

# Linguïstische en culturele bronnen van itembias in de Eindtoets Basisonderwijs voor leerlingen uit etnische minderheidsgroepen

M. DE JONG en T. VALLEN\*

Werkverband Taal & Minderheden, Faculteit der Letteren, Katholieke Universiteit Brabant

## Samenvatting

*In deze bijdrage wordt op basis van literatuurstudie nagegaan in hoeverre linguïstische en culturele factoren zijn te traceren die mogelijk de oorzaak zijn van lagere prestaties van allochtone subgroepen op de Eindtoets Basisonderwijs, terwijl het kennis- en vaardigheidsniveau van die subgroepen niet lager ligt dan dat van vergelijkbare autochtone. Dit fenomeen dat de construct-validiteit van een toets aantast, wordt 'itembias' genoemd. De uit de literatuur naar voren gekomen mogelijke bronnen van itembias zijn toegepast op items uit de Eindtoets Basisonderwijs 1987. Daarbij is gebleken dat het uitermate belangrijk is om onderscheid te maken tussen (door de toetsconstructeurs) bedoelde moeilijkheid en onbedoelde moeilijkheid (itembias). Immers, een goede toets bevat ook moeilijke onderdelen om te kunnen discrimineren tussen goede en minder goede leerlingen. Tevens is de kwalificatie 'gebiased' sterk afhankelijk van de soort vaardigheid die in een item wordt gemeten: een moeilijk woord is terecht moeilijk in een woordenschat-item, maar kan onbedoeld moeilijk zijn in een rekenvraagstuk. Daarnaast blijkt niet altijd duidelijk te zijn wat precies de vaardigheid is die in een item wordt gemeten en blijken gemeten vaardigheden uiteen te lopen van enkelvoudig tot meervoudig-complex. Bovendien worden items meestal in een (cultuurspecifieke) context aangeboden die onbedoeld moeilijk kan zijn in de ene situatie, maar juist bedoeld moeilijk in de andere. Met name de rol die vocabulaire lijkt te spelen bij het vaststellen van mogelijke bias-bronnen, springt*

*in het oog. Om de Eindtoets succesvol te kunnen maken, lijkt het beheersen van de gemeten vaardigheden niet voldoende; naast een relatief hoog vocabulairniveau lijkt een goede kennis van vaak onzichtbaar blijvende culturele aspecten vereist om cultuurgeladen tekstelementen te kunnen begrijpen en taken goed te kunnen interpreteren.*

## 1 Inleiding

Om de kwaliteit van de toets op peil te houden wordt in het kader van de Eindtoets Basisonderwijs regelmatig onderzoek gedaan. Een van de lopende onderzoeksprojecten draagt als titel 'Item- en testbias voor etnische groepen in de Eindtoets Basisonderwijs', een samenwerkingsproject van de Afdeling Basisonderwijs van het CITO en de Letterenfaculteit van de Katholieke Universiteit Brabant.

De doelstelling van het onderzoek is na te gaan of er bij bepaalde onderdelen van de Eindtoets sprake is van onbedoelde benadelingen ('biases') voor leerlingen uit etnische minderheidsgroepen die samenhangen met hun talige en/of culturele achtergronden. Indien dergelijke biases worden opgespoord, ligt het in de bedoeling op termijn voorstellen voor bijstellingen van toetsonderdelen te formuleren, zodat de kwaliteit van de toets voor uiteenlopende etnische groepen vergelijkbaar wordt. Het is natuurlijk *niet* de bedoeling de kwaliteit en het niveau van de toets aan te tasten. Evenmin wordt beoogd om voor (etnische) subgroepen van toetsdeelnemers verschillende normen te gaan hanteren. Noch het onderwijs noch de perspectieven van allochtone leerlingen in het vervolgonderwijs zouden bij dergelijke maatregelen gebaat zijn. Bovendien zouden deze maatregelen in strijd zijn met de doelstellingen van de Eindtoets (zie 2).

Het onderzoek is opgezet in meerdere cycli. Allereerst is op basis van literatuurstudie geprobeerd verwachtingen inzake linguïstische

\*Wij danken M. Coenen, A. Kerkhoff en H. Uiterwijk voor hun kritisch commentaar op een eerdere versie van dit artikel.

en culturele biases voor leerlingen uit etnische minderheden in de Eindtoets 1987 te formuleren, die vervolgens statistisch worden getoetst. Op basis van die toetsing worden de verwachtingen bijgesteld en vervolgens in 1989 opnieuw getoetst. Deze bijdrage geeft een beeld van de voor het formuleren van verwachtingen noodzakelijke theorievorming.

In het onderzoek spelen de noties 'itembias' en 'testbias' een centrale rol. Het deelonderzoek naar itembias heeft betrekking op de construct-validiteit van de toets. Bij testbias gaat het om de predictieve validiteit. In deze bijdrage wordt uitsluitend ingegaan op itembias, met name op theoretische overwegingen en dilemma's die opdoemen wanneer men itembias bestudeert in relatie tot talige en culturele achtergronden van toetsdeelnemers. Allereerst wordt in paragraaf 2 stilgestaan bij het fenomeen itembias en bij de manieren waarop dat tot nu toe in onderzoek veelal is benaderd. Mogelijke linguïstische bronnen van itembias staan centraal in paragraaf 3. Mogelijke cultureel bepaalde biases komen aan de orde in paragraaf 4. In paragraaf 5 worden enkele eerste voorzichtige conclusies geformuleerd en wordt een mogelijk perspectief voor onderzoek geschetst.

## 2 Itembias

Er is sprake van itembias wanneer de kans op een goed antwoord voor leerlingen met gelijke kennis en vaardigheden, maar behorend tot onderscheiden subgroepen, verschillend is (Shepard, 1982; Scheuneman, 1988). Met andere woorden, een toets of een toetsitem met een bias is 'partijdig' (Kok, 1988): systematisch wordt de ene subgroep (meestal de meerderheidsgroep) bevoordeeld en worden andere subgroepen (meestal minderheidsgroepen) benadeeld.

Het meeste onderzoek naar itembias kenmerkt zich door het achteraf (meestal statistisch) analyseren van gegevens met als doel daarin aanwijzingen te vinden voor noodzakelijke wijzigingen, mede met het oog op het samenstellen van nieuwe toetsen. Veel gebruikte statistische procedures daarbij zijn de Mantel-Haenszel-techniek (Holland & Thayer, 1986) en op Item-Response-Theorie gebaseerde benaderingen. Deze werkwijze leidt tot een aantal problemen die aan de hand van een

voorbeeld uit de Eindtoets 1986 kunnen worden geïllustreerd. De bedoeling van het item is een eventuele spellingsfout aan te kruisen. Bij dit item wijkt de verschillscore tussen allochtonen en autochtonen significant af van de gemiddelde verschillscore (op basis van een p-waarde-analyse; zie De Jong, Uiterwijk, Kerkhoff & Vallen, 1987).

- a. Er worden veel gevallen van *fraude* ontdekt.
- b. Hij bekeek de *voorstelling* met groeiende *ernst*.
- c. De *fonken* vlogen eraf.
- d. Geen fout.

Indien er bij dit item van statistische bias voor allochtonen sprake zou zijn, kan deze op verschillende manieren verklaard worden, variërend van benodigde kennis uit een ander kennisdomein dan spelling (in dit geval kennis van uitdrukkingen) tot interferentie van de eerste taal (bijvoorbeeld het stemloos realiseren van de 'v'). Omdat hypothesevorming vooraf ontbreekt, hebben dergelijke verklaringen een ad hoc karakter en is interpretatie van de resultaten problematisch. Omdat subgroepen in zeer veel opzichten van elkaar verschillen, is er in principe een aanzienlijk aantal interpretaties van geobserveerde gedragsverschillen mogelijk. Als uitgangspunt voor het verbeteren van toetsitems lijken dergelijke verklaringen achteraf dus minder geschikt. Scheuneman (1988) wijst in dit verband op het ontbreken van een theoretisch kader. Als dat er zou zijn, zou het mogelijk zijn om via een 'logisch'-inhoudelijke analyse items te selecteren die naar verwachting problemen voor bepaalde subgroepen opleveren. Op deze manier krijgt zo'n eerste inhoudelijke analyse een hypothesevormend en de statistische analyse die erop volgt een hypothesetoetsend karakter.

