



Cito Instituut voor Toetsontwikkeling

Richtlijnen voor de ontwikkeling van onpartijdige toetsen

K. Bügel
P.F. Sanders

Cito

**RICHTLIJNEN VOOR DE ONTWIKKELING
VAN ONPARTIJDIGE TOETSEN**

K. Bügel

P.F. Sanders



Auteurs: P.F. Sanders (links) en K. Bügel

© Cito, Arnhem, 1998

Het is toegestaan deze tekst geheel of gedeeltelijk over te nemen, mits u de volgende bron vermeldt: Bügel, K. en P.F. Sanders, Richtlijnen voor de ontwikkeling van onpartijdige toetsen. Cito, Arnhem, 1998.

INHOUD

1	Inleiding	1
2	Begrippen en procedures	3
2.1	Toetsbias en itembias	3
2.2	Detectie van toetsbias en itembias	3
2.3	Kwetsende inhoud	4
3	Toetsinhouden die bias kunnen veroorzaken	5
3.1	Verschillen in achtergrondkennis en cultuur	5
3.1	Verschillen in taalkennis	7
4	Kwetsende inhoud	13
4.1	Negeren	13
4.2	Etnocentrische inhoud	15
4.3	Negatieve inhouden en stereotypen	17
5	Maatregelen in verschillende fasen van het proces van toetsconstructie en toetsanalyse	20
5.1	Maatregelen vooraf	20
5.2	Toetsspecificatie	20
5.3	Itemconstructie	22
5.4	Samenstellen van de toets	24
5.5	Toetsafname	24
6	Literatuur	26
	Checklist	28

1 Inleiding

De samenstelling van de Nederlandse maatschappij verandert en daarmee ook de samenstelling van de klassen in het Nederlandse onderwijs. Een toenemend aantal leerlingen is van niet-Nederlandse afkomst en spreekt thuis vaak een andere taal dan het Nederlands. Goed onderwijs is een van de belangrijkste middelen tot integratie en emancipatie van deze leerlingen. Sinds enige jaren worden dan ook aanzienlijke inspanningen verricht om het onderwijs ook af te stemmen op allochtone leerlingen. Er is onderwijs in Nederlands als tweede taal, er zijn schakelklassen om de doorstroming naar het reguliere onderwijs te bevorderen, enz. De problemen die allochtone leerlingen ondervinden zijn overigens voor een deel dezelfde en voor een deel vergelijkbaar met de problemen van leerlingen uit sociaal zwakke milieus.

Vergeleken met de problemen die allochtone leerlingen vaak ondervinden op school, zijn de problemen voor meisjes gering. Toch zijn er nog een aantal hardnekkige verschillen die in het nadeel van meisjes werken: vrouwelijke examenkandidaten halen gemiddeld lagere cijfers bij de eindexamens, vooral bij de maatschappijgerichte en exacte vakken. Bovendien beperkt de vakkenkeuze van meisjes hun latere studie- en beroepsmogelijkheden.

De in sommige opzichten achterblijvende prestaties van allochtone en vrouwelijke leerlingen hebben in een aantal gevallen geleid tot een beschuldigende vinger in de richting van het onderwijs. Het onderwijs zou niet voldoende bijdragen aan het verkleinen van de verschillen, maar deze juist in stand houden. Hetzelfde verwijt wordt ook wel gemaakt ten aanzien van de toetsen en examens die door het onderwijs gebruikt worden.

Onderzoek naar de vraag of de geringere toetsprestaties van allochtone en vrouwelijke leerlingen wellicht het gevolg zijn van onbedoelde kenmerken van de items of opgaven heeft ook in Nederland plaatsgevonden. Uit dat onderzoek bleek dat sommige opgaven moeilijker waren voor allochtone dan voor autochtone leerlingen als gevolg van hun andere culturele en linguïstische achtergrond. Ook bleek dat sommige opgaven moeilijker waren voor meisjes dan voor jongens als gevolg van hun andere interesses en achtergrondkennis. Het verschijnsel dat een verschil in moeilijkheidsgraad van een opgave voor verschillende groepen leerlingen veroorzaakt wordt door aspecten van de opgave die niet relevant zijn voor wat de opgave beoogt te meten, wordt in de testliteratuur 'itembias' genoemd.

Doel van de Richtlijnen

De Richtlijnen hebben in de eerste plaats tot doel er voor te zorgen dat de toetsen die door het Cito ontwikkeld worden geen opgaven met itembias bevatten. Daarnaast beogen de Richtlijnen te voorkomen dat de toetsen opgaven met kwetsende inhoud bevatten. De Richtlijnen kunnen zowel gebruikt worden bij de constructie van nieuwe toetsen als voor de beoordeling van bestaande toetsen.

In de Richtlijnen wordt de term onpartijdige toetsen gehanteerd voor toetsen die noch opgaven met bias noch opgaven met kwetsende inhoud bevatten. Dat de Richtlijnen met name betrekking hebben op allochtone leerlingen en meisjes heeft vooral te maken met de omvang van de betrokken groepen. Overigens zijn de meeste suggesties die gedaan worden op het gebied van taalgebruik ook van toepassing op leerlingen uit sociaal zwakke milieus.

Opbouw van de Richtlijnen

De Richtlijnen zijn als volgt opgebouwd. In het tweede hoofdstuk worden de begrippen toetsbias en itembias gedefinieerd en worden procedures voor de detectie van bias beschreven. Inhoudelijke kenmerken van opgaven die itembias kunnen veroorzaken worden in hoofdstuk drie beschreven, terwijl hoofdstuk vier een overzicht bevat van inhoudelijke kenmerken van opgaven die kwetsend kunnen zijn voor bepaalde groepen leerlingen. In hoofdstuk vijf worden maatregelen beschreven die in verschillende fasen van het toetsconstructieproces genomen kunnen worden om partijdigheid te voorkomen. Tot slot is een checklist opgenomen waarmee de partijdigheid van toetsen beoordeeld kan worden.

2 Begrippen en procedures

In dit hoofdstuk van de *Richtlijnen voor de ontwikkeling van onpartijdige toetsen* worden de belangrijkste begrippen en procedures voor de detectie van bias en kwetsende inhoud kort toegelicht. Dat de *Richtlijnen* zich vooral op etnische of culturele partijdigheid en sekse-partijdigheid richten, heeft te maken met het feit dat met name deze vormen van partijdigheid veel voorkomen en dus ook relatief grote aantallen leerlingen treffen.

2.1 Toetsbias en itembias

In de testliteratuur wordt de Engelse term 'bias' gebruikt voor het verschijnsel dat prestatieverschillen tussen onderscheiden groepen leerlingen, bijvoorbeeld allochtone en autochtone leerlingen, op toetsen als geheel of op afzonderlijke opgaven niet veroorzaakt worden door verschillen in vaardigheid maar door factoren die niet relevant zijn voor wat met de toets of opgaven beoogd wordt te meten. Onder opgave wordt een onderdeel van een toets verstaan. Een opgave bevat altijd een stimulus, bijvoorbeeld een vraag, waarop een respons (het uitgelokte gedrag) mogelijk moet zijn. Een toets bestaat uit een aantal bij elkaar behorende opgaven plus een voorschrift voor het bepalen van de toetsscore. Een toets is een instrument voor het meten van iemands kennis, houdingen, inzichten en vaardigheden.

Er is sprake van toetsbias als op grond van een toets als geheel de vaardigheid van een bepaalde groep leerlingen ten onrechte systematisch te hoog of te laag wordt ingeschat. Toetsbias impliceert dat de toets niet dezelfde vaardigheid bij de onderscheiden groepen leerlingen meet.

Er is sprake van itembias als verschillen in prestaties op een item of opgave veroorzaakt worden door kenmerken van de opgave die niet relevant zijn voor wat de opgave beoogt te meten. Itembias impliceert dat de opgave niet dezelfde vaardigheid bij de onderscheiden groepen leerlingen meet.

2.2 Detectie van toetsbias en itembias

Voor het opsporen van toetsbias moeten de prestaties van verschillende groepen leerlingen op de te onderzoeken toets vergeleken worden met een relevant criterium: een andere toets of een andersoortige beoordeling. Mogelijke toetsbias van de Eindtoets Basisonderwijs kan onderzocht worden door de prestaties van bijvoorbeeld meisjes en jongens te vergelijken met het rapportcijfer aan het einde van de brugklas van het voortgezet onderwijs. Indien voor meisjes en jongens met vergelijkbare prestaties op de Eindtoets, het gemiddeld rapportcijfer voor meisjes in de brugklas hoger zou blijken te zijn dan dat voor jongens, dan zou dit op mogelijke toetsbias van de Eindtoets kunnen wijzen. Bij het onderzoek naar toetsbias geldt echter een belangrijke conditie, namelijk dat het criterium, in het voorbeeld het gemiddeld rapportcijfer, vrij van bias is.

