

ontwikkeling van een model voor domeinbeschrijving van taalonderwijs. Dissertatie. Arnhem, Cito, 1991.

Sijtsma, J., *Taal voor iedereen. Raamplan voor een geïntegreerde NT1/NT2-methode voor de basisschool*. 's Hertogenbosch, Katholiek Pedagogisch Centrum, 1994.

Vegt, N. van der, *Op vooroordelen bekeken. Het toetsen van leermiddelen voor intercultureel onderwijs*. Enschede, SLO, 1987.

Veenman, S., P. Lem, E. Roelofs & F. Nijssen (red.), *Effectieve instructie en doelmatig klasmanagement. Een schoolverbeteringsprogramma voor enkelvoudige en combinatieklassen*. Amsterdam/Lisse, Swets & Zeitlinger, 1992.

Veenman, S. (red.), *Effectieve Instructie. Leren onderwijzen met behulp van het Directe-Instructie-model*. Hoevelaken, Christelijk Pedagogisch Studiecentrum, 1993.

Verhallen, M. & S. Verhallen, *Woorden leren woorden onderwijzen. Handreiking voor leraren in het basis- en voortgezet onderwijs*. Hoevelaken, CPS, 1994.

Henny Uiterwijk & Ton Vallen

Hoe worden toetsen minder partijdig voor allochtonen?

De schoolresultaten van allochtone leerlingen in het basisonderwijs en hun doorstroming naar MAVO/HAVO/VWO ontwikkelen zich langzaam maar zeker in positieve zin. Dat neemt niet weg dat hun onderwijspositie nog steeds beduidend ongunstiger is dan die van hun autochtone leeftijdgenoten. Wetenschappelijk onderzoek laat zien dat er voor de minder voorspoedig verlopende schoolloopbanen van allochtone leerlingen vele en zeer uiteenlopende oorzaken zijn aan te wijzen. Vooral de volgende, veelal niet onafhankelijk van elkaar opererende factoren komen daarbij veelvuldig naar voren: migratiegeschiedenis, sociaal-economische, sociaal-culturele en talige kenmerken van het gezin, kenmerken van de processen van leren en instructie op school, en kenmerken van de keuzeprocessen die zich op cruciale momenten in de schoolloopbaan afspelen. Aan de laatstgenoemde factor, en dan met name aan de vraag of de toetsen die ter ondersteuning bij de te maken keuzes worden gebruikt, allochtone leerlingen onbedoeld bevoordelen dan wel benadelen, wordt slechts weinig aandacht geschonken. In dit artikel staat die vraag echter centraal. Een eerste aanzet tot beantwoording ervan gebeurt aan de hand van de resultaten van een door de auteurs uitgevoerd onderzoek naar de bruikbaarheid van de CITO Eindtoets Basisonderwijs voor allochtone leerlingen.

Onderzoeksopzet

Het onderzoek dat het Werkverband Taal en Minderheden van de Letterenfaculteit van de KUB en het CITO in verband met de Eindtoets Basisonderwijs gezamenlijk uitvoeren, bestaat uit twee onderdelen. Het eerste heeft betrek-

king op de Eindtoets als geheel, het tweede op de afzonderlijke toetsopgaven. In het eerste onderdeel gaat het om de *schoolresultaten* van allochtone en autochtone leerlingen. Met schoolresultaten worden de toetsscores op (onderdelen van) de Eindtoets Basisonderwijs 1987 en 1989 bedoeld en de schoolloopbaangegevens van deze leerlingen wat betreft hun toelating tot en doorstroming in de beginfase van het voortgezet onderwijs. Tegen die achtergrond wordt ook nagegaan in welke mate de Eindtoets Basisonderwijs het schoolsucces in het voortgezet onderwijs kan voorspellen in vergelijking met het advies van de basisschool. Dit onderzoeksgedeelte heeft dus betrekking op een vergelijking van de trefzekerheid waarmee door toets en advies het schoolsucces van autochtone en allochtone leerlingen in het voortgezet onderwijs wordt ingeschat.

Het tweede onderdeel gaat over *itembias* of *vraagpartijdigheid*. Bij itembias gaat het om scoreverschillen tussen beide groepen leerlingen die niet veroorzaakt worden door de uiteenlopende mate van beheersing van de vaardigheden die toetsopgaven beogen te meten, maar door verschillen die toetsopgaven niet beogen te meten. Wanneer voor het juist beantwoorden van de items nog andere vaardigheden nodig zijn dan uitsluitend de vaardigheden die de opgaven beogen te meten, kan afbreuk gedaan worden aan de validiteit van het meetinstrument: de toets meet immers meer dan ze pretendeert te meten. De validiteit is in het geding wanneer die benodigde additionele vaardigheden niet bij alle autochtone en allochtone leerlingen in vergelijkbare mate aanwezig zijn. Er kan bijvoorbeeld sprake zijn van itembias wanneer het niet direct tot het rekenjargon behorende taalgebruik in een rekenitem voor bijvoorbeeld allochtone leerlingen dermate ingewikkeld is, dat ze daardoor niet of in onvoldoende mate aan het uitvoeren van de feitelijk beoogde rekenoperatie toekomen of daaraan onvoldoende aandacht kunnen besteden. Allochtone leerlingen die over dezelfde rekenvaardigheid beschikken als autochtone hebben dan ten onrechte en onbedoeld een geringere kans op een goed antwoord.