In nagenoeg alle discussies over itembias duikt veelvuldig de term 'moeilijkheid' op. Moeilijkheid mag niet als synoniem voor bias worden opgevat. De functie van toetsen als de Eindtoets is immers te discrimineren tussen succesvolle en minder succesvolle leerlingen. Als een item moeilijk is voor bepaalde leerlingen, wordt meestal voldaan aan de bedoeling van de toets: het meten van verschillen in vaardigheid. Pas wanneer die moeilijkheid *onterecht of onbedoeld* is, is er mogelijk sprake van bias. Onderzoek naar itembias heeft nu als grootste probleem om vanuit voorspelde be-

doelde moeilijkheid te komen tot voorspelde onbedoelde moeilijkheid (bias). Daarbij speelt de vaardigheid die in een specifiek item wordt gemeten uiteraard een cruciale rol.

Bronnen van itembias kunnen worden gezocht in culturele en talige factoren. Toetsitems kunnen voor alloctonen onterecht moeilijker zijn dan voor autoctonen, omdat de inhoud voor hen vreemd of onbekend is. Ook de formulering van een instructie of van een item kan onbedoeld moeilijk zijn (Brière, 1973). Bij culturele biases lijkt vooral ontbrekende voorkennis een centrale rol te spelen; bij talige biases gaat het met name om taalelementen en teksteigenschappen.

De inhoud van de Eindtoets is gebaseerd op communale cognitieve doelstellingen, die geacht worden te gelden voor alle leerlingen aan het eind van het basisonderwijs (Cito, 1986). Deze communale doelstellingen worden in toetsitems geconcretiseerd; de items worden gerealiseerd in contexten. Daarbij wordt er stilzwijgend van uitgegaan dat deze contexten eveneens communaal zijn. Wanneer dat echter niet het geval is, is er sprake van itembias: voor leerlingen uit bepaalde subgroepen wordt het item daardoor immers onbedoeld moeilijk, waardoor de kans op een goed antwoord voor die leerlingen afneemt, terwijl hun vaardigheidsniveau hetzelfde is.

### 3 *Mogelijke linguïstische bronnen van itembias*

Met welke elementen van de tweede taal (verder T2) alloctone leerlingen aan het eind van het basisonderwijs specifieke problemen hebben, is nauwelijks bekend. Meestal zijn het gevorderde T2-leerders die waarschijnlijk in hoge mate dezelfde problemen met het Nederlands hebben als autoctone leerlingen. Sommige taaleigenschappen zijn immers relatief moeilijker voor iedereen of worden relatief laat in zowel het T1- als het T2-leerproces verworven. Zo laat bijvoorbeeld Verhoeven (1984, 1987) zien dat clausale en verbale referentie in het Nederlands voor zowel Turkse als Nederlandse kinderen moeilijker is en later verworven wordt dan referentie naar personen, tijd en ruimte.

Eén onderdeel van het T2-verwervingsproces komt recentelijk steeds als problematisch ook voor gevorderde T2-leerders naar voren,

nl. *woordenschat* (zie o.a. Appel, 1985; Coenen & Vermeer, 1988; Kerkhoff, 1988; Vermeer, 1986). Lange tijd is men ervan uitgegaan dat het leren van grammaticaregels moeilijk is en dat het leren van woorden vrij gemakkelijk is en als het ware vanzelf als een soort bijproduct daarvan plaatsvindt (Van Ginkel, 1988). Meera (1980) noemt woordenschatverwerving "a neglected aspect of language learning" en Levenston (1979, p. 147) beweert zelfs dat "second language lexical acquisition has been a victim of discrimination".

#### 3.1 *Moelijkheidsgraad van woorden*

Op het moment van afname van de Eindtoets beschikken de meeste alloctone leerlingen over een geringere Nederlandse woordenschat dan de meeste autoctone leerlingen. Voor het succesvol kunnen maken van de Eindtoets is een goede woordenschat echter zeer belangrijk; bekendheid met de in de teksten, instructies en items voorkomende woorden is een absolute voorwaarde om de opgaven te kunnen begrijpen en adequaat te kunnen beantwoorden. De in de Eindtoets gebruikte woorden zijn afgestemd op de gemiddelde woordenschat en het gemiddelde vocabulaire-niveau van de doorsnee autoctone leerling. Een woordenschattoets of een Eindtoets-onderdeel waarmee woordkennis wordt gemeten beoogt nu juist te differentiëren tussen leerlingen op dit specifieke gebied: de (moeilijke) woorden maken deel uit van de gemeten vaardigheid. Bij dergelijke toetsen of toetsonderdelen is er dan ook volstrekt geen sprake van een bias voor alloctone kinderen, wanneer ze lagere scores behalen dan autoctone kinderen. Bevatten daarentegen rekenopgaven of opgaven inzake kaartlezen moeilijk vocabulaire, of wordt kennis van interpunctie gemeten met behulp van zinnen die een moeilijk vocabulaire bevatten, dan zou er in termen van bias voor alloctonen wel wat aan de hand kunnen zijn. In dergelijke gevallen beschikken de betreffende leerlingen misschien wel over de inhoudelijke vaardigheden, maar ze komen daar niet of in onvoldoende mate aan toe, omdat allereerst de barrière van de voor hen moeilijke woorden moet worden genomen.

##### 3.1.1 *Woordfrequentie*

De moeilijkheidsgraad van vocabulaire is lastig vast te stellen. Vaak wordt ervan uitgegaan

dat infrequente woorden moeilijker zijn dan frequente. Daarbij zijn minstens twee kanttekeningen te plaatsen. Ten eerste gaat een dergelijke redenering niet op voor veel functiewoorden. Deze komen zeer frequent voor, maar worden in zowel T1 (Van Wijk & Kempen, 1982, 1985) als T2 (Broeder, Coenen, Extra, Van Hout & Zerrouk, 1986) relatief laat verworven. De reden daarvoor schuilt misschien in het feit dat veel functiewoorden als verbindende elementen tussen zinnen en tekstdelen opereren, linguïstische niveaus die later worden verworven dan het klank-, woord- en zinsniveau. Ten tweede blijven bij deze redenering verschillen binnen de groep inhoudswoorden buiten beschouwing. Sommige frequente inhoudswoorden kunnen juist lastig zijn, omdat hun betekenis varieert al naar gelang de situationele context. Zo heeft het woord 'onderwerp' binnen het grammatica-onderwijs een meer abstracte betekenis die afwijkt van die in het alledaagse taalgebruik. Zowel functie- als inhoudswoorden kunnen bovendien voor verschillende groepen leerlingen een verschillende semantische lading hebben. Zo laat Gerritsma (1987) zien dat het functiewoord 'ik' in het Maleis naast een individuele ook een collectieve betekenis kan hebben.

Overigens mag niet uit het oog verloren worden dat het bij het begrijpen van woorden niet alleen gaat om het (her)kennen van de woordbetekenis alleen. Carpenter en Just (1986) merken op dat kinderen die moeite hebben met lezen zoveel energie nodig hebben voor de techniek van het lezen zelf dat daardoor het begrijpen van *wat* ze lezen wordt belemmerd. Ook blijken goede lezers beter in staat om de betekenis van moeilijke en onbekende woorden uit de context af te leiden dan zwakke lezers.

De rol van de moeilijkheidsgraad van een woord bij toetsonderdelen die geen woordenschat beogen te meten kan wellicht verduidelijkt worden met een voorbeeld (rubriek 'meten'), dat, evenals de verdere voorbeelden in deze bijdrage, afkomstig is uit de Eindtoets 1987.