De methoden waarmee men itembias kan opsporen, berusten alle op hetzelfde principe. De groepen leerlingen die men wil vergelijken, bijvoorbeeld allochtone en autochtone leerlingen, worden opgesplitst in subgroepen die nagenoeg even vaardig zijn, dat wil zeggen nagenoeg dezelfde toetsscore behaald hebben. Voor elk item gaat men vervolgens na of de gemiddelde itemscore van de even vaardige subgroepen allochtone en autochtone leerlingen al of niet gelijk is. Indien, bijvoorbeeld op een rekenopgave, de

prestaties van allochtone leerlingen die even rekenvaardig zijn als autochtone leerlingen, statistisch significant verschillen, dan functioneert die opgave blijkbaar anders voor autoctonen dan voor alloctonen. In de testliteratuur wordt dit verschijnsel aangeduid met de term 'differential item functioning', wat veelal afgekort wordt als DIF (Holland en Wainer, 1993).

Teneinde te kunnen concluderen dat er sprake is van itembias, moet men nagaan waardoor de verschillen in prestaties tussen de groepen leerlingen veroorzaakt kunnen zijn. Er is sprake van itembias wanneer de verschillen veroorzaakt worden door aspecten van de opgave die niet relevant zijn voor wat men met de opgave wil meten. Bijvoorbeeld, als blijkt dat de prestatieverschillen van even vaardige Turkse en Nederlandse leerlingen bij een rekenopgave veroorzaakt worden door het feit dat sommige Turkse leerlingen de opdracht bij de opgave niet begrijpen. Bij de Turkse leerlingen wordt dan kennelijk ook taalkennis getoetst.

Het voorgaande maakt duidelijk dat onderzoek naar toetsbias en itembias begint met statistisch onderzoek en daarna gevolgd wordt door inhoudelijk onderzoek. Het is gebleken dat bias vaak veroorzaakt wordt door inhoudelijke aspecten van toetsen, zoals contextmateriaal of het onderwerp van de opgave (zie o.a. Scheuneman en Gerritz, 1990). Contextmateriaal wordt ook wel uitgangsmateriaal genoemd. Hiermee worden krantenartikelen en dergelijke genoemd die de leerlingen moeten gebruiken bij het maken van een opgave. Meisjes en jongens hebben verschillende interesses en hun kennis verschilt met betrekking tot een aantal kennisdomeinen, zoals techniek, menselijke relaties en kunst (Bügel, 1993). Het maakt dan ook verschil uit voor de leerling of het uitgangsmateriaal een tekst over een stoommachine of over vermageringsdiëten gaat. Leerlingen van verschillende etnische afkomst kunnen verschillende opvattingen, attitudes en ervaringen hebben ten aanzien van sommige onderwerpen. De manier waarop ouders en kinderen met elkaar omgaan verschilt per cultuur.

2.3 Kwetsende inhoud

Inhoudelijk onderzoek is ook om andere redenen noodzakelijk. Men moet altijd rekening houden met het feit dat een toets voor verschillende groepen leerlingen bedoeld is. De inhoud van toetsen mag niet beledigend zijn voor bepaalde groepen, vermeden moet worden dat de inhoud heftige emoties oproept bij sommige groepen kandidaten of dat sommige groepen leerlingen zich buitengesloten of gekleineerd voelen. Dergelijke gevoelens kunnen een goede inzet en motivatie tegengaan of op een andere manier de prestaties van de leerlingen beïnvloeden. Zo dienen in de toets niet uitsluitend mannen of vertegenwoordigers van de etnische meerderheid voor te komen. Etnocentrische zinsneden waarbij de eigen groep of cultuur als superieur of universeel wordt afgeschilderd dienen te worden vermeden evenals stereotyperende beschrijvingen van (leden van) minderheden of van een bepaalde sekse.

Om te beoordelen of toetsen kwetsende inhoud bevatten, moeten ze beoordeeld worden door vertegenwoordigers van de betrokken groepen.

3 Toetsinhouden die bias kunnen veroorzaken

Wanneer bepaalde items bias vertonen, is de oorzaak vaak niet zonder meer duidelijk. Soms is het moeilijk een verklaring te vinden. Toch is er wel iets bekend over de oorzaken van bias. Onderzoek heeft aangetoond dat bias veroorzaakt kan worden door verschillen in achtergrondkennis en cultuur en door verschillen in taalvaardigheid van de kandidaten.

3.1 Verschillen in achtergrondkennis en cultuur

Kandidaten kunnen verschillende ervaringen en hoeveelheden kennis hebben ten aanzien van een aantal onderwerpen. Jongens zijn bijvoorbeeld veel vaker lid van een voetbalclub dan meisjes. Jongens weten dan ook gemiddeld meer over voetbal dan meisjes. Verschillen in achtergrondkennis behoren tot de meest genoemde oorzaken van itembias (Langer, 1984; Johnston, 1984, e.a.). Nederlandse kinderen kennen meer Nederlandse begrippen en situaties dan kinderen van niet-Nederlandse afkomst. Het is bijvoorbeeld niet zeker dat alle leerlingen weten wat B.T.W. is, een begrip dat voorkwam in een rekenopgave van de Eindtoets Basisonderwijs (zie voorbeeld 13). Leden van etnische minderheidsgroepen hebben soms andere ervaringen, gewoontes en dergelijke dan de meerderheid van de Nederlandse bevolking. Turkse kinderen kijken bijvoorbeeld vaak naar Turkse televisiezenders. Ook kunnen er verschillen zijn in cultuur die leiden tot andere waarden en normen. Hierdoor kunnen sommige gebeurtenissen, beschrijvingen of woorden een verschillende betekenis hebben voor leden van verschillende etnische groepen (Steffensen, Joag-Dev & Anderson, 1978; Dais, 1993). De kans op bias is groter naarmate de achtergrond en het waardepatroon van de kandidaten meer verschillen van die van de toetsconstructeurs.

De invloed van achtergrondkennis, waaronder culturele kennis, op tekstbegrip wordt door veel onderzoek ondersteund (Voor een overzicht van de literatuur met betrekking tot de invloed van achtergrondkennis, zie Bügel K., 1993). Kerkhoff en Vallen (1984) lieten bijvoorbeeld zien dat de culturele gekleurdheid van teksten invloed heeft op de toetsprestaties van Nederlandse, Turkse en Molukse kinderen. Bügel (1993) toonde aan dat het onderwerp van teksten een grote rol speelt bij sekseverschillen in tekstbegrip moderne vreemde talen. Meisjes hebben gemiddeld andere interesses en tijdsbestedingen dan jongens. Ook verschilt hun kennis over sommige onderwerpen. Meisjes weten gemiddeld meer over menselijke relaties, verzorging, huishoudelijke zaken, kunst en filosofie. Jongens weten meer over sport, economische, technische en politieke onderwerpen. Het zal duidelijk zijn dat bias door verschil in achtergrondkennis niet altijd te voorkomen is. Sommige achtergrondkennis is nu eenmaal meer relevant voor kennis die op school geleerd wordt dan andere.

Aanbeveling

Men dient bij de toetsconstructie rekening te houden met het feit dat de achtergrondkennis van de kandidaten kan verschillen. Toetsen dienen de diversiteit in achtergrondkennis en ervaringen te reflecteren van de kandidaten voor wie de toets bedoeld is. De inhoud van het contextmateriaal moet niet zo specifiek zijn dat deze bij bepaalde groepen kandidaten volledig onbekend is. Ook mag de opgave geen informatie bevatten die specifiek is voor één bepaalde cultuur als de toets bestemd is voor kandidaten uit verschillende culturen. Deze aspecten kunnen beoordeeld worden door experts afkomstig uit verschillende culturen.

Voorbeeld¹⁾

Het volgende item zou bias kunnen veroorzaken door een verschil in ervaringen tussen Nederlandse leerlingen en leerlingen met een andere nationaliteit.

- (1) Els mag met haar tante op vakantie naar Spanje. Daarvoor heeft zij een paspoort nodig. Waar kan zij dat paspoort krijgen?
- A bij de Spaanse grens
 - B* bij het gemeentehuis van haar woonplaats
 - C bij het hoofdbureau van politie
 - D bij het reisbureau dat haar reis verzorgt.

