Taal met Ouders</b>	Vooraf NLS	693	22	40	755	99		79	44	92	0,18	0,12	0,15	0,17	0,22	0,20	0,25	0,22
	Vooraf Anders	1	3	15	19	0		11	17	2	-0,02	0,07	0,06	0,06		0,32	0,34	0,34
	Even vaak NLS/Anders	6	3	35	44	1		11	39	5	0,18	0,17	-0,04	0,01	0,13	0,15	0,20	0,19
<b>VL v2.2 Taal met Bekenden</b>	Vooraf NLS	679	26	65	770	97		93	72	94	0,18	0,13	0,08	0,17	0,22	0,19	0,27	0,22
	Vooraf Anders	3	0	3	6	0		0	3	1	0,06	.	-0,13	-0,03	0,21		0,12	0,22
	Even vaak NLS/Anders	18	2	22	42	3		7	24	5	0,22	-0,04	0,01	0,10	0,22	0,35	0,20	0,21
<b>VL v3.1 Hoe lang BO</b>	Heel BO (1-8)	693	25	80	798	99		89	89	97	0,18	0,13	0,07	0,16	0,22	0,20	0,26	0,22
	Deel BO (<8jr)	7	1	7	15	1		4	8	2	0,11	-0,29	-0,04	0,01	0,19		0,22	0,20
	Geen BO (0 jr in NL)	0	2	3	5	0		7	3	1	.	0,20	-0,07	0,04		0,02	0,07	0,08

## Bijlage 7 - deel 2 van 2

		Aantallen				Percentages (per kolom)				Gemiddelden				Standaardafwijking				
		GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	pooled	
VL v4.1 Lezen boek	(bijna) NOOIT	373	14	32	419	53		50	36	51	0,16	0,15	0,08	0,15	0,20	0,14	0,32	0,21
	Paar keer per MAAND	211	8	39	258	30		29	43	31	0,17	0,03	0,05	0,14	0,23	0,25	0,22	0,23
	Paar keer per WEEK	73	3	15	91	10		11	17	11	0,23	0,02	0,02	0,19	0,24	0,14	0,22	0,23
	(bijna) ELKE DAG	43	3	3	49	6		11	3	6	0,26	0,28	0,18	0,26	0,25	0,31	0,26	0,25
VL v4.2 Lezen krant	(bijna) NOOIT	334	14	42	390	48		50	47	48	0,17	0,17	0,10	0,16	0,21	0,17	0,23	0,21
	Paar keer per MAAND	240	6	24	270	34		21	27	33	0,18	0,15	-0,01	0,17	0,22	0,17	0,30	0,23
	Paar keer per WEEK	85	7	13	105	12		25	14	13	0,12	0,03	0,02	0,10	0,19	0,22	0,24	0,19
	(bijna) ELKE DAG	41	1	10	52	6		4	11	6	0,31	-0,29	0,10	0,26	0,23	.	0,28	0,24
VL v4.3 Lezen nieuw internet	(bijna) NOOIT	47	1	2	50	7		4	2	6	0,16	-0,07	0,13	0,15	0,18	.	0,19	0,18
	Paar keer per MAAND	91	3	8	102	13		11	9	12	0,16	0,24	0,02	0,15	0,20	0,31	0,20	0,20
	Paar keer per WEEK	219	8	20	247	31		29	22	30	0,13	0,18	0,06	0,12	0,19	0,16	0,29	0,20
	(bijna) ELKE DAG	343	16	59	418	49		57	66	51	0,21	0,07	0,06	0,19	0,24	0,20	0,26	0,24
VL v5.1 Betrokkenheid	Plaats / Regio NL	459	13	48	520	65		46	53	63	0,15	0,12	0,08	0,14	0,20	0,20	0,26	0,21
	Nederland(als geheel)	221	10	35	266	31		36	39	32	0,22	0,07	0,07	0,20	0,24	0,23	0,27	0,25
	Land/gebied buiten NL	18	5	6	29	3		18	7	4	0,21	0,20	-0,12	0,14	0,16	0,17	0,09	0,15
VL v6.1 Opleiding MOEDER	basisonderwijs	5	0	13	18	1		0	14	2	0,54	.	0,16	0,26	0,33		0,39	0,40
	vmbo, mavo, mbo 1-3	221	9	24	254	31		32	27	31	0,16	0,06	0,01	0,15	0,21	0,14	0,20	0,21
	havo, mbo 4 of vwo	220	7	21	248	31		25	23	30	0,17	0,12	0,06	0,16	0,21	0,17	0,25	0,22
	hbo of universiteit	241	9	23	273	34		32	26	33	0,18	0,27	0,11	0,18	0,22	0,16	0,24	0,22
	Weet ik niet	11	3	8	22	2		11	9	3	0,22	-0,17	-0,10	0,05	0,22	0,21	0,16	0,19
VL v6.2 Opleiding VADER	Basisonderwijs	16	1	11	28	2		4	12	3	0,23	0,02	0,13	0,18	0,27	.	0,42	0,33
	vmbo, mavo, mbo 1-3	229	6	19	254	33		21	21	31	0,15	0,11	-0,04	0,14	0,20	0,17	0,18	0,20
	havo, mbo 4 of vwo	213	5	18	236	30		18	20	29	0,16	0,10	0,04	0,15	0,20	0,16	0,24	0,20
	hbo of universiteit	220	14	28	262	31		50	31	32	0,21	0,19	0,09	0,19	0,24	0,18	0,23	0,23
	Weet ik niet	19	2	13	34	3		7	14	4	0,21	-0,29	0,09	0,13	0,25	0,00	0,26	0,25