Bij bovenstaand voorbeeld is er sprake van itembias ten nadele van allochtone leerlingen,

maar het kan natuurlijk ook voorkomen dat een toets opgaven bevat die partijdig zijn ten voordele van deze subgroep van leerlingen. Dat is het geval wanneer voor het correct kunnen oplossen van een item additionele vaardigheden vereist zijn waarover allochtone leerlingen in grotere mate beschikken dan gelijkpresterende autochtone.

Het onderzoek naar itembias is in verschillende stappen uitgevoerd. In een eerste fase zijn met verschillende statistische technieken die toetsopgaven opgespoord waarbij autochtone en allochtone leerlingen die over dezelfde relevante vaardigheid beschikken toch een ongelijke kans hebben om de betreffende opgave goed te beantwoorden. In de tweede fase van dit onderzoeksonderdeel is een poging ondernomen om te na te gaan wat bij een bepaalde opgave de inhoudelijke oorzaak van itembias zou kunnen zijn. Een fundamenteel probleem daarbij was dat niet alleen in Nederland maar ook daarbuiten bijzonder weinig onderzoek naar inhoudelijke oorzaken van itembias voor allochtone leerlingen is gedaan. In feite moest vanaf nul worden begonnen en tentatief te werk worden gegaan. Bijgevolg kunnen de resultaten van het onderhavige project dan ook slechts als voorlopig worden beschouwd.

Om bij de inhoudsanalyses van de statistisch partijdige items zoveel mogelijk gericht te kunnen kijken, is allereerst op basis van een uitgebreide literatuurstudie een inventarisatie gemaakt van de problemen die allochtone basisschoolverlaters met het Nederlands (als tweede taal) ondervinden. Ook zijn de problemen geïnventariseerd waarmee ze mogelijk geconfronteerd worden op grond van het feit dat hun culturele achtergrond op z'n minst voor een deel verschillend is van de autochtone *mainstream* in Nederland. Het gaat hierbij dus om uit onderzoek naar voren gekomen verschillen in moeilijkheidsgraad voor autochtonen en allochtonen op een aantal (deel)vaardigheden en niet direct om itembias. Deze verschillen in moeilijkheidsgraad (wat de Nederlandse taal betreft onder andere op het terrein van de woordenschat en de zins- en tekstcomplexiteit) zijn in het kader van ons onderzoek echter toch van belang, omdat ze te beschouwen zijn als hypotheses omtrent

potentiële linguïstische en culturele bronnen van itembias. Elders (met name in De Jong & Vallen 1989 en Uiterwijk 1994) is uitgebreid verslag gedaan over dit deel van het onderzoek.

Na afronding van het literatuuronderzoek en de eerste inhoudelijke interpretaties door de projectmedewerkers zijn voor het achterhalen van mogelijke inhoudelijke oorzaken van itembias externe experts en leerlingen uit groep 8 van het basisonderwijs ingeschakeld. De experts is gevraagd om de items die in statistische zin partijdig zijn in het voor- en nadeel van allochtone leerlingen, op biasbronnen te beoordelen. Daarnaast is door middel van een kleinschalig hardop-denken-experiment, waarin allochtone en autochtone leerlingen uit groep 8 van het basisonderwijs zo uitgebreid en nauwkeurig mogelijk moesten vertellen hoe ze de hen voorlegde items aangepakt hadden, getracht in de redeneringen van de leerlingen verdere aanwijzingen te vinden omtrent bronnen van itembias. Een en ander gebeurde vanuit het idee dat uit het totale onderzoek wellicht zou kunnen blijken met welke aanpassingen de bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen vergroot zou kunnen worden.

Schoolresultaten allochtone leerlingen

In het onderstaande wordt eerst gerapporteerd over de toetsscores van allochtone leerlingen op (onderdelen van) de Eindtoets Basisonderwijs en vervolgens over de gegevens van deze leerlingen omtrent hun toelating tot en doorstroming in het voortgezet onderwijs.

De gemiddelde CITO-totaalscore van de onderscheiden etnische groepen op de Eindtoets Basisonderwijs 1987 en 1989 verschillen aanzienlijk. De leerlingen van Marokkaanse en Turkse herkomst behalen in beide jaren de laagste gemiddelde totaalscore, gevolgd door de leerlingen van Surinaamse, Antilliaanse en Molukse herkomst. De Chinese leerlingen behalen vergeleken met alle andere etnische groepen (inclusief de autochtone leerlingen) de hoogste gemiddelde rekenscore. Dit wordt bevestigd door ander onderzoek in Nederland en de Verenigde Staten (Dossey e.a. 1988, Driessen 1990). De verschillen tussen de gemiddelde scores van autochtone en alloch-

tone leerlingen zijn bij de spellingopgaven, met name bij de spelling van werkwoordsvormen, relatief klein.