(1) Je ziet hier vier gewichtsaanduidingen.

Welke aanduiding geeft ongeveer  $\frac{1}{2}$  kg aan?

A. 0,51 g. B. 5,1 g. C. 49 g. D. 496 g.

Het woord 'gewichtsaanduiding' heeft een frequentie van 0 in Uit den Boogaart (1975).

Bovendien is het een samengesteld woord: voor een goed begrip daarvan moeten de woorddelen als informatie-eenheden worden herkend. Verder moet de lezer afleiden dat met het woord 'aanduiding' hetzelfde wordt bedoeld als met 'gewichtsaanduiding'. Tot de doelstelling van de rubriek 'meten' hoort onder andere "het kennen en kunnen gebruiken" van maateenheden als gewicht, kilogram e.d. (Cito, 1986, p. 29). De term 'gewichtsaanduiding' wordt daarbij niet genoemd en valt dus binnen de contextuele inbedding. De formulering van de vraagstelling bij dit item zou derhalve een talige bias kunnen bevatten.

Ook bij taalrubrieken zou een moeilijk woord een potentiële bron van itembias kunnen zijn (zie voorbeeld (2), rubriek 'spelling van niet-werkwoorden').

(2) A. Je moet niet zo *snauwen*.

B. Jij bent een echte *optimist*.

C. De *technische* dienst kwam meteen.

D. Geen fout.

De leerling moet nagaan of het item een spellingfout bevat of niet. Nog afgezien van de vraag of spelling (een op afspraken berustende technische vaardigheid) thuis hoort onder de rubriek 'taal', kan men zich bij dit item afvragen of voor het meten van spellingvaardigheid gebruik moet worden gemaakt van een laagfrequent woord als 'optimist'. Kennis van de betekenis van een woord is op zichzelf echter niet per definitie noodzakelijk voor correcte spelling. Toch worden bij spelling veelvuldig infrequente woorden gebruikt, waardoor het bepalen van een mogelijke biasbron uitermate lastig wordt.

### 3.1.2 Woordambigüiteit

Ook ambigue of vage woorden c.q. formuleringen zouden een potentiële bron van bias voor allochtone (sub)groepen kunnen vormen. Of dit het geval is, is moeilijk aan te geven, omdat er op dit terrein tot nu toe geen onderzoek voorhanden is. Wel kan geconstateerd worden dat er in de Eindtoets vaak uiteenlopende formuleringen voorkomen bij items die dezelfde vaardigheid beogen te meten. Dit gebeurt uiteraard om te voorkomen dat leerlingen afhankelijk zijn van slechts één type formulering. Alternatieve formuleringen zijn echter niet steeds even helder, zoals blijkt uit de voorbeeldzinnen (3) en (4) (beide betreffen vragen uit de rubriek 'interpreteerbaar taalgebruik' en behoren bij een tekst).

(3) Wat had Callista ook kunnen schrijven voor: "niet te laat"?

(4) Wat had Casper in plaats van "te kalmeren" ook kunnen gebruiken?

De formulering 'schrijven voor' is in vergelijking met 'in plaats van' relatief vaag en voor meerdere interpretaties vatbaar. Optredende verschillen tussen de scores van autochtonen en allochtonen bij vage of ambigue itemformulering hoeven niet te betekenen dat er van itembias sprake is. De foutscore kan immers ook veroorzaakt worden door het niet beheersen van de (gemeten) vaardigheid. Daarnaast kan ook de mate van leesvaardigheid van invloed zijn: voordat het item kan worden beantwoord, moet immers de bijbehorende tekst worden gelezen en begrepen. Kortom: naarmate de gemeten vaardigheid complexer wordt en meerdere vaardigheden (tegelijk) een rol spelen bij het oplossen van een item, ontstaat een onontwarbaar geheel van oorzaken, waarvan bias er mogelijksterwijs één kan zijn.

### 3.2 Moeilijkheidsgraad van zinnen

Het begrijpen van de syntactische structuur vormt een voorwaarde voor het ordenen, interpreteren en reconstrueren van de semantische inhoud van een zin. Taalgebruikers blijken kennis te bezitten over conventies inzake het produceren en verwerken/begrijpen van syntactische structuren: bijvoorbeeld hoofdzinnen staan vóór bijzinnen; twee gebeurtenissen worden in chronologische volgorde verteld; bevestigende zinnen zijn 'normaler' dan ontkennende zinnen (Clark & Clark, 1977).

Verwerkbaarheid van structuren heeft behalve met conventionele kennis ook te maken met andere factoren, zoals de beperkte capaciteit van het menselijk werkgeheugen: slechts een beperkt aantal proposities kan tegelijk verwerkt worden. Zinsstructuren die een relatie te grote belasting vormen voor het werkgeheugen, bemoeilijken het verwerken. Tangconstructies als (5) vormen de meest bekende voorbeelden van structuren die het geheugen zwaar belasten.

(5) "De agent die de dief die de poen die de vrek vergaarde, stal, nazat, viel." (ontleend aan Maureau 1979, p. 95)

Het is lastig te voorspellen welke zinsstructuren moeilijk zijn. De adviezen van schrijfadviseurs (Renkema, 1982; Steehouder & Jansen, 1979) zijn meestal niet op toetsing gebaseerd

en maken voorspellingen onmogelijk. In de tweede plaats speelt bij zinsverwerking de kennis van de wereld een rol. Maureau (1979) bespreekt psycholinguïstisch onderzoek waarin zinnen met een identieke syntactische structuur door de daarbij vereiste kennis van de werkelijkheid toch verschillen in begrijpelijkheid. In de derde plaats hebben contextuele factoren invloed op het verwerken van syntactische structuren. Zo werd bijvoorbeeld enige tijd op grond van onderzoek aangenomen dat actieve zinnen sneller worden verwerkt dan passieve. Clark en Clark (1977) nuanceren dit echter door contextuele factoren bij de syntactische analyse te betrekken. In de vierde plaats bevatten zinnen meestal elementen die vooruit- of terugverwijzen in de tekst. Zinnen en zinselementen kunnen daarom beter in samenhang worden bestudeerd. Omdat zoveel verschillende factoren (tegelijk) werkzaam zijn, is het nagenoeg onmogelijk om aan te geven waar mogelijke biases op zinsniveau voor allochtone (sub)groepen uit kunnen bestaan. De vraagwoorden waarmee vraagzinnen beginnen bieden echter misschien enig perspectief (waarmee we weer gedeeltelijk terugkeren op woordniveau).

#### 3.2.1 Vraagwoorden

Schuurs (1984) bestudeerde de ontwikkeling van een aantal typen vraagzinnen in T1 en T2 bij Turkse en Marokkaanse kinderen van 9,5 jaar en Nederlandse kinderen van 4 jaar. Hij vond overeenkomsten die sprekender zijn dan de verschillen tussen T1- en T2-leerders en tussen T2-leerders met uiteenlopende uitgangstalen. Bij de acht gebruikte vraagwoordcategorieën werd de volgende verwervingsvolgorde gevonden: of, wie, ja/nee, waar, wat, hoe, waarom en wanneer. De laatste drie worden volgens dit onderzoek relatief laat verworven en zouden dus in verhouding moeilijker kunnen zijn dan de eerste vijf. Wanneer dat ook nog geldt voor de leeftijds categorie waartoe allochtone Eindtoetsdeelnemers behoren, zou verondersteld kunnen worden dat de vraagwoorden 'hoe', 'waarom' of 'wanneer' een potentiële bron van bias voor deze kinderen zouden kunnen zijn. Bekijkken we opnieuw twee voorbeelden (rubriek 'hoofdrekennen'), die dezelfde vaardigheid beogen te meten:

(6)  $2178,45 : 479 = *$

Hoe kun je deze deling controleren?