# Bijlage 8 Natuur & Techniek

Bijlage 8 - deel 1 van 2

	NT	Aantallen				Percentages (per kolom)				Gemiddelden				Standaardafwijking				
		GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	pooled	
<b>TOTAAL</b>		776	38	85	899	100		100	100	100	0,09	0,05	-0,06	0,08	0,18	0,20	0,18	0,18
<b>MOMENT</b>	5e	382	15	24	421	49		39	28	47	0,10	0,10	-0,06	0,09	0,17	0,19	0,17	0,17
	6e	394	23	61	478	51		61	72	53	0,09	0,01	-0,06	0,07	0,18	0,20	0,18	0,18
<b>AV Dyslexie</b>	Regulier	725	36	82	843	93		95	96	94	0,09	0,04	-0,06	0,08	0,18	0,20	0,18	0,18
	Dyslectisch	51	2	3	56	7		5	4	6	0,12	0,09	-0,14	0,10	0,16	0,08	0,04	0,15
<b>AV Geslacht</b>	Man	225	11	15	251	29		29	18	28	0,13	0,14	0,00	0,12	0,17	0,21	0,17	0,17
	Vrouw	551	27	70	648	71		71	82	72	0,08	0,01	-0,07	0,06	0,18	0,18	0,18	0,18
<b>AV Leeftijd 5 Eq.Size.Cats</b>	16-18jr	194	6	6	206	25		16	7	23	0,08	0,04	0,02	0,08	0,14	0,08	0,04	0,14
	19jr	165	10	9	184	21		26	11	20	0,05	-0,03	-0,05	0,04	0,14	0,17	0,16	0,14
	20-21jr	190	8	21	219	24		21	25	24	0,08	0,00	-0,14	0,05	0,17	0,14	0,14	0,16
	22-27jr	122	7	24	153	16		18	28	17	0,09	0,00	-0,08	0,06	0,19	0,16	0,19	0,19
	28-55jr	105	7	25	137	14		18	29	15	0,22	0,27	0,01	0,19	0,22	0,26	0,19	0,22
<b>AV Opleiding 6 code (DET)</b>	Anders	107	10	6	123	14		26	7	14	0,11	0,08	0,01	0,11	0,20	0,23	0,25	0,21
	Havo EM/CM	270	13	18	301	35		34	21	33	0,10	0,04	0,01	0,09	0,15	0,14	0,15	0,15
	Havo NT/NG																	0,00
	MBO Anders	105	5	18	128	14		13	21	14	0,13	0,13	0,01	0,11	0,20	0,30	0,21	0,20
	Onderwijsassistent	261	9	38	308	34		24	45	34	0,06	-0,04	-0,12	0,03	0,17	0,18	0,13	0,17
	Pedagogische werk	33	1	5	39	4		3	6	4	0,15	0,14	-0,18	0,11	0,19		0,15	0,19
<b>VL v2.1 Taal met Ouders</b>	Vooraf NLS	766	27	44	837	99		71	52	93	0,10	0,07	-0,04	0,09	0,18	0,21	0,17	0,18
	Vooraf Anders	1	4	15	20	0		11	18	2	0,26	-0,09	-0,04	-0,04		0,16	0,18	0,17
	Even vaak NLS/Anders	5	7	24	36	1		18	28	4	0,04	0,05	-0,11	-0,06	0,15	0,16	0,19	0,18
<b>VL v2.2 Taal met Bekenden</b>	Vooraf NLS	740	34	69	843	95		89	81	94	0,09	0,06	-0,04	0,08	0,18	0,20	0,18	0,18
	Vooraf Anders	2	1	2	5	0		3	2	1	0,15	0,14	-0,19	0,01	0,16		0,02	0,09
	Even vaak NLS/Anders	30	3	14	47	4		8	16	5	0,17	-0,08	-0,11	0,07	0,19	0,21	0,16	0,18
<b>VL v3.1 Hoe lang BO</b>	Heel BO (1-8)	765	31	73	869	99		82	86	97	0,10	0,04	-0,06	0,08	0,18	0,18	0,18	0,18
	Deel BO (<8jr)	6	2	5	13	1		5	6	1	0,10	0,15	-0,02	0,06	0,10	0,66	0,12	0,16
	Geen BO (0 jr in NL)	0	5	7	12	0		13	8	1	.	0,05	-0,13	-0,05		0,10	0,14	0,13