Wanneer we bij de toelating tot het voortgezet onderwijs naar de instroom van leerlingen met een vergelijkbaar prestatieniveau kijken, dan blijkt dat er meer autochtone dan allochtone leerlingen naar schooltypen met een lager gemiddeld prestatieniveau (IVBO, VBO en VBO/AVO) gaan. Allochtone leerlingen hebben over het algemeen de overhand in het AVO in vergelijking met vergelijkbaar presterende autochtone leerlingen. Deze bevindingen zijn door andere onderzoekers herhaaldelijk bevestigd: allochtone leerlingen worden tot een hoger schooltype toegelaten dan op grond van toets- of testscores verwacht zou kunnen worden (Driessen 1991, Mulder & Tesser 1991, Van Langen & Jungbluth 1992, Meijnen & Riemersma, 1992).

De relatieve voorsprong die allochtone leerlingen bij de start in het voortgezet onderwijs ten opzichte van autochtone leerlingen hebben, wordt volgens de gegevens uit het onderhavige onderzoek voor een deel weer teniet gedaan aan het einde van het eerste leerjaar. Uit de doorstroomgegevens van leerlingen van een vergelijkbaar prestatieniveau blijkt dat allochtone leerlingen meer dan hun autochtone klasgenoten afstromen (blijven zitten of doorstromen naar een onderwijstype met een 'lager' gemiddeld prestatieniveau). Ook blijkt dat allochtone leerlingen die toegelaten zijn tot het VBO en het MAVO iets meer dan autochtone leerlingen doorstromen naar onderwijstypen met een hoger gemiddeld prestatieniveau. Er zijn dus ook allochtone opstomers. Het meer algemene beeld blijft echter toch dat allochtone leerlingen vanuit het eerste leerjaar van een scholengemeenschap in vergelijking met autochtone naar de schooltypen met een lager gemiddeld prestatieniveau gaan.

De voorspellende waarde

Tegen de achtergrond van bovenstaande gegevens is ook onderzocht hoe hoog de voorspellende waarde van de Eindtoets Basisonderwijs voor allochtone en autochtone leerlingen is in vergelijking met die van het advies van de basisschool. Hierbij is de samenhang (correla-

tie) berekend tussen de CITO-score respectievelijk het advies basisschool en de posities die de leerlingen na één jaar in het voortgezet onderwijs innemen (Uiterwijk, 1994). Op grond van de onderzoeksgegevens kan gesteld worden dat de voorspellende waarde van zowel de CITO-score als het advies basisschool hoog is. De voorspellende waarde van de Eindtoets voor allochtone leerlingen kan, uitgedrukt in een schoolcijfer, met een 7,5 gewaardeerd worden, voor autochtone met een 8. De voorspellende waarde van het schoolkeuze-advies van de basisschool voor allochtone leerlingen kan met het cijfer 8 gewaardeerd worden, voor autochtone met een 8,5. Hieruit blijkt dat de voorspellende waarde van zowel de Eindtoets als het advies basisschool bij allochtone leerlingen iets lager is dan bij autochtone. Het voorspellen van schoolsucces voor allochtone leerlingen gaat dus minder trefzeker dan voor autochtone. Het advies basisschool voorspelt het schoolsucces beter dan de CITO-score, maar de verschillen zijn klein. Verder is gebleken dat bij autochtone leerlingen de voorspellende waarde van het toetsonderdeel 'taal' het grootst is, bij allochtone zijn zowel de taal- als de rekenscores van belang.

Het is moeilijk om precies aan te geven waarom het schoolsucces van allochtone leerlingen in het voortgezet onderwijs minder trefzeker is te voorspellen dan dat van autochtone leerlingen. Op de schoolkeuze en op de toelating tot en doorstroming in het voortgezet onderwijs zijn immers veel factoren van invloed. Onderzoek geeft wel aan dat het risico op afstromen het grootst is als bij de schoolkeuze en toelating weinig rekening wordt gehouden met het advies van de basisschool en met de behaalde toetsscore (De Wit, Suhre & Mulder 1993, Uiterwijk 1994).

Over het algemeen kan gesteld worden dat meer aandacht voor de schoolkeuze, toelating en doorstroming van allochtone leerlingen geboden is. Om de schoolloopbanen van allochtone leerlingen in de beginfase van het voortgezet onderwijs te verbeteren is het in de eerste plaats gewenst om de geldende toelatingsregelingen te evalueren. Hierbij kan met empirische gegevens nagegaan worden of een toelatingsregeling in het schoolkeuzeproces van allochtone en allochtone leerlingen op

vergelijkbare wijze functioneert. Aan de hand van de afstroomgegevens kan ingeschat worden of de toelatingsregeling voor zowel allochtone als autochtone leerlingen niet te streng of te soepel is. Verder is het van groot belang dat het voortgezet onderwijs zijn verantwoordelijkheid ten opzichte van de toegelaten leerlingen effectueert. Gezien de relatief grote afstroom van allochtone leerlingen dient het voortgezet onderwijs na te gaan of het onderwijsaanbod niet beter afgestemd kan worden op de mogelijkheden van allochtone leerlingen.

Resultaten van het itembias-onderzoek

Zoals eerder opgemerkt kunnen in het onderzoek naar vraagpartijdigheid twee elkaar aanvullende fasen worden onderscheiden. In de eerste fase zijn met behulp van verschillende statistische procedures vraagpartijdige items opgespoord, in de tweede fase is nagegaan wat bij statistisch gebiaste items de inhoudelijke oorzaak daarvan zou kunnen zijn.