A.  $* : 479 = 2178,45$

B.  $* \times 479 = 2178,45$

$$C. * : 2178,45 = 479$$

$$D. * \times 2178,45 = 479$$

$$(7) \quad 49537 \\ \quad \quad 698 -$$

Hans en Miranda controleren de uitkomst van deze aftrekking. Ze doen het ieder op een andere manier.

$$\text{Hans: } 49537 - \text{  } = 698$$

$$\text{Miranda: } \text{  } - 698 = 49537$$

Wie doet het juist?

- A. alleen Hans                      C. beiden  
B. alleen Miranda                  D. geen van beiden

In beide items wordt kennis en toepassing van rekenkundige controleregels gemeten. In (6) kan kennis van het mogelijk relatief moeilijke vraagwoord 'hoe' daarbij misschien interfereren. Op grond van het voorafgaande lijkt derhalve (7) gemakkelijker te verwerken dan (6): er wordt een concrete context aangeboden alsmede een relatief gemakkelijk vraagwoord ('wie'). In dit laatste item wordt echter ook een sterker beroep gedaan op begrijpend lezen. De formulering bevat bijvoorbeeld vijf verwijswaarden waarvan er zeker één relatief moeilijk verwerkbaar is (zie 3.3). Ook moet de lezer zien dat alle hokjes hetzelfde getal representeren (zie 4.3). Hierdoor is het toch mogelijk dat ook (9) een bias bevat, zij het door andere oorzaken dan (6).

### 3.3 Referenties in teksten

De linguïstische middelen om vooruit- of terug te verwijzen (referenties) naar elementen binnen dezelfde zin of (elementen in) voorafgaande of volgende zinnen worden zowel in T1 als T2 relatief laat verworven. Vooral minder taalvaardige kinderen blijken moeite met verwijswaarden te hebben. Naast de reeds genoemde problemen bij clausale en verbale referenties blijkt vooral voornaamwoordelijke verwijzing (pronominale referentie) over lange afstanden in teksten problematisch.

Bartlett (1984) doet verslag van het gebruik van referenties in geschreven verhalen van succesvolle en minder succesvolle basisschoolleerlingen (klas 5 t/m 7). Naarmate de leerlingen ouder en/of 'beter' zijn, neemt het aantal referentiële ambiguïteiten af. Bovendien blijken goede schrijvers in staat hun taal aan te passen bij moeilijke contexten. Zo komen in (8) en (9) twee personen van dezelfde leeftijd en hetzelfde geslacht voor. Om ambiguïteiten

te voorkomen, is het beter om ze nominaal aan te duiden. Goede schrijvers blijken in dergelijke gevallen vaker nominaal te verwijzen zoals in (8), terwijl zwakke schrijvers pronominale vormen als in (9) blijven gebruiken.

(8) Mik en Mak draaiden zich om.

*Mik* lachte.

(9) Mik en Mak draaiden zich om.

\**Hij* lachte.

Het gebruik van verwijswaarden in teksten hangt nauw samen met de wijze waarop de door de schrijver als bekend veronderstelde (Given) en als nieuw bedoelde (New) informatie wordt aangeboden (Clark & Clark, 1977; Noordman & Vonk, 1987). In conventionele zinnen blijkt bekende informatie vooraf te gaan aan nieuwe. In andere zinnen ligt het informatiepatroon juist andersom. Sommige informatiepatronen vragen een kortere verwerkingstijd dan andere (Glatt, 1982). In deze 'favoriete' (conventionele) patronen blijkt in de vervolgzin Given steeds vooraf te gaan aan New. Bekende informatie wordt redundant gemarkeerd; de overlapping helpt de lezer bij het identificeren van materiaal als Given of als New. Lezers blijken kennis te hebben van genoemde patronen en hun leesgedrag erop af te stemmen. Hieronder volgen vier verschillende patronen, ontleend aan Glatt (1982), waarvan (10) en (11) sneller verwerkt werden door de proefpersonen.

(10) *Becky* houdt van tropische vissen. *Zij* bezoekt elke week het aquarium.

(11) *Budd* helpt *Sandra* in de auto. *Zij* kan niet lopen.

(12) *De scheidsrechter* legde de spelregels uit. Een deelnemer begreep *hem* niet.

(13) *De scheidsrechter* legde *de spelregels* uit. Een deelnemer begreep *ze* niet.

Op grond van het bovenstaande zou kunnen worden verwacht dat in bepaalde toetsonderdelen moeilijke en ambigue referenties een bron van bias zouden kunnen zijn voor allochtone (sub)groepen. *Ambigue* referenties bevatten verwijzingen als in (9). *Moelijke* referenties zijn: referenties waarbij Given niet voor New komt, referenties waarbij het verwijswaard getalsmatig afwijkt van de referent, referenties waarbij over relatief grote tekstdelen heen wordt verwezen, clausale en/of verbale referenties. Het volgende voorbeeld stamt uit de rubriek 'vraagstukken'.

(14) Per jaar gaf een gezin gemiddeld f1500,- uit aan aardappels en groenten.

*Ze wilden bezuinigen. Ze huurden een jaar een tuin van 200 m<sup>2</sup>. voor f2,- per vierkante meter. De overige kosten waren f80,-. Ze moesten nog voor een bedrag van f400,- aan aardappelen en groenten in de winkel kopen. De rest kwam uit de tuin. Hoeveel had dat gezin bespaard met tuinieren?* (cursivering MdJ/TV)

A. f620,-

B. f820,-

C. f920,-

D. f1020,-

In (14) wijkt het verwijzwoord 'ze' (meervoud) getalsmatig af van de referent 'een gezin' (enkelvoud); in de laatste zin wordt vervolgens weer overgegaan op enkelvoud. Bovendien zou de lezer bij het laatste 'ze', volgens de gangbare verwachting, de referent in eerste instantie in de voorafgaande zin kunnen zoeken ('de overige kosten'). Van de lezer wordt in dit geval een extra zoekactie geëist: over de verwachte referent heen moet - via de twee eerste pronomina - de referent worden opgespoord in de eerste zin. Dit kost relatief veel tijd en kan het tekstbegrip nadelig beïnvloeden.

### 3.4 Tekstsignalen

In teksten kunnen expliciterende middelen worden aangebracht, die verondersteld worden het tekstbegrip te vergemakkelijken. Op globaal niveau, dat wil zeggen de zinsgrens overschrijdend, is enige empirische steun vergaard voor de faciliterende werking van signaalzinnen die de globale tekststructuur expliciteren: de zogenoemde structuuraanduiders (bijvoorbeeld 'In de volgende alinea worden hiervoor argumenten gegeven'). Sanders (1986) vond dat de leestijden bij proefpersonen afnamen naarmate meer expliciet werd aangeduid voor de lezer.

Teksten zonder structuuraanduiders kunnen problemen opleveren voor minder taalvaardige kinderen. Met dergelijke 'structuurimpliciete' teksten kan uiteraard geselecteerd worden tussen goede en minder goede lezers. In toetsen moet echter onnodige complicering vanwege foutieve of niet-gangbare structuuraanduiders worden vermeden; (15) is daarvan een voorbeeld. Het betreft een passage uit een tekst van de Eindtoets 1987 over de bedreiging van de weidevogels met een daarbij behorende vraag.

(15) Verstedelijking

De oorzaken liggen in stad en boerderij.

De eerste komt met de gevolgen van de

verstedelijking als de grootste vogelverschrikker uit de geschiedenis aanzetten. Sterk veranderde landbouwmethoden gaan uiteraard ook al niet van de belangen van de weidevogels uit. (cursivering MdJ/TV)

Wat of wie wordt bedoeld met: *de grootste vogelverschrikker uit de geschiedenis?*

A. de boerderij

B. de moderne boer

C. de stad

D. de veranderde landbouw

In (15) wijken de structuuraanduiders af van wat gangbaar is: 'de eerste' en 'ook' in plaats van bijvoorbeeld 'ten eerste' en 'ten tweede'. Daarnaast is de gebruikte formulering 'de eerste komt aanzetten' onnodig gecompliceerd en kan daardoor verwarrend werken (zie 3.1.2). Tenslotte past de titel niet bij de tekstinhoud: 'verstedelijking' betreft slechts één van de twee oorzaken; de tekst blijkt bovendien uitsluitend in te gaan op de andere oorzaak. Op grond van het voorafgaande is het voorstelbaar dat niet-gangbare of foutieve structuuraanduiders een bron van bias kunnen opleveren voor allochtone (sub)groepen. Het gaat daarbij om foutieve en/of minder gangbare signaalwoorden, alsmede niet passende titels, beide onder voorwaarde dat het foutieve/minder gangbare niet bevestigd wordt in een van de bijbehorende items.