Bijlage 8 - deel 2 van 2

		Aantallen				Percentages (per kolom)				Gemiddelden				Standaardafwijking				
		GMA	WMA	NWMA	Totaal	GMA		WMA	NWMA	Totaal	GMA	WMA	NWMA	Totaal	GMA	WMA	NWMA	pooled
VL v4.1 Lezen boek	(bijna) NOOIT	399	15	25	439	51		39	29	49	0,07	0,12	-0,01	0,07	0,16	0,18	0,15	0,16
	Paar keer per MAAND	256	12	41	309	33		32	48	34	0,12	-0,05	-0,09	0,08	0,19	0,16	0,18	0,19
	Paar keer per WEEK	81	6	12	99	10		16	14	11	0,11	-0,02	-0,08	0,08	0,19	0,13	0,19	0,19
	(bijna) ELKE DAG	35	5	7	47	5		13	8	5	0,15	0,17	-0,01	0,13	0,18	0,28	0,21	0,19
VL v4.2 Lezen krant	(bijna) NOOIT	348	14	31	393	45		37	36	44	0,10	0,03	-0,07	0,08	0,18	0,12	0,17	0,17
	Paar keer per MAAND	271	11	22	304	35		29	26	34	0,09	0,07	-0,07	0,08	0,18	0,17	0,16	0,18
	Paar keer per WEEK	113	7	16	136	15		18	19	15	0,09	0,05	-0,08	0,07	0,17	0,37	0,19	0,18
	(bijna) ELKE DAG	39	6	15	60	5		16	18	7	0,12	0,04	0,01	0,09	0,19	0,17	0,21	0,19
VL v4.3 Lezen nieuw internet	(bijna) NOOIT	48	0	0	48	6		0	0	5	0,08	.	.	0,08	0,15			0,16
	Paar keer per MAAND	104	6	3	113	13		16	4	13	0,10	0,06	-0,13	0,09	0,18	0,11	0,23	0,18
	Paar keer per WEEK	218	11	16	245	28		29	19	27	0,07	0,00	-0,11	0,06	0,16	0,12	0,15	0,16
	(bijna) ELKE DAG	401	21	66	488	52		55	78	54	0,11	0,07	-0,04	0,08	0,19	0,24	0,18	0,19
VL v5.1 Betrokkenheid	Plaats / Regio NL	510	21	39	570	66		55	46	63	0,10	0,01	-0,06	0,08	0,17	0,19	0,16	0,17
	Nederland(als geheel)	244	10	38	292	31		26	45	32	0,09	0,09	-0,05	0,07	0,19	0,26	0,20	0,19
	Land/gebied buiten NL	17	7	8	32	2		18	9	4	0,13	0,08	-0,09	0,07	0,20	0,10	0,16	0,17
VL v6.1 Opleiding MOEDER	basisonderwijs	11	1	12	24	1		3	14	3	0,18	-0,11	0,02	0,09	0,20		0,20	0,20
	vmbo, mavo, mbo 1-3	238	14	13	265	31		37	15	29	0,10	0,00	-0,05	0,09	0,18	0,18	0,19	0,18
	havo, mbo 4 of vwo	253	7	22	282	33		18	26	31	0,09	-0,02	-0,07	0,08	0,18	0,12	0,15	0,17
	hbo of universiteit	240	14	26	280	31		37	31	31	0,09	0,16	-0,07	0,08	0,17	0,20	0,19	0,18
	Weet ik niet	29	2	12	43	4		5	14	5	0,09	-0,10	-0,11	0,02	0,16	0,31	0,14	0,16
VL v6.2 Opleiding VADER	Basisonderwijs	20	0	6	26	3		0	7	3	0,17	.	0,05	0,15	0,18		0,24	0,20
	vmbo, mavo, mbo 1-3	232	9	22	263	30		24	26	29	0,09	0,03	-0,10	0,07	0,18	0,23	0,13	0,18
	havo, mbo 4 of vwo	230	7	19	256	30		18	22	28	0,08	-0,01	-0,08	0,07	0,18	0,12	0,18	0,18
	hbo of universiteit	258	21	24	303	33		55	28	34	0,11	0,05	-0,05	0,09	0,17	0,17	0,19	0,17
	Weet ik niet	29	1	14	44	4		3	16	5	0,07	0,61	-0,04	0,05	0,16		0,17	0,17