Zonder hier nader in te gaan op voors en tegens van de verschillende typen statistische procedures (zie daarvoor Uiterwijk 1994) kan opgemerkt worden dat van meerdere statistische technieken gebruik werd gemaakt, omdat uit eerder onderzoek was gebleken dat diverse relevante procedures niet tot identieke resultaten leiden.

De resultaten van de statistische analyses maken duidelijk dat niet exact is aan te geven hoeveel opgaven van de Eindtoets 1987 en 1989 (in totaal 360 items) partijdig zijn. Uit de analyses met de ene procedure blijkt dat 20 van de in totaal 360 geanalyseerde opgaven (6 procent) partijdig zijn voor Turkse en/of Marokkaanse leerlingen. De analyses met de andere procedure leveren een aantal van 45 partijdige items op (13 procent) voor Turken en/of Marokkanen. In totaal zijn er 13 items (4 procent) partijdig met zowel de ene als de andere procedure, waarvan er drie partijdig zijn in het voordeel van één of beide etnische groepen en tien in het nadeel. In concluderende zin kan worden opgemerkt dat uitspraken over de partijdigheid van items gradaties moeten kennen: sommige items zijn partijdig bij alle analyses, andere bij één analyse en de meeste bij geen enkele analyse.

Bij alle statistisch gebiaste items is vervolgens nagegaan wat de inhoudelijke oorzaak van de partijdigheid zou kunnen zijn. De betreffende opgaven zijn, rekening houdend met de beoogde te meten vaardigheid, eerst door drie projectmedewerkers onafhankelijk van elkaar inhoudelijk in detail geanalyseerd op mogelijke bronnen van partijdigheid in de diverse item-onderdelen (introductie, uitleg, eventuele afbeelding, vraag, afleiders). Hierbij is vooral aandacht besteed aan die elementen die voorkomen in items die partijdig zijn in het nadeel van Turkse en Marokkaanse leerlingen en ontbreken in opgaven die in het voordeel zijn van deze leerlingen en omgekeerd. Bij deze inhoudsanalyse hebben de eerder op basis van de uitgevoerde literatuurstudie geformuleerde potentiële bronnen van item-bias voor allochtone leerlingen als hypotheses gefungeerd (De Jong & Vallen 1989). Vervolgens zijn de gemeenschappelijkheden in de analyses van de afzonderlijke projectmedewerkers in kaart gebracht.

Vervolgens is aan zestien externe experts (elf autochtone en vijf allochtone) gevraagd om partijdige items in het voor- en nadeel van allochtone leerlingen op biasbronnen te beoordelen (Van de Waal 1992). Tussen de oordelen van de experts over de itemelementen die moeilijk zijn voor allochtone leerlingen blijkt niet alleen een hoge mate van onderlinge overeenstemming te zijn, maar bovendien sluiten deze oordelen over het algemeen ook sterk aan bij de bevindingen van de projectmedewerkers.

Het hardop-denken-experiment

In een volgende stap van de inhoudelijke bias-analyse werd door middel van een kleinschalig hardop-denken-experiment onderzocht hoe vaak allochtone en autochtone leerlingen partijdige items fout beantwoorden tengevolge van een element in het item dat waarschijnlijk de itembias veroorzaakt. Verder is nagegaan hoe vaak bij *gemanipuleerde* items ten gevolge van de itemmanipulatie een goed antwoord wordt gegeven. Gemanipuleerde items zijn opgaven waarbij het itemelement dat vermoedelijk de biasbron vormt (bijvoorbeeld: 'Hoeveel moet hij betalen *inclusief* B.T.W.?'), is vervangen door een itemelement waarvan ver-

wacht wordt dat het geen bias veroorzaakt ('Hoeveel moet hij betalen *met* B.T.W.?). De allochtone en autochtone leerlingen moesten eerst de hen voorgelegde oorspronkelijke, respectievelijk gemanipuleerde items goed bestuderen, daarna konden ze het naar hun oordeel goede antwoord aankruisen. Tot slot moesten ze zo uitgebreid en nauwkeurig mogelijk aangeven hoe ze aan hun antwoord gekomen waren. In de redeneringen van de leerlingen is getracht aanwijzingen te vinden omtrent inhoudelijke bronnen van itembias (Coenen & Vallen 1991).

In het hardop-denken-experiment werden zeventien opgaven uit de Eindtoets Basisonderwijs 1987 (vijf taalitems, acht rekenitems en vier opgaven voor informatieverwerking) die volgens de statistische analyses sterk ten nadele van de Turkse en/of Marokkaanse leerlingen gebiast waren, in hun oorspronkelijke vorm (versie A) en in een talig gemanipuleerde vorm (versie B) opgenomen. Bij de keuze van de opgaven speelde mede een rol dat de vaardigheid die ze in hun oorspronkelijke vorm pretenderen te meten door de (minimale) manipulaties niet zou worden aangetast. De items werden ter oplossing en bespreking voorgelegd aan 44 leerlingen uit groep 8 van de basisschool: 22 paren van telkens qua niveau vergelijkbare autochtone en allochtone (Turkse en Marokkaanse) leerlingen. Aan de ene helft van de paren werd versie A voorgelegd, aan de andere helft de gemanipuleerde versie B.