Toelichting

In dit item wordt er kennelijk van uitgegaan dat Els de Nederlandse nationaliteit heeft. Voor buitenlandse kinderen is de situatie anders. Voor hun situatie is er geen goed antwoord. Het kan zijn dat de kennis die met dit item getoetst wordt wel in overeenstemming is met het toetsdoel, bijvoorbeeld 'de taken van de gemeentelijke instanties kennen'. In dat geval zou beter aan de vraag kunnen worden toegevoegd dat Els de Nederlandse nationaliteit heeft.

Voorbeeld

Item 2 bevoordeelt waarschijnlijk leerlingen die katholiek of protestant zijn.

- (2) Hieronder staan een aantal feestdagen in Nederland.
Welke feestdagen hebben te maken met het christelijk geloof?
(Genoemd worden: Nieuwjaar, Pasen, Koninginnedag, Bevrijdingsdag, Moederdag, Hemelvaartdag, Pinksteren, Dierendag, Kerstmis)

Toelichting

De christelijke feesten zullen niet bij alle groepen in Nederland even bekend zijn. Bij het beoordelen van dit soort items zou men moeten nagaan of de kennis die er aan de orde komt een goede uitwerking is van de domeinen die genoemd worden in de toetspecificatie, met andere woorden of deze kennis noodzakelijk is voor de vaardigheid die men wil toetsen. In dit geval ging het om 'Wereldoriëntatie'.

Het volgende item zal wellicht niet tot itembias leiden, maar zou toch beter anders geformuleerd kunnen worden.

- (3) Vader heeft 12 flessen wijn gekocht. Elke fles bevat driekwart liter wijn. Hoeveel liter heeft hij nu?

Toelichting

Het gebruik van 'vader' in combinatie met 'wijn' maakt dit zinnetje misschien minder acceptabel voor islamitische leerlingen. Gezien het geringe aantal vrouwelijke personages in de toets waar dit item in voorkomt, zou een vrouw bovendien beter passen. Mogelijke verbetering:

Paula van den Berg heeft 12 flessen wijn (of: vruchtensap) gekocht, etc.

Noot ¹ Alle voorbeelden die in deze Richtlijnen voorkomen, zijn ontleend aan bestaande toetsen, tenzij vermeld wordt dat het voorbeeld fictief is.

3.2 Verschillen in taalkennis

Allochtone leerlingen beheersen de Nederlandse taal vaak minder goed dan autochtone leerlingen (Verhallen-van Ling, 1994). Hierdoor ontstaan vaak problemen voor allochtone leerlingen. Taal is namelijk niet alleen doel maar ook middel. Leerlingen leren taal op school, maar taal is ook het middel om kennis te verwerven. Wanneer leerlingen de taal die op school gebruikt wordt onvoldoende beheersen, kunnen ze ook minder leren. Een soortgelijk probleem doet zich voor bij toetsen. Bij alle toetsen wordt taal gebruikt. Soms is dit beperkt tot de vraagstelling. In andere gevallen wordt extra informatiemateriaal toegevoegd. In dat geval spreekt men van 'contexten'. Dit zijn teksten, tekeningen, tabellen of andere informatie die men aan de eigenlijke vraagstelling toevoegt om een situatie te schetsen die de vraag aannemelijk moet maken, om de kandidaat zelf het probleem te laten vinden in een bepaalde situatie of om andere redenen.

'Talige' contexten

Wanneer 'talige' contexten worden gebruikt, blijkt een grotere kans op bias te bestaan. Niet alleen vanwege de verschillen in kennis en interesse van de leerlingen ten aanzien van die contexten, maar ook door verschillen in taalkennis en leesvaardigheid van de kandidaten. In dat geval kan bias optreden, wanneer taalkennis of leesvaardigheid niet het onderwerp van de toets is. Wees daarom zeer kritisch bij het gebruik van talige contexten wanneer geen taalvaardigheid getoetst moet worden.

Echter, ook bij taaltoetsen kan DIF en soms bias ontstaan omdat bepaalde taalonderdelen een verschillende moeilijkheidsgraad kunnen hebben voor leerlingen voor wie Nederlands de moedertaal is en voor leerlingen voor wie dat niet het geval is. Dit geldt bijvoorbeeld voor het gebruik van spreekwoorden, vaste uitdrukkingen, woordgrapjes en beeldspraak.

Een dergelijk probleem bij een taalitem kan geïllustreerd worden aan de hand van een item uit de Eindtoets Basisonderwijs dat na analyse relatief veel DIF bleek te vertonen:

- (4) Wat kun je het beste doen met: *Gelukkig was het goed met de wind...*(r. 4, 5)?
- A Zo laten staan
 - B* Vervangen door: Gelukkig hadden we de wind mee,...
 - C Vervangen door: Gelukkig hadden we goede wind,...
 - D Vervangen door: Gelukkig was de wind met ons,...

Toelichting

Zowel de aangeboden zin als de aangeboden verbeteringen in B, C en D zijn grammaticaal correct en betekenen hetzelfde. Om het item te kunnen beantwoorden moet men de uitdrukking 'de wind mee hebben' kennen en bovendien weten dat 'goede wind hebben' en 'de wind met zich hebben' geen bestaande uitdrukkingen zijn in het Nederlands. Dit vereist een grote kennis van het Nederlands, die grotendeels buiten de school zal zijn verkregen. Leerlingen die minder mogelijkheden hebben gehad om buiten de school Nederlands te leren, worden hierdoor benadeeld. Dit betekent niet dat een dergelijke vraag niet gesteld mag worden. Kennis van gezegden en spreekwoorden kan onderdeel uitmaken van een toetsspecificatie. Men moet er echter wel voor zorgen dat de kennis en vaardigheden die men toetst, op school onderwezen worden.

Onderwijsniveau

Natuurlijk is het zo dat het taalgebruik ook bepaald wordt door het onderwijsniveau. De eisen die gelden voor het taalgebruik op de basisschool zijn anders dan die voor het vwo. De taal die gebruikt wordt in toetsen bestemd voor basisschooll leerlingen dient zo eenvoudig mogelijk te zijn. Op de basisschool zitten bovendien relatief meer allochtone leerlingen dan op havo en vwo. Onderzoek naar het functioneren van etnische minderheden is ook vaak toegespitst op de basisschool. Het ligt echter in de verwachting dat het aandeel van leerlingen van buitenlandse afkomst in het vervolgonderwijs in de toekomst sterk zal toenemen. Er zou daarom meer onderzoek gedaan moeten worden naar partijdigheid van toetsen ten aanzien van deze leerlingen.

Ook op 'hogere' schooltypen kan men er niet zonder meer van uitgaan dat de taalvaardigheid van alle leerlingen hoog is. Leerlingen die op latere leeftijd naar Nederland komen en in eigen land een relatief hoog schooltype bezochten of allochtone vwo-leerlingen die thuis geen Nederlands spreken, hebben over het algemeen een aanzienlijke geringere kennis van de Nederlandse taal dan Nederlandse leerlingen. Door gebrek aan onderzoek op dit gebied met betrekking tot de Nederlandse situatie, is het vaak niet mogelijk bias te voorkomen.

Aanbevelingen

Schoolkennis

Afsluitende toetsen moeten vooral gaan over zaken die op school geleerd zijn en niet de nadruk leggen op kennis en vaardigheden die buitenschools verkregen zijn.

Functie van contexten

Men dient zich bij alle toetsen af te vragen wat de functie is en wat het effect zou kunnen zijn van de zinnen of de tekstjes die worden gebruikt of die worden toegevoegd aan de eigenlijke vraagstelling. Zijn de contexten nodig en zijn ze functioneel? Wanneer men context- of uitgangsmateriaal gebruikt bij de vraagstelling, heeft men minder controle over de vaardigheden die men toetst. Alleen al bij tekstbegrip spelen verschillende vaardigheden een rol, zoals achtergrondkennis, eigen opvattingen en taalkennis.

Redundantie en illustraties

Dit betekent overigens niet dat vraagstelling en eventuele situatieschets zo beknopt mogelijk moeten zijn. Voor minder taalvaardige leerlingen is redundantie van groot belang. Zorg ervoor dat het taalmateriaal zo expliciet mogelijk is en waar mogelijk redundant. Probeer 'talige' context te vervangen of aan te vullen door een foto, een tekening of een grafiek. Wanneer het niet duidelijk is wat de functie is van de toegevoegde context, laat deze dan weg.