## Bijlage 9 Het bepalen van gecorrigeerde effectgroottes

We hebben in de analyses effectgroottes bepaald voor groepen die van elkaar verschillen op een kenmerk. In deze bijlage wordt kort uitgelegd hoe deze effecten bepaald kunnen worden als gecorrigeerd wordt voor een ander kenmerk.

Stel we hebben twee groepen, A en B, die van elkaar verschillen wat betreft het kenmerk waarnaar we onderzoek doen. Voor deze twee groepen hebben we de vaardigheidsverdelingen bepaald. In dit voorbeeld is het gemiddelde van vaardigheid A gelijk aan 0,275, en dat van B gelijk aan 0,175. Beide groepen hebben een standaardafwijking van de vaardigheid gelijk aan 0,20. Met deze informatie is de effectgrootte van het verschil tussen A en B te bepalen: dat is het verschil tussen de twee gemiddelden gedeeld door de standaardafwijking<sup>11</sup>. De effectgrootte is hier dus  $[(0,275-0,175)/0,20=]$  0,50.

Nu kunnen we verder kijken naar de verdeling van de groepen A en B op een andere variabele. Hier nemen we opleiding, waar we twee verschillende niveaus aannemen: X en Y. Als we kijken naar het verschil tussen A en B per onderwijsniveau (zie tabel B5.4.1) dan zien we dat binnen hetzelfde onderwijsniveau de gemiddelden tussen A en B gelijk aan elkaar zijn. Binnen niveau X zijn de gemiddelden gelijk aan 0,30, en binnen niveau Y gelijk aan 0,15. Het verschil tussen de gemiddelden van A en B is per opleidingsniveau geheel verdwenen. Het totale verschil tussen A en B is het gevolg geweest van de verschillen in de verdeling naar opleiding voor groep A en B naar opleiding.

Tabel B5.4.1: Voorbeeld van verdelingen van groep A en B, totaal en naar opleiding

	Kandidaten				Vaardigheid				
	Aantal		% per groep		Gemiddelde		Standaardafwijking		
AV Opleiding	A	B	A	B	A	B	A	B	Pooled
<b>Totaal</b>	<b>720</b>	<b>90</b>	<b>100,0</b>	<b>100,0</b>	<b>0,275</b>	<b>0,175</b>	<b>0,20</b>	<b>0,20</b>	<b>0,20</b>
X	589	15	83,1	16,7	0,300	0,300	0,19	0,19	0,19
Y	120	75	16,9	83,3	0,150	0,150	0,19	0,19	0,19

Het gegeven voorbeeld is zeer extreem. Als we kijken naar GMA (geen migratieachtergrond) en NWMA (niet-westerse migratieachtergrond, en de werkelijke opleidingsniveaus dan ziet de tabel er B5.4.2 er anders uit.