## 4 Culturele bronnen van itembias

In deze paragraaf schetsen we kort enkele dilemma's in verband met de vraag waar het bij 'cultuur' in toetsen nu eigenlijk om gaat en gaan we in op culturele ladingen van teksten en op mogelijk cultureel bepaalde toetservaring.

### 4.1 Cultuur

De studie van cultuur kent vele benaderingen. Naast beschrijvingen en vergelijkingen van specifieke culturen zijn er veel theoretische (begrips)beschouwingen (Hagendoorn, 1986; Tennekes, 1985, 1986; Poortinga, 1982). Daarin wordt met name ingegaan op de volgende problemen:

- Naast een waarneembaar niveau (gedrag) is er een veel belangrijker, doch onzichtbaar blijvend niveau (waarden en normen, bete-

kenis). De beschrijvingen van dit laatste niveau variëren sterk en zijn (noodgedwongen) vaag.

- De relatie tussen gedrag en betekenisstelsel is in veel opzichten niet duidelijk. Vooral omdat mensen ook keuzes kunnen maken, zijn er grenzen aan de deterministische verklaarbaarheid en voorspelbaarheid van gedrag.
- Zijn alle menselijke gedragingen cultuurspecifiek of zijn er ook gemeenschappelijke (universele) gedragingen (cultuur-natuurdiscussie)?
- Veelal wordt uitgegaan van een relatief vaststaand betekenisstelsel, terwijl culturen voortdurend blijken te veranderen. Een dergelijke statische benadering leidt vaak tot oversimplificaties en stereotyperingen. Ook wordt de rol van machtsverhoudingen buiten beschouwing gelaten.

Samenvattend blijkt uit de literatuur een voorkeur voor dynamische cultuuromschrijvingen, waarin rekening wordt gehouden met onderliggende, onzichtbare betekenissen, met verschillen tussen individuen en subgroepen, alsmede met veranderings- en machtsprocessen. Omdat er nauwelijks beschrijvingen in die zin voorhanden zijn, wordt in de gangbare onderzoekspraktijk cultuur (of 'culturele oriëntatie') veelal verengd tot het zichtbare gedragsniveau. Daarbij beperkt men zich meestal tot die aspecten die als centraal voor de betreffende cultuur gelden en die men onderscheidend ten opzichte van andere culturen acht. Daarvoor worden termen gebruikt als 'herkenbaarheidscriteria' (Hagendoorn, 1986), 'cultuursymbolen' (Rinsampessy, 1984) of 'sleutelwaarden' (Watson-Gegeo & Gegeo, 1986). Opvallende culturele kenmerken worden op instrumentele wijze gehanteerd om "de eigenheid, het anders zijn, te benadrukken" (Tennekes, 1986, p. 3). Dit laatste wordt beschouwd als kenmerkend voor etnische groepen. Etniciteit wordt dus gekenmerkt door een bepaald type sociale identiteit dat verankerd is in een gemeenschappelijke culturele erfenis. Culturele gegevens worden in deze opvatting gezien als markeerders van de groeps grenzen en de groepsidentiteit ("het gevoel bij elkaar te horen"; a.w., p. 5). Met de maatschappelijke situatie van de groep kunnen zij veranderen, maar zij blijven hun functie als markeerders

houden. Vanwege de grote mate van wetenschappelijke overeenstemming over de term etniciteit en omdat in deze term zowel de culturele (identiteits-)factor als de sociaal-economische (situatieve) factor een plaats hebben, geeft Tennekes de voorkeur aan 'etniciteit' boven 'cultuur' en 'minderheid'. De meeste (im)migrantengroeperingen in Nederland kunnen dan ook het best worden gekarakteriseerd als etnische groep.

Op grond van de besproken inzichten en standpunten vatten wij de notie cultuur binnen ons onderzoek op in de zin van etniciteit. Daarbij beperken we ons tot culturele factoren voor zover die tekstbegrip of het begrijpen van de taak mogelijk bemoeilijken.

#### 4.2 Culturele lading van teksten

Het maken van toetsopgaven waarbij een tekst als context is gebruikt, veronderstelt dat de leerling deze tekst heeft begrepen/verwerkt. Een van de variabelen bij tekstbegrip is de mate waarin voorkennis over de tekstinhoud aanwezig is bij de lezer. Spilich, Vesonder, Chiesi en Voss (1979) laten zien hoe lezers met veel voorkennis in vergelijking met lezers met weinig voorkennis in staat waren belangrijke informatie te onthouden en een verslag te produceren dat veel samenhangender en vollediger was. Johnson (1981) vergeleek de prestaties van Iraanse studenten op tekstbegrijptaken bij een Amerikaanse folkloristische tekst met die bij een Iraanse folkloristische tekst. De cultuurbepaalde inhoud van de teksten bleek meer van invloed op het tekstbegrip dan het niveau van de linguïstische complexiteit. Kerkhoff en Vallen (1984) vergeleken de taalvaardigheid van Turkse, Molukse en Nederlandse kinderen op Nederlandstalige toetsen die een identieke linguïstische complexiteit bezaten maar qua inhoud cultureel verschillend gekleurd waren. Op de 'Nederlands-gearde' toets scoorden de Nederlandse kinderen significant hoger dan de Molukse en Turkse kinderen. Op de 'Turks-gearde' toets bleken de Turkse kinderen significant beter te scoren dan op de 'Nederlands-gearde' toets. Ditzelfde werd mutatis mutandis gevonden bij de Molukse kinderen. Johnston (1984) bestudeerde de effecten van voorkennis op tekstbegrip bij stads- en plattelandskinderen. Hij concludeert eveneens dat slechts weinig gezegd kan worden over de leesvaardigheid van een proefpersoon, wanneer de voorkennis niet



gemeten is. De cultuurbepaalde voorkennis die een tekst vereist, kan een taak dus onbedoeld moeilijk maken, waardoor er sprake zou kunnen zijn van culturele bias.

Een probleem dat zich bij dit type onderzoek voordoet, is de wijze waarop de in een tekst aanwezige voorkennis vastgesteld wordt. Daarbij gaat het om de (on)mogelijkheid van de onderzoeker inzake het vaststellen van culturele inhoudsaspecten (Richardson, 1987) en om de variabele wijze waarop voorkennis wordt gemeten (Johnston, 1984; Johnson, 1981; Kerkhoff & Vallen, 1984).

In de hierboven besproken onderzoeken werd het effect van voorkennis aangetoond met behulp van teksten waarvan de inhoud zeer cultuurspecifiek was. Het is de vraag of de in de Eindtoets gebruikte teksten een dermate sterke culturele lading hebben. Daarnaast speelt de volgende vraag: welke kennis van de wereld is bij het uitvoeren van een opdracht nog acceptabel en welke kennis niet? Mag bijvoorbeeld bij het beantwoorden van een rekenvraagstuk met procenten (zie (16)) verondersteld worden dat allochtone kinderen weten wat B.T.W. is en dat B.T.W. altijd 'erbij komt'? Met andere woorden: hoe cultuurspecifiek is de notioneel-functionele betekenis van B.T.W. en in hoeverre behoort die kennis tot het communale curriculum? In hoeverre spelen daarnaast andere elementen een rol bij het beantwoorden van dit item, zoals het verwerken van moeilijke onderdelen uit de tekst, bijvoorbeeld het verwijswoord *deze*, het vraagwoord *hoeveel* en het vocabulairniveau van *B.T.W.* en *inclusief*?