Voorbeelden

- (5) Opa is zuinig met water. Hij verbruikte dit jaar maar 9 m³. Hoeveel liter is dat?
- A 9 liter
 - B 90 liter
 - C 900 liter
 - D* 9000 liter

- | | |
|-----|--|
| (6) | Drie jongens willen twee bananen eerlijk verdelen. Hoeveel moet elke jongen dan krijgen? |
| A | 1/3 banaan |
| B | 1/4 banaan |
| C* | 2/3 banaan |
| D | 1 1/2 bananen |

Toelichting

In het eerste geval (5) is het niet duidelijk wat de functie is van de inleidende zin ten aanzien van het gevraagde. De tekst is eerder verwarrend. De indruk wordt gewekt dat de leerlingen moeten inschatten wat 'weinig' is, waardoor ze wellicht A kiezen. Een neutrale vraag zou zijn:

Hoeveel liter is 9m³ water?

Bij het tweede voorbeeld (6) kan men zich wel voorstellen wat de functie van de tekst is. Leerlingen moeten hier een probleem oplossen in een relevante context. Deze vraag toetst iets anders dan de som: $2:3 = \dots$. Om te kunnen beoordelen of de toevoeging van de context zinvol is, moet men echter wel het toetsdoel kennen.

Begrijpelijkheid

Wanneer men heeft vastgesteld dat de talige elementen in het item noodzakelijk zijn, dient er een tweede afweging plaats te vinden. Zijn de woorden, zinnen en uitdrukkingen die gebruikt worden voor iedereen begrijpelijk? Allochtone leerlingen hebben vaak een minder grote kennis van de Nederlandse taal. Ze kennen niet alleen minder grammaticale structuren, maar vooral ook minder woorden.

Aanbeveling

Bij vakken als rekenen of aardrijkskunde, doet men er goed aan zoveel mogelijk dezelfde termen te gebruiken als in schoolboeken en examenprogramma's. Bij de vraagstelling dient men te streven naar uniformiteit en duidelijkheid. Probeer het woordgebruik in de vraagstelling zoveel mogelijk te standaardiseren. Dus niet bij rekenopgaven woorden als uitkomst, oplossing, resultaat, antwoord e.d. door elkaar gebruiken, als men er hetzelfde begrip mee bedoelt. Dit kan voor iemand die niet alle nuances van het Nederlands beheerst, zeer verwarrend zijn. Vermijd variatie om stijlredenen.

Vergelijk bijvoorbeeld de volgende twee vragen uit de Eindtoets Basisonderwijs

- | | |
|-----|--|
| (7) | Wat had Callista ook kunnen schrijven voor 'niet te laat'? |
| (8) | Wat had Casper in plaats van 'te kalmeren' ook kunnen gebruiken? |

Toelichting

Met beide vragen wordt hetzelfde bedoeld, maar door de verschillende formuleringen wordt de suggestie gewekt dat dit niet zo is.

De volgende mogelijke hinderpalen voor minder taalvaardige leerlingen worden in de literatuur genoemd (Coenen en Vallen, 1991; De Jong en Vallen, 1989).

- 1 Woordgebruik: veel allochtone leerlingen hebben een Nederlandse woordenschat die aanzienlijk geringer is dan die van leerlingen van Nederlandse afkomst van

dezelfde leeftijd. Deze leerlingen kennen vaak niet de precieze betekenis van minder frequente woorden en functiewoorden. Functiewoorden zijn woorden die in een tekst de logische verbanden en de verbanden tussen zinnen en zinsdelen weergeven, zoals voegwoorden (bijvoorbeeld 'omdat', 'want', 'indien') en woorden als 'voorts', 'daarna', 'tenslotte'.

- 2 Voegwoorden die bijna alleen in de schrijftaal voorkomen, zoals 'ofschoon' en 'mits' en voegwoorden die uit twee delen bestaan zoals 'noch ...noch' en 'zowel...als' zijn vaak onbekend.
- 3 Verwijswoorden: verwijzingen met voornaamwoorden zijn vaak problematisch, vooral ook in de volgende gevallen:
 - Als inhoudelijk of grammaticaal niet duidelijk is naar welk woord verwezen wordt, zoals in de volgende itemstam:

Voorbeeld

- (9) Carla's ouders zijn gescheiden. Carla's moeder heeft geen baan en het geld dat **ze** iedere maand van Carla's vader krijgen is veel te weinig om van te leven.

- Wanneer het verwijswoord getalsmatig afwijkt van de referent: '**De familie Van Zuijlen** ging naar Amsterdam. **Ze** namen de trein.'
 - Als de afstand tussen het verwijswoord en de referent erg groot is. Bijvoorbeeld: Dit probleem is al vakeresignaleerd in de dierentuin. Ook olifanten en giraffen hebben er last van. **Het** doet zich trouwens ook wel voor bij leden van een apenfamilie.
 - Als het verwijswoord vooruit wijst en niet terug zoals meestal het geval is. Bijvoorbeeld: De zaak is **deze**: sommige leerlingen snappen zulke zinnen niet.
- 4 Voorzetsels en voorzetseluitdrukkingen: ten aanzien van, met betrekking tot, door middel van.
 - 5 Werkwoorden met een vast voorzetsel zoals: voorkomen ('menseneters **komen** hier niet **voor**'), afnemen, toenemen.
 - 6 Samengestelde woorden: 'milieuveranderingstechnieken'. 'gewichtsaanduiding'.
 - 7 Lange, samengestelde zinnen.

Aanbeveling

Bij het formuleren van het item, moet men nagaan of de woorden die men gebruikt in het onderwijs geleerd zijn of bekend zijn bij de kandidaten. Vermijd taalconstructies en woorden die begripsmoeilijkheden kunnen veroorzaken. Wanneer men de keuze heeft uit twee synoniemen, kies dan het meest frequent voorkomende woord. Raadpleeg eventueel frequentielijsten (bijvoorbeeld het *Basiswoordenboek Nederlands* (De Kleijn & Nieuwborg, 1987). Wanneer men in de vraagstelling vaktermen gebruikt die niet tot het curriculum behoren, moeten die termen bij iedereen bekend zijn.

Uiterwijk (1994) geeft een aantal voorbeelden van items die DIF vertonen in het nadeel van Turkse en Marokkaanse leerlingen:

.	(10) Welke twee getallen liggen even ver van 1 af?
.	A 0,99 en 1,99
.	B 1,01 en 1,10
.	C* 0,9 en 1,1
.	D 0,95 en 1,5
.	(11) Op de Arkschool is 1 van elke 2 kinderen lid van een club. Hoeveel procent is dat?
.	A 1/2%
.	B 33 1/3%
.	C* 50%
.	D 100%
.	(12) Iemand moet 3 km wandelen. Zij gaat na hoelang zij doet over honderd meter. Dat duurt ongeveer 1 minuut. Hoelang zal zij ongeveer onderweg zijn?
.	A 3 minuten
.	B 20 minuten
.	C* 30 minuten
.	D 50 minuten

Toelichting

Zoals ook Uiterwijk (1994) opmerkt, is niet met zekerheid te zeggen wat bij deze items de oorzaak van de DIF is. Wel valt op dat in alle drie de items constructies met voorzetsel voorkomen die moeilijk zouden kunnen zijn voor leerlingen voor wie Nederlands niet de moedertaal is: in het eerste item 'liggen even ver af', in het tweede '1 van elke 2 kinderen' en in het derde: 'Zij gaat na' en misschien ook 'onderweg zijn'. Omdat het hier om rekenopgaven gaat, is het belangrijk ervoor te zorgen dat de moeilijkheid van het item niet bepaald wordt door het taalgebruik in de vraagstelling.

Onduidelijke vraagstelling

Wanneer een vraag voor tweeërlei uitleg vatbaar is, is itembias soms onvermijdelijk. De ene leerling kan de opdracht anders opvatten dan de andere, waardoor verschillende antwoorden gegeven kunnen worden die een beroep doen op verschillende vaardigheden. In zijn algemeenheid valt niet te zeggen welke groep leerlingen hierdoor benadeeld wordt. Meestal de minst taalvaardige.

Voorbeeld

(Bij dit item staat een plaatje van een man die tot zijn knieën in het water staat)

(13) Iedere zes uur staat dit gebied onder water.

Bij welk water kun je dit zien?

- A de Maas
- B* de Waddenzee
- C het Noordzeekanaal
- D Het IJsselmeer

Toelichting

Onduidelijk is wat met 'dit gebied' bedoeld wordt. Ook 'iedere zes uur' roept vragen op: Wat wordt bedoeld: 'om de zes uur?' of 'zes uur lang'? Eerst wordt over een gebied gepraat, dan over een water. En ten slotte: waar verwijst 'dit' in de vraag naar? Naar 'dit gebied'? of naar het onder water staan van het gebied? Of naar de man op de foto die onder water staat?