De resultaten van het experiment laten zien dat de *allochtone* leerlingen die de gemanipuleerde versie B maakten, in totaal gemiddeld 15,5 procent meer items goed maken dan de (volgens de leerkrachten) even vaardige allochtone leerlingen die de oorspronkelijke items (versie A) maakten. Bij de qua vaardigheidsniveau vergelijkbare *autochtone* leerlingen bedraagt het scoreverschil tussen beide versies slechts 3,7 procent ten voordele van de B-versie.

Het verschil in het gemiddeld percentage goede antwoorden tussen de allochtone en autochtone leerlingen die toetsversie A maakten is 22,5 procent ten voordele van laatstgenoemden. Bij de gemanipuleerde versie B is het gemiddelde verschil in percentage goede antwoorden tussen beide groepen meer dan

gehalveerd en bedraagt het nog maar 10,7 procent.

Deze bovenstaande gegevens laten zien dat de allochtone leerlingen grosso modo meer van de itemmanipulaties geprofiteerd hebben dan de autochtone. Voor de drie toetsonderdelen (Taal, Rekenen en Informatieverwerking) lag dat trouwens verschillend. Bij Informatieverwerking maakten de allochtone leerlingen alle gemanipuleerde opgaven beter dan de oorspronkelijke (wat een toename in het gemiddeld aantal goede antwoorden van ruim 20 procent oplevert), terwijl de score voor de autochtonen hetzelfde bleef. De toename in de correctscore bij de gemanipuleerde rekenitems bedraagt voor de allochtonen ruim 17 procent, terwijl de correctscore voor de autochtonen slechts zo'n 2 procent stijgt. Zowel op de ongemanipuleerde als op de gemanipuleerde opgaven voor Rekenen en Informatieverwerking blijven de autochtone leerlingen wel beter scoren dan de allochtone, maar het verschil tussen beide subgroepen is bij de gemanipuleerde items veel kleiner geworden: bij Rekenen is het verschil gedaald van 19,3 naar 4,6 procent en bij Informatieverwerking van 31,8 naar 11,4 procent.

De itemmanipulaties bij het onderdeel Taal waren nauwelijks succesvol. Niet alleen was er sprake van een veel geringere toename in de correctscore, maar bovendien was die toename van 9 procent voor allochtonen en autochtonen precies gelijk. Dit zou erop kunnen wijzen dat met de talige manipulatie van de taalitems niet de mogelijke biasbron geëlimineerd is, maar dat de moeilijkheidsgraad van de items veranderd is, waardoor ze voor iedereen gemakkelijker zijn geworden. Manipulatie van taalopgaven lijkt nauwelijks mogelijk zonder de moeilijkheidsgraad van die opgaven te veranderen of wezenlijk aan te tasten. Bij itembias van taalopgaven zijn wellicht andere biasbronnen in het geding zijn dan talige.

Over de vraag of de veronderstelde inhoudelijke biasbronnen in de 17 bij het hardop-denken-experiment betrokken opgaven ook daadwerkelijk voor allochtonen een rol speelden, moesten de geregistreerde leerlinguitspraken en de op basis daarvan gemaakte protocolanalyses uitsluitsel geven. Dat was geen eenvoudige opgave, mede gezien de diversiteit

Oorspronkelijke versie (1987; Rekenen 57)

Vader koopt een naaimachine. Deze kost f 800,- zonder b.t.w. De b.t.w. is 20%. Hoeveel moet vader betalen inclusief b.t.w.?

- a f 160,-
- b f 640,-
- c f 820,-
- d f 960,-

Gemanipuleerde versie

Vader koopt een naaimachine. Deze kost f 800,- zonder b.t.w. De b.t.w. is 20%. Wat moet vader voor de naaimachine betalen met b.t.w.?

- a f 160,-
- b f 640,-
- c f 820,-
- d f 960,-

Afbeelding 1: Een gemanipuleerde rekenopgave

en complexiteit van de opgaven. Om een zo duidelijk mogelijke indruk te geven van de procedure worden in het onderstaande bij wijze van voorbeeld twee rekenopgaven en een opgave voor Informatieverwerking behandeld. De doelstelling van die drie statistisch sterk in het nadeel van allochtone leerlingen gebiaste items heeft primair met de meting van rekenen respectievelijk informatieverwerking van doen en met de meting van taalvaardigheid Nederlands. De kwestie van interferentie van taalbias en beoogde moeilijkheid behoort daarom bij deze items dus niet aan de orde te zijn. De items worden telkens eerst in hun oorspronkelijke en daarna in hun gemanipuleerde vorm gepresenteerd. Het gemanipuleerde deel van het item is steeds gecursiveerd.

De talige manipulatie bestaat met name hierin dat de vraag anders geformuleerd is (onder andere een *wat*-vraag in plaats van een *hoe*-vraag) en dat het woord *inclusief* vervangen is door het equivalente alternatief *met*.

De oorspronkelijke versie van rekenopgave 57 (afbeelding 1) wordt door zes van de elf allochtone leerlingen fout beantwoord en door twee van de elf autochtone. Bij de gemanipu-

Oorspronkelijke versie (1987; Rekenen 52)

Per jaar gaf een gezin gemiddeld f 1500,- uit aan aardappels en groenten. Ze wilden bezuinigen. Ze huurden een jaar een tuin van 200 vierkante meter voor f 2,- per vierkante meter. De overige onkosten waren f 80,-. Ze moesten nog voor een bedrag van f 400,- aan aardappelen en groenten in de winkel kopen. De rest kwam uit de tuin. Hoeveel had dat gezin bespaard met tuinieren?