- (16) Vader koopt een naaimachine. Deze kost f800,- zonder B.T.W. De B.T.W. is 20%. Hoeveel moet vader betalen inclusief B.T.W.?  
A. f160,- B. f640,- C. f820,- D. f960,-

Het is duidelijk dat het voorspellen van culturele bronnen van itembias veel en complexe problemen met zich meebrengt. De theoretisch-inhoudelijke inschatting door de onderzoeker zal dan ook zeker moeten worden gecombineerd met aanvullende informatie die bij de leerlingen zelf wordt verzameld.

#### 4.3 Cultureel bepaalde toetservaring

Bij 'testwiseness' gaat het in het algemeen om ervaring in het maken van (gestandaardiseerde) toetsen, die op formele wijze worden

afgenomen, inclusief kennis van/vertrouwdheid met de soort taken die in dergelijke toetsen te verwachten zijn en kennis van effectieve oplossingsstrategieën. In hoeverre testwiseness cultuurspecifiek genoemd kan worden, wordt hieronder nagegaan.

Van de Vijver en Joe (1986) beschrijven onderzoek naar culturele verschillen in abstract denken. Niet-westerse, ongeschoolde groepen zouden een 'achterstand' vertonen wat betreft hun vermogen tot abstract denken (het vermogen om hypothesen en veronderstellingen te genereren over de werkelijkheid en deze op hun accuratesse te toetsen). Toch blijken genoemde groepen in het alledaagse leven wel degelijk abstract te kunnen redeneren. Hieruit is onder meer geconcludeerd dat abstract denken op zich universeel is, maar dat niet iedereen op alle gebieden daartoe in staat is c.q. daarin ervaren is.

Het waarnemen van en het denken over de werkelijkheid lijkt mede afhankelijk te zijn van de culturele groep waartoe iemand behoort. Van de Koppel en Schoots (1986) spreken in dit verband over 'conventies', dat wil zeggen impliciete en expliciete vormen van overeenstemming tussen leden van een culturele groep. In produkten van een cultuur, bijvoorbeeld in toetsen, manifesteren zich de conventies van die cultuur. Bij westerse toetsen kan daarbij gedacht worden aan de volgende conventies:

- de taakstructuur van een item wordt niet altijd expliciet (genoeg) uitgelegd in een opgave (Van de Vijver & Joe, 1986);
  - items uit een toets lijken onderling op elkaar, zodat nieuwe problemen naar analogie van eerdere problemen kunnen worden opgelost (Van de Vijver & Joe, 1986);
  - het oplossen van items vereist een bepaald type cognitieve stijl ('veldonaafhankelijkheid'), waarbij gebruik wordt gemaakt van externe referentiekaders (Van de Koppel & Schoots, 1986).
- Met name onduidelijkheid over de taakstructuur blijkt een belangrijke oorzaak te zijn voor het kiezen van foutieve antwoorden. Van de Vijver en Joe (1986) beschrijven werk van Luria die aan analfabete Russische boeren onder andere nonsens-syllogismen voorgelegde als in (17):
- (17) Alle ijsberen zijn zwart. Mijn vriend zag gisteren een ijsbeer. Was deze ijsbeer zwart?

De boeren gaven antwoorden als: "Dat zult u uw vriend moeten vragen" of "Er bestaan alleen maar witte ijsberen". De informanten interpreteerden het syllogisme analoog aan alledaagse vragen, dat wil zeggen gekoppeld aan het hier-en-nu. Door de onbekendheid met toetsen en taakstructuren wordt teruggegrepen naar bekende oplossingsstrategieën die in het leven van alledag (in een orale cultuur) uitstekend voldoen, maar die in een toetsituatie leiden tot foutieve antwoorden.

Algemeen wordt aangenomen dat scholing de cognitieve groei van leerlingen stimuleert, hetgeen wil zeggen dat "kinderen op school gegeneraliseerde oplossingsstrategieën leren hanteren (...) die gebruik maken van abstracte, verbale probleemrepresentaties" (Van de Vijver & Joe, 1986, p. 162). Dit veronderstelt dat de problemen die Luria tegenkwam bij de analfabete boeren, bij geschoolde groepen niet (meer) voorkomen. Dat dit laatste geenszins het geval is, blijkt uit Aronowitz (1984). Hij vergeleek interpretatiestrategieën die normaal gesproken gelden bij het lezen van teksten met de strategieën die vereist zijn voor een correcte interpretatie van leesvaardigheidsteksten. Lezers hanteren normaal gesproken verwachtingen als: een tekst is betekenisvol en informatief; een tekst is afgeleid uit een conventionele communicatieve gebeurtenis/context; om de tekst te kunnen begrijpen moet kennis van de wereld worden ingezet. Een succesvolle toetsperformance levert teksten op die volgens Aronowitz 'raar' overkomen. Dat komt omdat bij leesvaardigheidstaken de volgende relatief ongebruikelijke, niet geëxpliciteerde regels (maximes) van kracht zijn:

1. geef de voorkeur aan metalinguïstische teksten;
2. geef de voorkeur aan maximaal redundante, minimaal informatieve teksten;
3. ga uit van 'onschuldigheid'.

Met maxime 1 wordt bedoeld dat lezers moeten weten dat leesvaardigheidstaken vaak resulteren in teksten/uitingen die metalinguïstisch van aard zijn. In (18) wordt kennis van synoniemen getoetst. De tekst die resulteert uit het goede antwoord ("If you capture a rabbit, you catch it") is, vanuit het gangbare oogpunt van semantische structuur gezien, tautologisch; dit kan een acceptabele reden zijn om het goede antwoord niet te kiezen. Vanuit metalinguïstisch standpunt betreft het echter een acceptabele, welgevormde uiting.

- (18) If you capture a rabbit, you  
hurt it    kill it    tease it    catch it

Met maxime 2 wordt bedoeld dat lezers moeten weten dat leesvaardigheidsteksten vaak resulteren in teksten waarin informatie wordt herhaald die al in de tekst aanwezig was. In voorbeeld (19) is het vanuit het standpunt van informatiestructuur redundant om te stellen dat Braziliaans hout wordt gebruikt om verf te maken. Gezien vanuit de regel dat kennis van de wereld wordt ingezet bij het lezen van teksten, is de keuze van 'barns' plausibeler: dit levert immers een tekst op die informatiever en minder redundant, en daardoor coherenter is.

- (19) The country of Brazil was named for brazilwood from which a red dye is made. The wood grew abundantly in Brazil. It is used for making  
barns    dye    cakes    cheese

Met maxime 3 wordt bedoeld dat lezers moeten weten dat zij de tekst zo letterlijk en serieus mogelijk moeten interpreteren, tenzij de tekstvragen aanwijzingen bevatten dat de inhoud bijvoorbeeld ironisch bedoeld is. Bovendien moeten de beschreven handelingen consistent met expliciet in de tekst aanwezige informatie worden begrepen. In voorbeeld (20) is 'depend on him' consistent met maxime 3. Wordt de tekst ironisch opgevat, dan is echter het foutieve antwoord 'pay him' voor de hand liggend.

- (20) The postman always comes, regardless of weather. We can always  
write to him    pay him  
depend on him    hear him

Aronowitz beschouwt de maximes als buiten de leesvaardigheid vallend. Hoewel dit betwisteld kan worden (het vermogen om de tekstwerkelijkheid los te zien van de eigen ervaringswerkelijkheid kan immers ook worden beschouwd als behorend tot leesvaardigheid), laten de resultaten van zijn studie zien, dat het hier om belangrijke oorzaken van lage scores zou kunnen gaan. Bovenstaande en andere items zijn voorgelegd aan enkele middle-class-kinderen van Europese komaf en aan lowerclass-kinderen van etnische minderheden (Blacks & Hispanics). De verschillen in scores tussen beide groepen bleken significant.