Impliciet taalgebruik

Minder taalvaardige leerlingen hebben vaak moeite met impliciet taalgebruik en kunnen hierdoor een vraag verkeerd interpreteren. Dit komt vooral voor bij sommige groepen allochtone leerlingen die minder vertrouwd blijken te zijn met Nederlandse toetsconventies en daarom aan een half woord niet genoeg hebben.

Aanbeveling

Formuleer de vraag zo expliciet mogelijk. De vraagstelling moet duidelijk en ondubbelzinnig zijn. Het moet voor de kandidaten helder zijn wat er van hen verwacht wordt.

Voorbeeld

Het volgende rekenitem (14) bleek voor allochtone leerlingen gemakkelijker te worden na de vervanging van een moeilijk woord en het expliciteren van de vraag:

(14) 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.?

Vereenvoudiging van de laatste zin (in het onderzoek):

Wat moet vader voor de naaimachine betalen met B.T.W.?

Toelichting

Een meer realistische en daardoor wellicht nog eenvoudiger te begrijpen context zou overigens kunnen zijn (waarbij van de veronderstelling wordt uitgegaan dat het berekenen van de procenten essentieel is en niet de optelling):

Anita Sprengers koopt een naaimachine. Deze naaimachine kost normaal f 800,-. Omdat het uitverkoop is, krijgt zij 20% korting. Hoeveel moet Anita voor de naaimachine betalen?

Deze vraag zou met een tekening verduidelijkt kunnen worden: een etalage met de genoemde naaimachine en het opschrift '20% korting'. Dit hangt af van het niveau en van de doelgroep.

4 Kwetsende inhoud

De inhoud van toetsen mag niet kwetsend zijn voor kandidaten. Er kunnen verschillende oorzaken zijn waardoor de inhoud van een toets als beledigend wordt ervaren door leden van etnische minderheden of door andere groepen kandidaten. Vaak worden hierbij de volgende gevallen onderscheiden.

4.1 Negeren

In veel toetsen komen alleen blanke Nederlanders voor en wordt bijna uitsluitend gerefereerd aan mannelijke personages, zoals 'Jan de Wit' en 'Ondernemer J.E. Ans'. Dit laatste komt waarschijnlijk doordat mensen geneigd zijn de mannelijke vorm als algemeen of 'neutraal' te zien. Dit heeft echter tot gevolg dat de ene groep kandidaten meer mogelijkheden tot identificatie heeft dan de andere en dat sommige kandidaten de indruk krijgen dat de groep waar zij deel van uitmaken van minder belang is of in het geheel niet meetelt. Hierdoor kan bij leden van minderheids- of andere groeperingen een gevoel van uitsluiting en verminderde motivatie ontstaan. Zo blijkt uit onderzoek naar het effect van het gebruik van vooral mannelijke personages en verwijswaarden (voor een overzicht zie Porreca, 1984) onder andere

- dat mensen zich zelden een vrouwelijk personage voorstellen wanneer 'hij' gebruikt wordt of een andere mannelijke vorm, zoals 'mankracht', 'de kandidaat' of 'de bestuurder', terwijl in het algemeen gesproken wordt. Neutrale termen, zoals 'zij' (meervoud), 'mensen' e.d. worden daarentegen wel ervaren als geldend voor de twee seksen;
- dat meisjes tot 12 jaar vaak 'zij' (enkelvoud) gebruiken als zij in het algemeen spreken en moeten leren dat termen als 'de leerling' en 'hij' ook op hen kunnen slaan. Zij voelen zich dan ook vaak buitengesloten door het gebruik van algemeen bedoelde mannelijke vormen;
- dat mensen informatie beter onthouden als de eigen sekse in de informatie genoemd wordt.

Uit onderzoek naar aanleiding van advertenties tenslotte blijkt dat mensen meer geïnteresseerd zijn en hun eigen geschiktheid hoger inschatten als hun sekse expliciet genoemd wordt.

Aanbeveling

Zorg ervoor dat de toets als geheel de diversiteit reflecteert van de kandidaten voor wie hij bedoeld is. Dit betekent dat opgaven en afbeeldingen niet uitsluitend blanke Nederlanders tot onderwerp dienen te hebben. Ook vrouwen, meisjes en leden van etnische minderheids-groepen moeten vertegenwoordigd zijn. Vermijd het gebruik van 'hij' wanneer in het algemeen gesproken wordt. De mededeling dat 'hij' ook als 'zij' gelezen kan worden is niet zo'n goede oplossing. Kandidaten worden hierdoor ongelijk behandeld. Wanneer naar een persoon of naar personen verwezen wordt, gebruik dan duidelijke, correcte en onpartijdige termen.

Voorbeelden

15 en 16 zijn twee representatieve items uit een examen waarin 10 vragen worden gesteld naar aanleiding van polisvoorwaarden van een verzekering.

- (15) De verzekerde moet voor zaken in Parijs zijn. Bij Brussel begeeft de auto het; de reparatie duurt 1 dag. Om 15.00 uur heeft verzekerde echter een belangrijke zakelijke afspraak. Daarom huurt hij een auto om toch op tijd te komen.

Stelling: De schade van f 350,-, zijnde de prijs van de huurauto, betaalt Ohra uit.

- (16) De verzekerde en familie zijn op vakantie in het buitenland. Verzekerde moet plotseling terug naar Nederland in verband met het overlijden van zijn broer. Hij gaat per vliegtuig. De familie krijgt een chauffeur van de verzekering die hen (sic!) later terugrijdt naar Nederland. Totaal gemaakte kosten: f 2000,- (vliegticket/reis- en verblijfskosten/vervangende chauffeur/benzine e.d.).

Stelling: Ohra vergoedt f 2000,-.

Toelichting

In alle tien items wordt er automatisch van uitgegaan dat 'de verzekerde' een man is, waarmee de suggestie wordt gewerkt dat vrouwen zich niet (kunnen) verzekeren. Toch is juist het woord 'verzekerde' zeer neutraal en gaat het hier om voorbeelden. De toetsconstructeurs vinden deze mannelijke uitwerking zelfs zo vanzelfsprekend dat ze niet de noodzaak zien om 'de verzekerde' te introduceren. In het eerste item moeten de kandidaten er pas in de laatste regel achterkomen dat 'de verzekerde' een man is. In het tweede item wordt de sekse van de verzekerde pas in de derde regel geëxpliciteerd.

Het zou beter zijn in een aantal items een verzekerde vrouw op te voeren. Bijvoorbeeld:

De verzekerde moet voor zaken in Parijs zijn. Bij Brussel begeeft haar auto het; de reparatie duurt 1 dag. Om 15.00 uur heeft zij echter een belangrijke zakelijke afspraak. Daarom huurt zij een auto om toch op tijd te komen.

Stelling: De schade van f 350,-, zijnde de prijs van de huurauto, betaalt Ohra uit.

De verzekerde is met haar familie op vakantie in het buitenland. Verzekerde moet plotseling terug naar Nederland in verband met het overlijden van haar broer. Zij gaat per vliegtuig. De verzekering zorgt voor een chauffeur die de auto met de achtergebleven familieleden terugrijdt naar Nederland. Totaal gemaakte kosten: f 2000,- (vliegticket/reis- en verblijfskosten/vervangende chauffeur/benzine e.d.).

Stelling: Ohra vergoedt f 2000,-.

In het natuurkunde-examen waar het volgende item deel van uitmaakt, komen geen vrouwelijke personages voor. Wel enkele mannelijke en relatief veel gemotoriseerde vervoermiddelen.

- (17) Albert, een fanatieke motorrijder, moet op een koude winteravond een lange motorrit maken. Albert trekt voor deze rit dubbele kleren onder z'n motorpak aan. Van welk soort isolatie maakt Albert gebruik?

- A straling
- B stroming
- C geleiding
- D* stilstaande lucht

Toelichting

Op zich is er niets tegen motorrijders in een toetsopgave. De interesse voor motoren en auto's blijkt echter nogal te verschillen tussen meisjes en jongens. Daarom was er wellicht iets voor te zeggen geweest om het aantal motorrijders in het desbetreffende examen enigszins in te perken en ook eens een vrouwelijk personage op te nemen.

Bijvoorbeeld:

Eva moet op een koude winteravond nog een lange fietstocht maken. Ze trekt twee truien aan onder haar jas. Van welk soort isolatie maakt Eva gebruik?