<sup>11</sup> Om precies te zijn is dat de gepoolde standaardafwijking, maar die is – omdat de standaardafwijkingen in beide groepen hier exact gelijk zijn – gelijk aan de standaardafwijking zoals gevonden in beide groepen.

Tabel B5.4.2: Verdelingen van groep GMA en NWMA, totaal en naar opleiding

	Kandidaten				Vaardigheid				
	Aantal		% per groep		Gemiddelde		Standaardafwijking		
AV Opleiding	GMA	NWMA	GMA	NWMA	GMA	NWMA	GMA	NWMA	Pooled
<b>Totaal</b>	<b>703</b>	<b>90</b>	<b>100,0</b>	<b>100,0</b>	<b>0,17</b>	<b>0,06</b>	<b>0,22</b>	<b>0,26</b>	<b>0,22</b>
Havo EM/CM	129	13	18,3	14,4	0,20	0,13	0,20	0,15	0,19
Havo NT/NG	91	6	12,9	6,7	0,28	0,18	0,22	0,25	0,22
MBO Anders	101	18	14,4	20,0	0,20	0,17	0,24	0,33	0,25
MBO Z&W Onderwijsassistent	247	35	35,1	38,9	0,10	-0,04	0,19	0,21	0,19
MBO Z&W Pedagogische werk	37	7	5,3	7,8	0,21	-0,03	0,25	0,22	0,25
Anders	98	11	13,9	12,2	0,19	0,09	0,21	0,31	0,22

Op basis van tabel B5.4.2 kunnen de effectgroottes bepaald worden. Die staan in tabel B5.4.3. Daarin zien we dat de totale effectgrootte gelijk is aan 0,52. Kijken we per opleidingsniveau dan zie we wel duidelijke verschillen, met name bij de zorg & welzijn mbo-opleidingen tot onderwijsassistent en voor pedagogisch werk (relatief groot verschil tussen GMA en NWMA), maar juist een klein verschil bij de overige mbo-opleidingen (weinig impact van achtergrond). Er moet wel enige voorzichtigheid betracht worden bij het interpreteren van deze verschillen tussen de opleidingsniveaus vanwege de kleine aantallen kandidaten waar we het hier over hebben (neem bijvoorbeeld slechts 7 kandidaten met een NWMA bij pedagogisch werk). We kunnen een gemiddelde van de effecten bepalen, gewogen naar de frequentie van de opleiding. Dan komen we daarmee op een effectgrootte voor het verschil tussen GMA en NWMA dat vrijwel gelijk is aan het totaaleffect. De zogenaamde 'correctie' door opleiding mee te nemen heeft hier dus geen effect. Het verschil tussen de groepen blijft bestaan.

Tabel B5.4.3: Effectgroottes (GMA – NWMA), totaal en per onderwijsniveau

AV Opleiding	GMA - NWMA		GMA en NWMA	
	verschil	effect	N	%
<b>Totaal</b>	0,11	0,52	793	100,0
Havo EM/CM	0,07	0,38	142	17,9
Havo NT/NG	0,10	0,45	97	12,2
MBO Anders	0,03	0,11	119	15,0
MBO Z&W Onderwijsassistent	0,14	0,74	282	35,6
MBO Z&W Pedagogische werk	0,24	0,97	44	5,5
Anders	0,10	0,44	109	13,7
<b>Opleiding</b>		0,51	793	100,0

Op een vergelijkbare wijze zijn zo alle effecten bepaald zoals die gegeven zijn in tabel 5.5. Het is ook evident dat vanwege het lage aantal kandidaten met een westerse migratieachtergrond een opdeling daar minder zinvol is.

Wanneer een gemiddelde, noodzakelijk voor de berekening van de effectgrootte in een subgroep, bepaald is bij minder dan 5 personen, dan is deze effectgrootte niet gebruikt.