- a f 620,-
- b f 820,-
- c f 920,-
- d f 1020,-

Gemanipuleerde versie

Per jaar koopt mevrouw Knol voor f 1500,- groenten. Dat vindt ze teveel geld. Daarom huurt ze voor een jaar een groentetuin van 200 vierkante meter. Die tuin kost f 2,- per vierkante meter. De spullen die ze voor de tuin nodig heeft kosten nog f 80,-. Omdat de tuin veel groenten oplevert hoeft mevrouw Knol nog maar voor f 400,- aan groenten in de winkel te kopen. Hoeveel geld heeft mevrouw Knol met haar groentetuin bespaard?

- a f 620,-
- b f 820,-
- c f 920,-
- d f 1020,-

Afbeelding 2: Een gemanipuleerde rekenopgave

leerde versie geven nog twee allochtone en één autochtone leerling een fout antwoord. Uit de protocol-analyses blijkt dat bij drie allochtone en bij één autochtone leerling het woord *inclusief* duidelijk de oorzaak van de fout is. Een andere allochtone leerling heeft ook problemen met *inclusief*, maar ook met het berekenen van procenten. Als de aanduiding *met* in de oorspronkelijke opgave zou zijn gebruikt, had dat (naar het eigen oordeel van de bevraagde leerlingen) voor twee allochtone leerlingen geen verschil uitgemaakt, maar vier allochtone en twee autochtone leer-

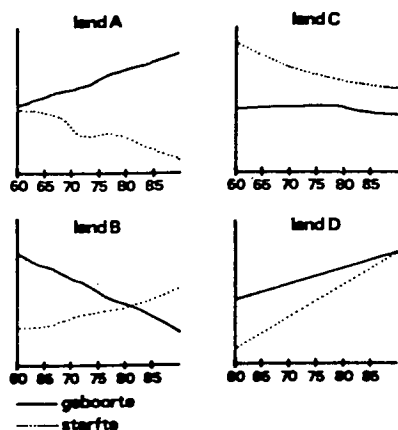
lingen geven te kennen dat de opgave daarvoor voor hen begrijpelijker zou zijn geweest.

De leerlingen moeten bij rekenopgave 52 (afbeelding 2) uit de tekst opmaken welke rekenoperaties ze achtereenvolgens moeten uitvoeren. Door het relatief grote aantal getallen dat in een omvangrijke talige context is ingebed en gebruikt moet worden, doet deze rekenopgave een groot beroep op taalvaardigheid Nederlands. Het is dan ook niet zo verwonderlijk dat de meerderheid van de allochtone leerlingen grote problemen met de oorspronkelijke opgave heeft. Daarom is in de gemanipuleerde versie (die met name door de allochtonen beduidend beter werd gemaakt) met meer expliciete en duidelijker op elkaar aansluitende formuleringen gewerkt. Daarnaast heeft personificatie plaatsgevonden (*mevrouw Knol* in plaats van de verzamelnaam *gezin*). Verder bevat de oorspronkelijke opgave niet alleen veel verwijzingen maar ook een aantal complexe verwijzingen (bijvoorbeeld het verwijswoord *ze* in de tweede zin, dat getalsmatig afwijkt van zijn antecedent in de eerste zin: *een gezin*). Uit de literatuur is het bekend dat de verwijzingsystematiek van het Nederlands voor allochtonen een specifiek probleem vormt dat ze pas in een vrij laat T2-verwervingsstadium onder de knie krijgen.

In de gemanipuleerde versie van informatie-opdracht 40 (afbeelding 3) is de vraagstelling meer expliciet gemaakt en zijn *toenemen* en *afnemen* vervangen door respectievelijk *stijgen* en *dalen*. De oorspronkelijke versie wordt door vijf van de elf allochtone leerlingen fout beantwoord en door geen enkele autochtone leerling. Uit de protocol-analyses blijkt dat twee allochtone leerlingen de legenda niet meteen opmerken, omdat die erg dicht op de grafieken staat. Vier allochtone leerlingen hebben problemen met de woorden *toenemen* en *afnemen*. Een van deze leerlingen, die aan *afnemen* de betekenis *iemand zijn spullen afpakken* toekent, komt via de redenering dat de drie andere grafieken 'raar' zijn toch uiteindelijk op het goede antwoord uit. Alle vier de allochtonen die problemen hebben met de betekenis van de oorspronkelijke combinatie van werkwoorden, begrepen het begrippenpaar *stijgen/dalen* wel. Bij de gemanipuleerde versie

Oorspronkelijke versie (1987; Informatieverwerking 40)

Sterfte- en geboortecijfers in vier landen

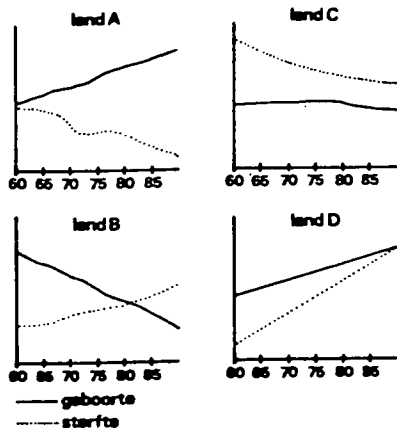


Van welk land kan men zeggen dat het aantal geboorten toeneemt en het aantal sterfgevallen afneemt?