4.2 Etnocentrische inhoud

Men noemt uitspraken 'etnocentrisch' indien de eigen groep of cultuur als superieur of universeel wordt afgeschilderd. Kenmerken van een etnocentrische houding zijn loyaliteit aan de eigen groep en een streven naar polarisatie tussen de eigen groep en vreemde groepen. Men ziet degenen die niet tot de eigen groep behoren vooral als vreemd en anders. Dit kan inhouden dat men 'wij' gebruikt in een als 'universeel' bedoelde uitspraak, terwijl uit de inhoud blijkt dat de uitspraak in feite slechts voor een beperkte groep geldt. ('We denken allemaal wel eens als we ons 's morgens staan te scheren...' of: 'als we dan zo'n Turkse vrouw met een hoofddoek op straat zien lopen'). Impliciet blijkt een dergelijke houding ook wanneer men andere landen, andere etnische groepen of vrouwen presenteert als hulp-behoevend, incompetent, 'achter bij', 'nog niet zo ver als', en dergelijke.

Gelijkwaardigheid

Vaak zijn de formuleringen waaruit men kan afleiden dat groepen niet als gelijkwaardig worden gepresenteerd erg subtiel. Het komt regelmatig voor dat vrouwen veel meer dan mannen worden aangeduid in een relatie tot een ander persoon, zoals 'moeder van' of 'vrouw van'.

'Neutrale' personages

In de meeste examens en toetsen is het mannelijke personage de norm. Wanneer men in het algemeen spreekt en een neutraal personage ter illustratie opvoert, is dat een man. Vrouwen worden voorgesteld als uitzonderingen op die norm en zij worden ook zeer nadrukkelijk als zodanig geïntroduceerd. In een natuurkunde-examen wordt bijvoorbeeld een aantal vragen gesteld naar aanleiding van een tekst van 17 regels over een quiz waaraan wordt deelgenomen door een 'quizmaster' en 'drie kandidaten A, B en C'. Deze worden niet nader geïntroduceerd, maar in de vierde regel komt de lezer erachter dat de kandidaten kennelijk mannelijk zijn: '..., drukt **hij** even op de knop'. Over de sekse van de quizmaster blijkt de kandidaat lange tijd in het ongewisse. Pas in de derde opgave wordt duidelijk dat ook dit een man is ('... heeft de quizmaster ook een drukschakelaar tot **zijn** beschikking'). In een volgende opgave komt 'een waarnemer' voor. Pas in de tweede opgave naar aanleiding van deze 'waarnemer' wordt bekend gemaakt dat de waarnemer mannelijk is ('**Zijn**' camera). In het bewuste examen komt één vrouwelijk personage voor. De examenmakers gaan dan heel anders te werk. Het eerste woord van de opgave zorgt ervoor dat er geen misverstand mogelijk is ('**Petra**' slaat een tennisbal horizontaal weg'.) In de opgave wordt haar naam dan nog vijf maal herhaald, terwijl ook nog vrouwelijke verwijswwoorden worden gebruikt. Een meer gelijke behandeling zou de voorkeur verdienen.

Voorbeeld

Het volgende item illustreert Europese ethnocentriciteit, ook wel eurocentrisme genoemd:

(18) De Europese gemeenschap helpt een arm land met 30 miljoen ecu. Een ecu is ongeveer een rijksdaalder waard. Welk bedrag is dat ongeveer in guldens?

- A 75 000
- B 7 500 000
- C 12 000 000
- D 75 000 000

Toelichting

Een ander land wordt hier op paternalistische wijze beschreven als hulpbehoevend ten opzichte van het 'rijke westen', niet als een gelijkwaardige partner. Mogelijke verbeteringen zouden kunnen zijn:

- De Europese gemeenschap leent een bedrag van 30 miljoen ecu aan een land in Oost-Europa, enz.
- De Europese munteenheid ecu is ongeveer f 2,50 waard. Hoeveel is 30 miljoen ecu ongeveer in guldens?

Aanbevelingen

Multiculturele samenleving

Stel de Nederlandse samenleving niet voor als een monoculturele gemeenschap. Inwoners van Nederland verschillen in vele opzichten van elkaar. Deze verschillen zijn overigens niet automatisch gekoppeld aan etniciteit, huidskleur of sekse. Wanneer de verschillende groepen in de samenleving ter sprake komen, beschrijf deze dan als gelijkwaardig. Dit houdt onder andere in dat leden van die verschillende groepen op dezelfde manier beschreven of aangeduid worden. De indruk moet niet gewekt worden dat leden van minderheden en vrouwen afwijkend zijn, minder of uitzonderingen op een bepaalde norm.

Geen polarisatie

Polarisatie zou vermeden moeten worden. Gebruik geen 'wij' voor blanke Nederlanders en 'zij' voor andere groepen zoals allochtonen. Vermijd termen die een oordeel bevatten. Etnische groepen kan men het beste aanduiden met de termen die zij zelf gebruiken. Hetzelfde geldt voor situaties en gebeurtenissen die specifiek zijn voor bepaalde etnische of andere groepen.

Bij het formuleren van items zou men ervan uit moeten gaan dat mensen een eigen identiteit hebben en hun bestaansrecht niet ontleen aan het feit dat ze getrouwd zijn met of familie zijn van iemand die lid is van een meer belangrijke of meer interessante groep. Een stam als 'De vrouw van Piet Jansen doet vrijwilligerswerk voor Amnesty International' is onaanvaardbaar, tenzij deze zin deel uitmaakt van een grotere tekst waarin Piet Jansen de hoofdpersoon is. Zo niet, dan kan de zin herschreven worden als 'Piet Jansen/Marie de Bruin/Malika/ doet vrijwilligerswerk enz.

4.3 Negatieve inhoud en stereotypen

De regel dat uitspraken en woorden niet beledigend mogen zijn voor bepaalde bevolkingsgroepen, spreekt eigenlijk voor zich, maar sommige itemschrijvers lijken een uitzondering te maken voor afleiders.

Negatieve afleiders

Bij de constructie van items bij teksbegripexamens worden soms afleiders geformuleerd op basis van vooroordelen.

Voorbeeld (fictief):

- (19) Wat zegt de auteur in de eerste alinea over vrouwelijke chauffeurs?
C Ze rijden minder goed dan mannen.

Kandidaten weten echter niet van te voren wat het goede antwoord is. De status van de alternatieven is dus in principe gelijk voor hen. Ook het argument dat wel gebruikt wordt dat zo racistische of seksistische kandidaten gestraft worden, is tegen de principes van een goede toets, tenzij het meten van dergelijke attitudes het doel van de toets is. Overigens blijkt uit de inhoudelijke analyse van items met bias dat sommige afleiders voor de ene groep veel aantrekkelijker zijn dan voor de andere (Bügel, 1993). Ook hierdoor kan DIF en bias ontstaan (Green, Crone & Folk, 1989).

Negatieve leerstof

Men kan verschillende soorten negatieve items onderscheiden. Het duidelijkst zijn items die als enig doel lijken te hebben te toetsen of de kandidaat de negatieve feiten kent over een bepaalde bevolkingsgroep of een bepaald land. Bijvoorbeeld een item dat vraagt in welke bevolkingsgroep de meeste criminaliteit voorkomt of een toets over Turkije waarin voornamelijk vragen staan over de percentages analfabeten en werklozen. In het algemeen kan men stellen dat het toetsen van dit soort kennis alleen zinvol is als het in een breder verband gebeurt, waarbij bijvoorbeeld ook de historische ontwikkelingen, de oorzaken of de positieve aspecten aan bod komen.

Negatieve contexten

Een ander type negatieve items zijn vragen naar aanleiding van negatief contextmateriaal, zoals vragen bij tekstbegrip of bij geschiedenis naar aanleiding van controversiële teksten, bijvoorbeeld van auteurs die van mening zijn dat sommige groepen of 'rassen' inferieur zijn aan andere. Het gebruik van dergelijk contextmateriaal kan natuurlijk wel zinvol zijn binnen een bepaald onderwijsdoel. Bijvoorbeeld bij maatschappijleer in het kader van een onderwerp als discriminatie of bij geschiedenis over bepaalde ideologieën.