- Dat was het geval bij dyslexie (bij alle vaardigheden minder dan 5 kandidaten in de groep NWMA met een dyslexie verklaring).
- Bij de opleiding hoeven leerlingen met een havo CM/EM, die geschiedenis in hun eindexamen hadden, de geschiedenisstoets niet te doen, en de havo NT/NG hoefden de toets Natuur & Techniek niet te doen. Er waren enkele leerlingen die desalniettemin de toetsen gemaakt hadden, maar te weinig om mee te nemen in deze analyses.
- Bij taalgebruik lag het aantal kandidaten zonder migratieachtergrond dat vooral een andere taal met de ouders sprak onder de 5.
- Bij alle groepen was dat het geval voor kandidaten die vooral een andere taal met vrienden en bekenden spraken.
- Ook bij de groepen gebaseerd op de ervaring met het basisonderwijs, waarbij een opdeling gemaakt tussen geen ervaring en minder dan 8 jaar ervaring zijn de groepen zeer klein.

Om de variabele op basis van taal en ervaring met basisonderwijs toch mee te nemen voor de minder vaak voorkomende groepen zijn er ook berekeningen gemaakt waarbij categorieën samengevoegd waren. Dat waren bij de twee taalvragen de categorieën 'vooral anders' en 'even vaak Nederlands als anders' die nu naast 'vooral Nederlands' staan. Bij de ervaring met onderwijs is de tweedeling gemaakt tussen wel en niet 8 jaar ervaring. Bij de vergelijking van de resultaten bij de alternatieve opdeling en in het geval van de exclusies door de lage aantallen, zien we dat gemiddeld de effecten 0,02 kleiner zijn na de samenvoeging. Dit is een zeer beperkt effect.

Tabel B5.4.4: Effectgroottes (GMA – NWMA), taal per vakgebied

Effectgroottes GMA - NWMA	AK	GS	NT
<b>Totaal</b>	<b>0,52</b>	<b>0,31</b>	<b>0,87</b>
Taal met Ouders	0,23	0,34	0,75
Taal met Ouders alternatief	0,18	0,30	0,75
Taal met Bekenden	0,47	0,31	0,86
Taal met Bekenden alternatief	0,44	0,27	0,83
Hoe lang BO	0,48	0,28	0,85
Hoe lang BO alternatief	0,47	0,28	0,85

# Bijlage 10 Eisen vaststelling

1. De vaststellingscommissie stelt items vast aan de hand van de volgende criteria:
  - Het item past binnen de kennisdoelen uit de handreiking.
    - AK: 8 kernthema's
    - GS: 10 tijdvakken
    - NT: 10 kernconcepten;
2. Het item past binnen het vastgestelde niveau een niveau van complexiteit dat past bij de eerste drie leerjaren havo en/of de bovenbouw van het vmbo-t.
  - niveau vergelijkbaar met havo-3: AK en NT
  - niveau vergelijkbaar met vmbo-t4: GS
3. Het item is inhoudelijk correct
  - De inhoud van de vraag klopt
  - De sleutel klopt
  - De alternatieven zijn aantrekkelijk

Als het item voldoet aan alle drie de criteria dan wordt het item vastgesteld.

Als het item niet voldoet aan criteria 2 en/of 3 dan is er aanpassing van het item mogelijk.

# Bijlage 11 Een aantal aandachtspunten voor toetsconstructeurs

De vier aandachtspunten die hierna worden gepresenteerd, zijn gebaseerd op de resultaten van de inhoudelijke analyse van 81 opgaven. Daarvan waren er 65 in meer of minder mate opgevallen in een van de DIF-analyses die in het kader van dit onderzoek zijn uitgevoerd. De lijst pretendeert dan ook niet meer te zijn dan een aanvulling bij bestaande hulpmiddelen en richtlijnen voor toetsconstructeurs. De aandachtspunten die hier volgen, zijn specifiek gericht op degenen die belast zijn met de ontwikkeling van opgaven voor de toelatingstoetsen voor de pabo. Zij vinden in deze lijst een aantal 'criteria' die zij náást uitgebreide aanwijzingen zoals die van Bügel & Sanders (1998), kunnen gebruiken bij het ontwikkelen en evalueren van opgaven voor de Toelatingstoetsen voor de pabo.