Hoewel er bij het doen van uitspraken potentiële linguïstische en culturele bronnen van itembias op basis van 'logisch'-inhoudelijke, c.q. theoretische overwegingen vooralsnog de nodige voorzichtigheid geboden is, kunnen er niettemin enkele eerste conclusies worden getrokken.

Allereerst moet worden geconstateerd dat het geen eenvoudige zaak is om itemmoeilijkheid, dat wil zeggen de gemeten vaardigheid, te scheiden van bias- en contextparameters. Ongetwijfeld speelt het feit dat gemeten vaardigheid niet adequaat genoeg te omschrijven valt daarbij een essentiële rol. Zo lang we immers niet precies weten welke (deel)vaardigheid door een item of door een set van items precies gemeten wordt, is het niet of nauwelijks mogelijk 'moeilijkheid' en 'onbedoelde moeilijkheid' van elkaar te scheiden. Daar komt bij dat, naarmate een itemcontext omvangrijker en (mede daardoor) complexer is, het steeds moeilijker wordt om op verantwoorde en adequate wijze bronnen van bias te isoleren, en bovendien dat een geconstateerde of plausibele bron van linguïstische of culturele bias in de ene context niet per definitie ook tot bias leidt in een andere context.

Wanneer er bij relatief eenduidige toetsrubrieken één specifiek taalvaardigheidsaspect aan de orde zou zijn, lijkt – wanneer tenminste volstrekt duidelijk is wat dat taalvaardigheidsaspect precies inhoudt – linguïstische itembias met meer kans op succes te kunnen worden geïsoleerd dan wanneer complexe rubrieken in het geding zijn, waarin meerdere taalvaardigheidsaspecten tegelijk aan bod komen. Met name bij rubrieken waarin de vaardigheid in tekstverwerking wordt gemeten, is het de vraag of het voorspellen van bronnen van itembias op dit moment überhaupt mogelijk is.

Het voorspellen en onderzoeken van linguïstische bronnen van itembias lijkt vooralsnog de meeste vruchten af te kunnen werpen bij toetsonderdelen als 'rekenen', omdat de gemeten vaardigheden bij dergelijke onderdelen primair op een heel ander terrein liggen dan taal. Bij onderdelen als 'taal' en 'informatieverwerking' ontstaan veel meer problemen en dilemma's.

De onzichtbaar blijvende en daardoor moeilijk vast te stellen (en te meten) betekenis-

niveaus van culturele aspecten en voorkennis spelen, evenals de verwevenheid van cultuur met vocabulairniveau en de semantische lading van woorden, een problematische rol bij het doen van uitspraken over de culturele lading van teksten en toetsitems.

Cultuurbepaalde toetservaring lijkt een niet te onderschatten rol te spelen bij het zoeken naar oorzaken van itembias. Een probleem hierbij is dat tot nu toe veel te weinig bekend is over (cultuurbepaalde) verschillen in redeneerprocessen en oplossingsstrategieën tussen (sub)groepen van toetsdeelnemers teneinde zinvolle voorspellingen omtrent culturele bronnen van itembias te kunnen doen. Daarnaast doen zich gelijksoortige problemen voor als bij het opsporen van linguïstische bronnen van itembias: de complexiteit van de gemeten vaardigheid en de verwevenheid ervan met zowel contexten als talige bronnen van itembias bemoeilijken het op adequate wijze onderscheiden van culturele bronnen van itembias.

Wanneer we de in deze bijdrage besproken factoren nog eens overzien, springt met name de rol die het vocabulaire lijkt te spelen in het oog. Daarbij gaat het niet alleen om het niveau van het vocabulaire, maar ook om woordkennis (woordenschat). Daarnaast blijkt ook de culturele lading van teksten nauw verbonden met (voor)kennis van woorden. Het is in dit verband dan ook niet toevallig dat in veel onderzoek voorkennis wordt gemeten met behulp van een woordenschattoets. Een te lage lexicale vaardigheid lijkt derhalve op uiteenlopende manieren te kunnen leiden tot het verkeerd interpreteren of het niet kunnen interpreteren van de taak in een toetsitem, ook al wordt de gemeten vaardigheid zelf misschien wel beheerst. Bovendien is het niet ondenkbaar dat ongewenst moeilijk vocabulaire gehanteerde oplossingsstrategieën kan beïnvloeden en sturen. Vanwege de centrale rol die woordenschat lijkt te spelen komt het ons wenselijk voor om verder onderzoek naar itembias in de Eindtoets (en in andere toetsen) voor allochtone (sub)groepen te richten op lexicale vaardigheden. Het accent zou daarbij in eerste instantie kunnen liggen op inhoudswoorden als belangrijkste betekenisdragers in zinnen en op de relatie tussen inhoudswoorden en functiewoorden. Daarbij dient bovendien rekening te worden gehouden met betekenisverschillen tussen woorden in alle-

daags taalgebruik en schooltaalgebruik. Op zins- en tekstniveau zien we, gezien de wetenschappelijke stand van zaken, vooraansnog weinig mogelijkheden om met kans op succes onderzoek naar itembias te doen.

### Literatuur

Appel, R., Mustafa op de MAVO; een exploratief onderzoek naar Nederlandse taalvaardigheid van anderstalige leerlingen in het voortgezet onderwijs. *Toegepaste Taalwetenschap in Artikelen*, 1985, 22, 65-72.

Aronowitz, R., Reading Tests as Texts. In: D. Tannen (Ed.), *Coherence in Spoken and Written Discourse* (pp. 245-265). Norwood, N.J.: Ablex, 1984.

Bartlett, E., Anaphoric Reference in Written Narratives of Good and Poor Elementary Writers. *Journal of Verbal Learning and Verbal Behavior*, 1984, 23, 540-552.

Boogaart, P. uit den, *Woordfrequenties in Geschreven en Gesproken Nederlands*. Utrecht: Oosthoek, Scheltema & Holkema, 1975.

Brière, E., Cross-cultural Biases in Language Testing. In: J. Oller & J. Richards (Eds.), *Focus on the Learner*. Rowley Mass.: Newbury House Publishers, 1973.

Broeder, P., J. Coenen, G. Extra, R. van Hout & R. Zerrouk, Ontwikkelingen in het Nederlandstalig Lexicon bij Anderstalige Volwassenen: een Macro- en Micro-Perspectief. In: J. Creten, G. Geerts & K. Jaspaert (Eds.), *Werk-in-Uitvoering. Momentopname van de Sociolinguïstiek in België en in Nederland* (pp. 39-59). Leuven: Acco, 1986.

Carpenter, P. & M. Just, Cognitive Processes in Reading. In: J. Orasanu (Ed.), *Reading comprehension from Research to Practice* (pp. 11-31). Hillsdale, N.J./London: L. Erlbaum, 1986.

Cito, *Doelenboek, Inhoudsverantwoording van de Eindtoets vanaf 1987*. Arnhem: Centraal Instituut voor Toetsontwikkeling, 1986.

Clark, H. & E. Clark, *Psychology and Language. An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich, 1977.

Coenen, M. & A. Vermeer, *Nederlandse woordenschat allochtone kinderen*. Tilburg: Zwijsen, 1988.

Ginkel, L. van, Woorden leren in een vreemde taal. *Toegepaste Taalwetenschap in Artikelen*, 1988, 32, 60-93.

Gerritsma, H., *Problemen bij Tekstbegrip van Allochtone Leerlingen*. Enschede: Stichting voor de Leerplanontwikkeling, 1987.

Glatt, B., Defining Thematic Progression and their Relationship to Reader Comprehension. In: M. Nystrand (Ed.), *What Writers Know. The Lan-*

*guage, Process and Structure of Written Discourse* (pp. 87-104). New York: Academic Press, 1982.

Hagendoorn, L., *Cultuur conflict en vooroordeel. Essays over de waarneming en betekenis van cultuurverschillen*. Alphen a/d Rijn/Brussel: Samsom, 1986.