- a van land a
- b van land b
- c van land c
- d van land d

Gemanipuleerde versie

Sterfte- en geboortecijfers in vier landen



In welk land is tussen 1960 en 1985 het aantal geboorten gestegen en het aantal sterfgevallen gedaald?

- a van land a
- b van land b
- c van land c
- d van land d

Afbeelding 3: Een gemanipuleerde opgave voor Informatieverwerking

gaven twee allochtone leerlingen en één autochtone een fout antwoord.

Met inachtneming van de beperkingen van dit (tentatieve) onderzoeksonderdeel kan gezegd worden dat de hardop-denken-procedure de meeste van de op basis van de andere activiteiten geformuleerde voorlopige bronnen van itembias heeft bevestigd en bovendien enkele nieuwe opties heeft opgeleverd (voor meer informatie zie Coenen & Vallen 1991 en Uiterwijk 1994).

Inhoudelijke bronnen van itembias

Op basis van de besproken onderzoeksactiviteiten kan allereerst worden vastgesteld dat onderzoek naar itembias nog met veel onzekerheden is omgeven. Dat neemt niet weg dat

het onderzoek zeker inzichten heeft opgeleverd over de vraag in welke richting verder gewerkt zou moeten worden en enkele plausibele bronnen van talige itembias voor Turkse en Marokkaanse leerlingen heeft opgeleverd. Die plausibiliteit is vooral hierin gelegen dat de resultaten van de diverse gehanteerde procedures op een aantal punten steeds in de richting van dezelfde talige biasbronnen wijzen. Die plausibiliteit wordt nog vergroot door het gegeven dat de eerste resultaten van het vervolgonderzoek naar itembias in de Eindtoets van 1993 grotendeels de inhoudelijke bevindingen inzake de toetsen van 1987 en 1989 ondersteunen (zie Uiterwijk & Vallen, te verschijnen).

In het volgende overzicht wordt aangegeven welke elementen en welke soorten opgaven gezien het onderzoek tot nu toe een gereede

kans lopen partijdig te zijn in het *nadeel* van Turkse en Marokkaanse leerlingen.

- Items die op meso-/microniveau een beroep doen op tekstbegrip, zoals dat het geval is bij vragen naar de betekenis van zinnen of verbanden tussen zinnen of naar de betekenis van woordgroepen of combinaties van woordgroepen. Dit geldt in het bijzonder wanneer om een woordelijke of geparafraseerde herhaling van expliciet in de tekst gegeven informatie wordt gevraagd. In aansluiting hierop kan nog worden opgemerkt dat dit tevens geldt voor moeilijke verwijzingen (bijvoorbeeld over langere tekstpassages heen of verwijzingen waarbij het bijwoord *er* een functie vervult) of potentieel ambigue verwijzingen (zoals het verwijzende *het* en verwijzwoorden als *die*, *deze*, *ze* en dergelijke, die een enkelvoudig of een meervoudig antecedent kunnen hebben).
- Items die vragen naar de betekenis van moeilijke inhoudswoorden, functiewoorden en woordcombinaties waarbij de betekenis niet of moeilijk uit de context kan worden afgeleid.
- Items die betrekking hebben op de kennis van de vorm van idiomatische (letterlijk of figuurlijk gebruikte) woordcombinaties (bijvoorbeeld: ergens zonder kleerscheuren afkomen) en/of conventies op het gebied van de zinsbouw (zinsvolgorde, inversie/vraagvormen).
- Items die veel contextmateriaal bevatten. Voor het oplossen van complexe items moet de leerling vaak een aantal tussenstappen maken. Uit de context moet dan worden afgeleid welke tussenstappen gemaakt moeten worden en dit wordt voor allochtone leerlingen extra moeilijk wanneer er in dergelijke items (vooral op zinsniveau) veel verwijzwoorden en impliciete zins- en tekstverbanden voorkomen. Vooral bij realistische rekenopgaven en opgaven voor informatieverwerking kan dit een rol spelen.
- Items die betrekking hebben op contextmateriaal (bijvoorbeeld een tekst) dat specifieke Nederlandse cultuurkennis vereist. Zo kunnen Turkse en Marokkaanse leerlingen minder vertrouwd zijn met het onderwerp dat in een tekst aan de orde wordt gesteld. Dit kan tot *itembias* leiden wanneer dat contextmateriaal cruciaal is voor het oplossen van de opgave. Dit onderdeel vergt trouwens uitgebreid nader onderzoek (Klatte-Folmer & Vallen 1995).

Verder heeft het onderzoek ook aanwijzingen

opgeleverd dat er ook soorten opgaven zijn die partijdig kunnen zijn in het *voordeel* van allochtone leerlingen.

- Items die op macroniveau (tekst-/alineaniveau) naar de hoofdgedachte van een tekst vragen. Hierbij gaat het dus om globaal tekstbegrip zoals dat voorkomt in opgaven als: 'Wat wil de schrijver in de eerste tien regels duidelijk maken?' Of: 'Welke conclusie kun je uit de laatste alinea trekken?'
- Items die vragen spellingfouten in werkwoorden en in woorden met een vast woordbeeld aan te geven.