Stereotypen

Stereotypen kunnen beter vermeden worden. Mensen zijn soms geneigd bepaalde groepen - vooral die waartoe ze zelf niet behoren - als meer gelijkvormig waar te nemen dan andere. De generalisaties die hier het gevolg van zijn, kunnen echter beledigend zijn of een andere negatieve uitwerking hebben, omdat ze suggereren dat leden van bepaalde groepen minder geschikt zijn voor bepaalde rollen, werkzaamheden of functies. Leden van minderheidsgroeperingen en vrouwen zijn zich over het algemeen zeer bewust van de heersende stereotyperingen die hen betreffen en kunnen zich eraan ergeren. Stereotyperingen voor vrouwen zijn bijvoorbeeld: passief, emotioneel,

ijdel, niet logisch kunnen denken. Voor mannen: agressief, dominant, rationeel. Voor Surinamers: lui, meer aanleg voor fysieke dan voor intellectuele taken, enz. Wanneer kinderen afgebeeld worden, worden jongens vaak als ouder en/of groter weergegeven dan meisjes. Ook worden vrouwen en leden van minderheden vaak in vaste rolpatronen weergegeven. Maar vrouwen houden zich niet uitsluitend bezig met het huishouden of hun man en kinderen, allochtonen hoeven niet alleen met problemen geassocieerd te worden.

Omgekeerde wereld

Het is niet de bedoeling dat alle vrouwelijke personages als piloot, politieagente, enz. worden afgebeeld en leden van etnische minderheden uitsluitend in leidinggevende posities. In alle gevallen gaat het om het beeld dat door de toets als geheel gepresenteerd wordt.

Aanbeveling

Controleer of etnische minderheden en vrouwen niet overwegend negatief worden afgeschilderd of uitsluitend in verband worden gebracht met problemen. Vermijd polarisatie tussen verschillende groepen evenals een stereotyperende weergave van mensen. Dit laatste geldt zowel ten aanzien van typering en karaktertrekken als ten aanzien van rolpatronen en werkzaamheden. Controversiële teksten kunnen beter niet opgenomen worden, tenzij dergelijk materiaal deel uitmaakt van de toetsspecificatie of van het programma dat de leerlingen hebben bestudeerd.

Voorbeelden

De volgende twee items (20 en 21) voldoen niet aan de genoemde criteria.

(20) (schema: soort organogram)

Wie heeft in bovenstaand schema een lijnfunctie?

- A chef spinnerij en de secretaresse.
- B chef atelier en de boekhouder.
- C bedrijfsleider en inkoper.
- D* chef atelier en chef spinnerij.

(21) Zo ook een werknemer die gegevens inbrengt in een computer. Welke functie heeft deze werknemer?

- A Procuratiehouder.
- B Boekhouder.
- C* Datatypiste.
- D Receptioniste

Toelichting

In beide items wordt gesuggereerd dat bepaalde functies alleen voor mannen bestemd (of algemeen?) zijn en andere - lagere functies - alleen voor vrouwen.

Het eerste item kan alleen verbeterd worden als het organogram (dat hier niet is afgebeeld) sekseneutraal wordt gemaakt. Het tweede item valt ook op door de onduidelijke vraagstelling. Het zou ook als volgt geformuleerd kunnen worden:

Frits Sanders heeft als taak gegevens in te brengen in een computer. Welke functie heeft deze werknemer?

- A Procuratiehouder.
- B Boekhouder.
- C* Datatypist.
- D Receptionist.

5 Maatregelen in verschillende fasen van het proces van toetsconstructie en toetsanalyse

In het toetsconstructieproces kan men verschillende fasen onderscheiden (Eggen en Sanders, 1993). In de meeste daarvan kan men maatregelen nemen om partijdigheid te voorkomen. Het is beter al in een vroeg stadium - voor de constructie van de toets - rekening te houden met de diversiteit van de kandidaten om ingrijpen in latere fasen van het constructieproces zo veel mogelijk te beperken.

5.1 Maatregelen vooraf

Uitgangspunten Cito

Afspraken ter voorkoming van partijdigheid, kunnen het beste schriftelijk worden vastgelegd, als uitgangspunten van het Instituut, vóór het overleg met de opdrachtgevers en vóór de eigenlijke toetsconstructie. Hierin zou verwezen kunnen worden naar deze *Richtlijnen voor de ontwikkeling van onpartijdige toetsen*. Cito-medewerkers die commissies van itemschrijvers begeleiden kunnen zich dan beroepen op het instituutsbeleid, waardoor persoonlijke discussies vermeden kunnen worden. Om er zeker van te zijn dat de richtlijnen ook daadwerkelijk worden toegepast, moet de screening op dergelijke ongewenste kenmerken deel uit maken van de procedure van toetsconstructie.

Bevordering van deskundigheid

Het voorkomen van partijdigheid vraagt om deskundigheid die meer inhoudt dan objectiviteit, goede bedoelingen en gezond verstand. De meeste mensen zijn niet op de hoogte van de gevoeligheden, gewoontes en cultuur van etnische minderheden of de verschillen in interesse en achtergrondkennis van meisjes en jongens. Toetsspecialisten moeten in de gelegenheid gesteld worden hun deskundigheid op dit terrein te vergroten. Medewerkers van het Cito die toetsen samenstellen of de toetsconstructie begeleiden zouden daartoe een cursus moeten kunnen volgen. Leden van Advies Commissies van Docenten en andere itemschrijvers zouden de beschikking moeten hebben over schriftelijk materiaal.

Deskundigen

Het verdient tevens aanbeveling een commissie van deskundigen samen te stellen die advies kan geven en informatie kan verstrekken indien dat noodzakelijk is, bijvoorbeeld op het gebied van taalgebruik, cultuurverschillen tussen bepaalde etnische groepen of verschillen in kennis en interesse tussen meisjes en jongens.

5.2 Toetsspecificatie

Kandidaten

Een van de vragen die in deze fase aan de orde komen, is:

- Voor wie is de toets bestemd?

Maken zowel jongens als meisjes de toets? Zijn er (veel) allochtone kandidaten? Heeft de toets betrekking op onderwerpen waar de rol of de opvattingen van etnische minderheden van belang zijn, bijvoorbeeld religie, immigratie of sociaal werk in stadswijken waar veel allochtonen wonen? De antwoorden op deze vragen hebben

consequenties voor de rest van het toetsconstructieproces. In principe moet een toets voor alle potentiële groepen kandidaten geschikt zijn. Leden van beide seksen en leden van minderheidsgroepen (bijvoorbeeld als 'screeners') dienen zoveel mogelijk ingeschakeld te worden in het toetsontwikkelproces. Dit is absoluut noodzakelijk als de toets betrekking heeft op stof over minderheidsgroeperingen of vrouwen of wanneer de kandidaten voor een groot deel bestaan uit leden van deze groepen.

Inhoud toets

Een andere belangrijke vraag is:

- Wat is de inhoud van de toets?

Deze wordt bij studietoetsen vaak bepaald door een tweedimensionale toetsmatrijs, waarbij het leerlinggedrag en de inhoudscomponent de twee dimensies zijn waarmee de toetsvragen geassocieerd kunnen worden. Men kan zich hier afvragen of de inhoudscomponent en ook het eventueel gebruikte contextmateriaal in gelijke mate interessant zijn voor de verschillende groepen kandidaten die de toets maken. Wanneer de toets direct of indirect betrekking heeft op groepen kandidaten is de manier waarop de desbetreffende onderwerpen worden gepresenteerd van belang. Bij de behandeling van etnische minderheden in Nederland hoeven deze groepen niet onnodig geproblematiseerd te worden. Wanneer er vragen gesteld worden over een land als Turkije dient ervoor gewaakt te worden dat niet alleen de problemen in het land aan de orde komen of dat iedere ontwikkeling in dat land beoordeeld wordt volgens Nederlandse of westerse maatstaven.

Voorbeeld

(22) Hieronder staan een aantal grafieken. De zinnen die onder de grafieken horen zijn door elkaar geraakt.

Dit zijn de zinnen. Schrijf het nummer van de juiste zin bij de juiste grafiek (op het antwoordblad!).

- I Aantal artsen per 1000 inwoners,
- II Aantal TV's per 1000 inwoners,
- III Aantal analfabeten in %,
- IV Aantal auto's per 1000 inwoners.

(afbeelding: vier staafdiagrammen waarin bovenstaande gegevens zijn weergegeven voor Turkije, Griekenland, Nederland, Syrië, Irak, Indonesië, Mali en Mexico.

Toelichting

De leerlingen kunnen het antwoord vinden als ze weten dat derde wereld landen meer analfabeten en minder auto's, t.v.'s en artsen hebben dan rijke westerse landen. Met andere woorden dat Nederland 'ontwikkelder' is dan Turkije en Turkije 'ontwikkelder' dan bijvoorbeeld Mali. Je kunt je afvragen of dit de belangrijkste kenmerken van landen zijn die leerlingen moeten weten. Een dergelijke vraag moet natuurlijk wel beoordeeld worden in relatie tot de andere vragen van de toets.