Holland, P. & D. Thayer, *Differential Item Performance and the Mantel-Haenzel Procedure*. Paper Presented at the American Educational Research Association Annual Meeting, San Francisco, April 1986.

Jong, M. de, H. Uiterwijk, A. Kerkhoff & T. Vallen, *Vooronderzoek naar item- en testbias in de Eindtoets Basisonderwijs*. Tilburg: Faculteit Letteren (TILL-paper 115), 1987.

Johnson, P., Effects on Reading Comprehension of Language Complexity and Cultural Background of a Text. *Tesol Quarterly*, 1981, 15, 2, 169-182.

Johnston, P., Prior Knowledge and Reading Comprehension Test Bias. *Reading Research Quarterly* 1984, 19, 2, 219-240.

Kerkhoff, A., *Taalvaardigheid en schoolsucces*. Amsterdam/Lisse: Swets & Zeitlinger, 1988.

Kerkhoff, A. & T. Vallen, Cultural Biases in Second Language Testing of Children. In: G. Extra & T. Vallen (Eds.), *Ethnic Minorities and Dutch as a Second Language* (pp. 133-147). Dordrecht: Foris, 1984.

Kok, F., *Vraagpartijdigheid. Methodologische verkenningen*. Amsterdam: SCO, 1988.

Koppel, J. van de & N. Schoots 'Why are all Trains in Holland Painted Yellow?' Conventies in het Acculturatieproces. *Nederlands Tijdschrift voor de Psychologie*, 1986, 41, 189-196.

Levenston, E., Second language acquisition: issues and problems. *Interlanguage Studies Bulletin*, 1979, 4, 2, 147-160.

Maureau, J.H., *Goed en Begrijpelijk Schrijven. Een Analyse van 40 Jaar Schrijfadvisen*. (Tweede druk.) Muiderberg: Coutinho, 1979.

Meara, P., Vocabulary acquisition: a neglected aspect of second language acquisition. *Language Teaching and Linguistic Abstracts*, 1980, 13, 4, 221-246.

Noordman, L. & W. Vonk, Selectieve Verwerking van Tekst. *Tijdschrift voor Taal- en Tekstwetenschap*, 1987, 7, 1, 57-71.

Poortinga, Y., Crossculturele Psychologie en Minderhedenonderzoek. *De Psycholoog*, 1982, 17, 708-720.

Renkema, J., *Schrijfwijzer. Handboek voor Duidelijk Taalgebruik*. 's-Gravenhage: Staatsuitgeverij, 1982.

Richardson, K., Critical Linguistics and Textual Diagnosis. *Text*, 1987, 7, 2, 145-165.

Rinsampessy, E., De Etnische Factor in Molukse Families. In: E. Rinsampessy (Red.), *Emancipatie van Molukkers in het Onderwijs en het Welzijnswerk* (pp. 211-273). Wijchen: Pattimura, 1984.

- Sanders, T., *De Invloed van Globale Teksteigenschappen op het Begrijpen en Onthouden van Teksten*. Doctoraalscriptie KUB, Tilburg. Facultet der Letteren, 1986.
- Scheuneman, J., Item Bias and Individual Differences. In: S. Irvine (Ed.), *Human Assessment in Computer Context*. Den Haag: Nijhoff, 1988.
- Schuurs, U., De Perceptieve Ontwikkeling van Vraagzinnen in T1 en T2: Overeenkomsten en Verschillen. *Gramma*, 1984, 8, 2, 143-155.
- Shepard, L., Definitions of Bias. In: R. Berk (Ed.), *Handbook of Methods for Detecting Test Bias*. London: The Johns Hopkins University Press, 1982.
- Spilich, G., G. Vesonder, H. Chiesi & J. Voss, Text Processing of Domain-related Information for Individuals with High and Low Domain Knowledge. *Journal of Verbal Learning and Verbal Behavior*, 1979, 18, 275-290.
- Steehouder, M. & C. Jansen, *Leren Communiceren. Procedures voor Mondelinge en Schriftelijke Communicatie*. Groningen: Wolters-Noordhoff, 1979.
- Tennekes, J., Cultuur en Cultuurverschillen. *Migrantenstudies*, 1985, 1, 17-39.
- Tennekes, J., Nederland een Multiculturele Samenleving? *Migrantenstudies*, 1986, 3, 2-23.
- Verhoeven, L., *Ethnic Minority Children Acquiring Literacy*. Dordrecht: Foris, 1987.
- Verhoeven, L., Structural and Temporal Aspects of Turkish Children's Reading Dutch as a Second Language. In: G. Extra & T. Vallen (Eds.), *Ethnic Minorities and Dutch as a Second Language. Studies on Language Acquisition I* (pp. 65-86). Dordrecht: Foris, 1984.
- Vermeer, A., *Tempo en structuur van tweede-taalverwerving bij Turkse en Marokkaanse kinderen*. Tilburg: uitgave in eigen beheer.
- Vijver, F. van de & R. Joe, Cognitie en Perceptie in Cross-cultureel Perspectief. *Nederlands Tijdschrift voor de Psychologie*, 1986, 41, 157-164.
- Watson-Gegeo, K. & D. Gegeo, The Social World of Kwara'ae Children: Acquisition of Language and Values. In: J. Cook-Gumperz, W. Corsaro & J. Streeck (Eds.) *Children's Worlds and Children's Language* (pp. 109-129). Berlin/New York/Amsterdam: Mouton de Gruyter, 1986.
- Wijk, C. van & G. Kempen, Syntactische Formulvaardigheid en het Schrijven van Opstellen. *Pedagogische Studiën*, 1982, 59, 126-136.
- Wijk, C. van & G. Kempen, *The Origin of Developmental Patterns in Sentence Form: Syntactic, Conceptual or Rhetorical?* Nijmegen: University of Nijmegen (Dept. of Psychology), 1985.

### Curricula vitae

M. de Jong studeerde Tekstwetenschap en Taal & Minderheden aan de Letterenfaculteit van de KUB. Na haar afstuderen werkte ze bij het Werkverband Taal & Minderheden van de KUB aan een onderzoek naar itembias in de Eindtoets. Momenteel is zij werkzaam als Redacteur Informatica bij Kluwer Bedrijfswetenschappen te Deventer.

T. Vallen, zie *Pedagogische Studiën*, 1989, 66, p. 83

Adres: Katholieke Universiteit Brabant, Faculteit der Letteren, Postbus 90153, 5000 LE Tilburg

Manuscript aanvaard 12-6-'89

### Summary

Jong, M. de & T. Vallen. 'Linguistic and cultural sources of item bias for ethnic minority pupils in the CITO final primary schooltests.' *Pedagogische Studiën*, 1989, 66, 390-402.

In this contribution the question is addressed whether linguistic and cultural factors can be detected which may be responsible for the lower results of non-indigenous subgroups on the final primary school tests of the National Institute for Educational Measurement (CITO), in spite of the fact that the level of knowledge of these subgroups is not lower than the level of comparable indigenous subgroups. This phenomenon, which affects the construct validity of a test, is called 'item bias'. The possible sources of item bias that emerge from the literature have been applied to the items of the 1987 final primary school tests. It appears that it is very important to make a distinction between intended difficulty and unintended difficulty (item bias). Two things should be borne in mind here. First, a good test contains difficult items in order to discriminate between successful and less successful pupils. Second, the qualification 'biased' depends strongly on the kind of ability which is measured in an item: a difficult word may be intentionally difficult in a vocabulary item, but can be not-intentionally difficult in an item on arithmetic. It is not always clear either what exactly is the ability which is measured in an item and measured abilities appear to vary from simple to multiply complex. In addition, items are usually offered in a (culturespecific) context which may be unintentionally difficult in one situation, but intentionally so in another. What is especially striking is the role played by vocabulary when possible sources of bias are pinpointed. In order to make a success of the final tests a possession of the skills measured doesn't seem sufficient; besides a relatively high level of vocabulary a good knowledge of often implicit cultural aspects is required to be able to understand text elements with a high cultural load and to be able to interpret tasks correctly.