Enkele praktische suggesties

Het onderzoek naar inhoudelijke bronnen van vraagpartijdigheid staat nog in de kinderschoenen en is vanwege zijn complexiteit werk van lange adem. Het lijkt niettemin zinvol dat toetsontwikkelaars en leerkrachten bij de ontwikkeling van toets- of proefwerkopgaven zich toch al bewust zijn van, en waar mogelijk rekening houden met, de behandelde talige bronnen van vraagpartijdigheid ten nadele van allochtone leerlingen. Dat is zeker relevant bij bijvoorbeeld de formulering en opbouw van opgaven/vragen voor (realistisch) rekenen/wiskunde (wat steeds meer een talig karakter krijgt) en zaakvakken als geschiedenis en aardrijkskunde. Maar al te vaak komt het voor dat allochtone leerlingen op deze terreinen onnodig en onbedoeld ingewikkeld geformuleerde vragen krijgen voorgelegd die een groot beroep doen op hun taalvaardigheid Nederlands, terwijl het niet de bedoeling is hun taalvaardigheid Nederlands te meten, maar een andere vaardigheid.

In dergelijke gevallen, met andere woorden wanneer het niet gaat om toetsing van taalvaardigheid Nederlands, zou op grond van het tot nu toe uitgevoerde onderzoek in talig opzicht nadrukkelijker gelet kunnen worden op bijvoorbeeld:

- het vermijden van moeilijke inhouds- en functiewoorden, ingewikkelde woordcombinaties of termen waar een eenvoudiger alternatief voorhanden is of waar met expliciterende omschrijvingen kan worden gewerkt;
- het gebruiken van korte zinnen waarin zo weinig mogelijk en zeker geen ambigue verwijzingen voorkomen;

- het vermijden van staande uitdrukkingen en figuurlijk en idiomatisch lastig taalgebruik;
- het duidelijk in de formulering aangeven van de te nemen tussenstappen in complexe en langere opgaven;
- het typografisch duidelijk aangeven van relevante kernbegrippen die voor beantwoording van opgaven van belang zijn.

Door rekening te houden met aandachtspunten in deze geest kunnen wellicht beter ongewenste talige stuikelblokken voor allochtone leerlingen worden voorkomen. Bovendien komt dat ook de validiteit van de instrumenten ten goede. Een toets behoort immers te meten wat ze pretendeert te meten en niet leerlingen te beoordelen op andere dan de bedoelde vaardigheden.

Literatuur

- Coenen, M. & T. Vallen, 'Itembias in de Eindtoets Basisonderwijs' in: *Pedagogische Studiën* 1991/68, p. 15-26.
- Dossey, J., I. Mullis, M. Lindquist & D. Chambers, *The Mathematics Report Card*. Princeton, Educational Testing Service, 1988.
- Driessen, G., *De onderwijspositie van allochtone leerlingen*. Nijmegen, Instituut voor Toegepaste Sociale Wetenschappen, 1990.
- Driessen, G., 'Discrepanties tussen toetsresultaten en doorstroomniveau. Positieve discriminatie bij de overgang basisonderwijs – voortgezet onderwijs?' in: *Pedagogische Studiën* 1991/68, p. 27-35.
- Jong, M. de & T. Vallen, 'Linguïstische en culturele bronnen van itembias in de Eindtoets Basisonderwijs voor leerlingen uit etnische minderheidsgroepen' in: *Pedagogische Studiën* 1989/66, p. 390-402.
- Klatte-Folmer, J. & T. Vallen, 'Een toets mag wel moeilijk zijn, maar niet partijdig: over talige en culturele bias voor allochtone leerlingen in Nederlandstalige toetsen' in: E. Huls & J. Klatte-Folmer (red.), *Artikelen van de Tweede Socio-linguïstische Conferentie*. Delft, Eburon, 1995, p. 357-371.
- Langen, A. van & P. Jungbluth, 'Leerkracht-overwegingen bij de vervolgadvisering van migranten' in: *Tijdschrift voor Onderwijswetenschappen* 1992/22, p. 69-79.
- Meijnen, G. & F. Riemersma, 'Schoolcarrières: een klassenkwestie?' in: Commissie Allochtone Leerlingen in het Onderwijs, *Ceders in de tuin, deel twee*. Zoetermeer, Ministerie van Onderwijs en Wetenschappen, 1992, p. 1-55.
- Mulder, L. & P. Tesser, *De schoolkeuze van allochtone leerlingen*. Nijmegen, Instituut voor Toegepaste Sociale Wetenschappen, 1991.
- Uiterwijk, H., *De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen*. Arnhem, Instituut voor Toetsontwikkeling (CITO), 1994.
- Uiterwijk, H. & T. Vallen, *Onderzoek naar bias voor allochtone leerlingen in de CITO-Eindtoets Basisonderwijs*, te verschijnen.
- Waal, M. van de, *Expert-oordelen over potentiële bronnen van itembias in de Eindtoets Basisonderwijs*. (doctoraalscriptie) Tilburg, KUB/Faculteit der Letteren, 1992.
- Wit, W. de, C. Suhre & L. Mulder, *De onderwijspositie van de OVB-doelgroep leerlingen na drie jaar voortgezet onderwijs*. Nijmegen, Instituut voor Toegepaste Sociale Wetenschappen, 1993.