Deze aandachtspunten zijn niet specifiek gericht op het voorkomen van culturele bias in opgaven voor aspirant-pabostudenten met een niet-westerse achtergrond. Het onderzoek waar deze lijst uit voortkomt, heeft géén gegevens opgeleverd die aanleiding geven om te vermoeden dat de toetsconstructeurs daar meer aandacht aan zouden moeten besteden.

De aandachtspunten zijn geformuleerd in termen van een aantal vragen die gesteld kunnen worden bij (concepten van) opgaven van de Toelatingstoetsen en zijn steeds voorzien van een korte toelichting. Uitgangspunt is steeds dat de opgaven zó geformuleerd dienen te zijn dat alle kandidaten optimale gelegenheid krijgen om te laten zien dat ze de getoetste kennis beheersen.

## Aandachtspunten

1. *Kan de doelgroep van de Handreiking bij deze toets, aspirant-pabostudenten die zich voorbereiden op de toelatingstoets, de opgave herkennen als een operationalisering van leerdoelen die beschreven zijn in de Handreiking?*

Kies bij de interpretatie van de inhoud van de Handreiking voor het perspectief van de primaire doelgroep van die Handreiking: de aspirant-student die zich wil voorbereiden op de Toelatingstoets. Wees je bewust van het feit dat jij als expert op het betrokken vakgebied, in die Handreiking 'tussen de regels' méér zult lezen dan de aspirant-student die nauwelijks of geen voorkennis heeft op het vakgebied waar de Handreiking betrekking op heeft. Om een valide opgave te maken, dien je uit te gaan van de letterlijke tekst van de Handreiking. Informatie en associaties die niet expliciet in de tekst van de Handreiking zijn terug te vinden, horen niet bevroegd te worden in de Toelatingstoetsen.

2. *Kan de doelgroep van de Toelatingstoetsen de context waarin de opgave wordt gepositioneerd, in verband brengen met de inhoud van de Handreiking?*

Houd bij het kiezen van contexten voor toetsvragen rekening met het beheersingsniveau van de stof dat je van de aspirant-studenten mag verwachten (havo 3, vmbo tl) en met verschillen in de referentiekaders tussen jou en de doelgroep en tussen de leden van de doelgroep onderling. Ga niet te creatief om met contexten en bedenk dat kandidaten geneigd zullen zijn alle informatie in de context te interpreteren als betekenisvol. Realiseer je dat een context misleidend kan zijn doordat een kandidaat die in een examensituatie verkeert, élk detail van de opgave zal interpreteren als relevant. Een vraag naar peristaltische bewegingen van de darm die wordt geplaatst in de context van een persoon die eet terwijl hij op zijn handen staat, is onnodig ingewikkeld.

3. *Zijn de afbeeldingen die in de opgave zijn verwerkt, functioneel en passend?*

Gebruik geen afbeeldingen die niet bijdragen aan het op de juiste manier interpreteren van de opgave en houd rekening met het feit dat kandidaten afbeeldingen verschillend kunnen interpreteren. Realiseer je daarbij dat de meeste kandidaten hebben geleerd om elke afbeelding als betekenisvol te behandelen. Geef bijvoorbeeld geen afbeelding van bevrijdingsfeesten in mei 1945 bij een opgave over slachtoffers die vielen in het laatste oorlogsjaar.

Vermijd ook dat er misleidende onderschriften blijven staan bij afbeeldingen (zoals 'Willibrord. Apostel der Friezen') en gebruik alleen kaartmateriaal dat voldoet aan alle eisen die in de Handleiding Aardrijkskunde worden beschreven.

4. *Heeft de opgave kenmerken van een 'puzzel'?*

Het interpreteren en vergelijken van antwoordmogelijkheden die méér dan een regel beslaan, is voor veel kandidaten verwarrend. Hetzelfde geldt voor vragen die een ontkenning bevatten ("Wat is géén voorbeeld van..."). Ook woordgebruik kan een opgave onbedoeld moeilijk maken. Kies wanneer beheersing van specifieke vaktaal geen specifiek toetsdoel is, voor gangbaar taal- en woordgebruik. Infrequent woordgebruik ('delven van grondstoffen' in plaats van 'winnen van aardgas') kan leiden tot ongewenste effecten.