De keuze van goed contextmateriaal is essentieel voor de kwaliteit van een toets. Het is vooral belangrijk dat een afweging kan worden gemaakt. Kennis van mogelijke gevolgen van bepaalde keuzes is hiervoor onontbeerlijk. Uit onderzoek blijkt dat de aard van het context-materiaal invloed heeft op het gemak waarmee de kandidaten tot een 'probleemrepresentatie' komen, dat wil zeggen begrijpen wat er precies gevraagd wordt en hoe ze een oplossing moeten vinden. Het contextmateriaal heeft ook invloed op de interesse en de motivatie van de leerlingen. Een specificatie van het te gebruiken contextmateriaal zou daarom in de toetsspecificatie moeten worden opgenomen, waarbij

de volgende vragen beantwoord zullen moeten worden: Is het beter bekende of onbekende context op te nemen? Gaat de voorkeur uit naar fictieve of naar authentieke situaties? Is de verdeling van domeinen en personages in overeenstemming met de verschillen in achtergrondkennis van de verschillende groepen kandidaten? (zie verder 4.3).

Toetstype

- Welke vorm?

Ook bij de keuze voor open en gesloten vragen dient men rekening te houden met de doelgroep. Uit onderzoek blijkt dat meisjes open vragen vaak relatief beter maken, jongens gesloten vragen (Murphy, 1983; Ouborg, 1987; Hellekant 1994). De geringere kennis van de Nederlandse taal van sommige groepen kandidaten kan ook een rol spelen bij de keuze voor open of gesloten vragen bij toetsen waar ook allochtone leerlingen aan deelnemen. Men kan hier echter geen eenduidig advies geven. Als de items eenvoudig geformuleerd kunnen worden, kan dat een reden zijn om voor meerkeuzevragen te kiezen. Bestaan de alternatieven echter uit lange, ingewikkelde zinnen en zijn de verschillen tussen de alternatieven subtiel, dan kunnen meerkeuze-items een probleem vormen voor minder taalvaardige leerlingen. Open vragen, waarbij de kandidaten zelf het antwoord kunnen formuleren, hebben dan wellicht de voorkeur.

5.3 Itemconstructie

Contexten

Bij veel toetsen worden de vragen gesteld naar aanleiding van een situatie, een tekst of een context. Het is beter de keuze van dit tekst- en contextmateriaal niet aan het toeval of aan de welwillendheid van de individuele itemschrijver over te laten, maar ze in de toetsspecificatie te vermelden. Men dient wel te bedenken dat iedere keuze bepaalde ongewenste gevolgen kan hebben. Een veel voorkomend misverstand is dat men - om bias te voorkomen - in de toets vooral materiaal moet opnemen dat voor niemand interessant is. Om tekstbegrip te toetsen kiest men dan bijvoorbeeld een tekst over de fabricage van bakstenen. Uit onderzoek blijkt echter dat oninteressant stimulusmateriaal vooral de prestaties van jongens negatief beïnvloedt (Asher, 1980). Intelligente leerlingen worden extra bevoordeeld indien men een zeer breed scala aan onderwerpen kiest vanwege hun grotere 'algemene kennis' (Johnston, 1984). Contexten die leerlingen interessant vinden, bevorderen de motivatie van de leerlingen en bekende contexten hebben invloed op het gemak waarmee de leerlingen tot het oplossen van problemen komen. De bekendheid van de verschillende onderwerpen is echter niet gelijk voor verschillende groepen leerlingen. Meisjes weten meer over onderwerpen die gaan over mensen, verzorging, huishoudelijke zaken, kunst en filosofie. Jongens hebben meer kennis over apparaten, sport, geld, geweld, politiek en economie. Over verschillen in achtergrondkennis tussen verschillende etnische groepen in Nederland is nauwelijks iets bekend. Men kan in ieder geval proberen variatie aan te brengen in de contexten en situatieschetsen. Daarom zou men in bepaalde gevallen kunnen overwegen bij een toets over informatieverwerking niet alle grafieken te laten gaan over neutrale en daardoor oninteressante onderwerpen zoals 'aantallen werknemers bij computerbedrijven', 'ledenaantallen van plaatselijke verenigingen', de 'fabricatie van fietsen en bromfietsen', en dergelijke, maar ook rekening te houden met verschillen tussen kandidaten. Men kan bijvoorbeeld de context toespitsen op bepaalde doelgroepen en onderwerpen kiezen als de ontwikkeling van het percentage Marokkaanse en Turkse leerlingen dat naar mavo en havo gaat, het percentage vrouwen dat een baan heeft, of een andere meer menselijke en voor sommige groepen meer aantrekkelijke inhoud.

Evenredige vertegenwoordiging

Het is opvallend dat in bijna alle toetsen waarin personen voorkomen, dit vaker mannelijke personages zijn dan vrouwelijke. Leden van etnische minderheden zijn over het algemeen geheel afwezig. Meestal worden personen opgevoerd in een bedachte context, kennelijk bedoeld om het item aantrekkelijker te maken of om het meer in te bedden in de dagelijkse werkelijkheid. In deze gevallen is het zeer eenvoudig om een meer evenwichtige en meer realistische samenstelling van de bevolking van 'toetspersonages' te bewerkstelligen.

Itembank

Wanneer de items deel uitmaken van een itembank, kan men bij de itemspecificatie gegevens opnemen over de aantrekkelijkheid van het item voor verschillende groepen kandidaten en over de aard van het contextmateriaal. In sommige gevallen kan het aanbeveling verdienen verschillende versies van een item te maken, dat wil zeggen dat men een bepaalde vraag van verschillende contexten voorziet (één uit de medische sfeer, één uit een meer technische sfeer, enz.).

Relevantie contexten

In alle gevallen dient echter het belangrijkste uitgangspunt te zijn dat zowel de vragen als het bijbehorende contextmateriaal relevant moeten zijn met betrekking tot de kennis of vaardigheden die gemeten moeten worden. De context of situatieschets moet met zorg gekozen of samengesteld worden. Deze moet zinvol zijn en een functie hebben, niet afleiden of de probleemstelling onduidelijk maken.

Voorbeeld

(23) Carla's ouders zijn gescheiden. Carla's moeder heeft geen baan en het geld dat ze iedere maand van Carla's vader krijgen is veel te weinig om van te leven. Wie kan ervoor zorgen dat ze iedere maand wat extra geld krijgen?

- A het Arbeidsbureau
- B de Burgerlijke Stand
- C het Leger des Heils
- D* de Sociale Dienst

Toelichting

In dit item worden veel stereotiepe vooronderstellingen gemaakt. Bijvoorbeeld dat 'Carla' aan haar moeder is toegewezen en dat Carla's moeder geen recht heeft op een werkloosheidsuitkering. Het gebruik van 'ze' is ook verwarrend. Het is niet duidelijk wat er in dit item getoetst wordt. Moeten leerlingen weten dat veel gescheiden vrouwen bijstand ontvangen? Gaat het erom dat ze weten welke mensen in aanmerking komen voor bijstand? Of is de vraag alleen welke instantie uitkeringen verstrekt? In het laatste geval verdient een meer neutrale stam, de voorkeur.

Carla is in juli afgestudeerd. Zes maanden later vond zij een baan. In die zes maanden leefde zij van een uitkering. Waar heeft zij die moeten aanvragen?

5.4 Samenstellen van de toets

Stereotypen

Pas bij het samenstellen van de hele toets kan men bepaalde aspecten beoordelen, zoals een stereotiepe weergave van bepaalde groepen. Hiervoor moet men namelijk alle items beoordelen. De volgende vragen komen hierbij aan de orde: Worden allochtonen vooral als 'anders' voorgesteld dan autochtonen? Worden vrouwelijke en mannelijke personages gelijk behandeld of zijn alle 'neutrale' items mannelijk? Worden vrouwen alleen als uitzonderingen op de regel afgebeeld of steeds als 'huisvrouw', of als incompetent of hulpbehoevend?

Voorbeeld

- (24) Al vier weken ligt de straat rond de winkel van mevrouw Gerritsen nu al opengebroken, omdat er een nieuwe riolering aangelegd moet worden. Mevrouw Gerritsen vindt dat er nu maar eens iets moet gebeuren. Ze wil een brief gaan schrijven. Een echte protestbrief. Ze weet alleen niet aan wie. Aan wie zal mevrouw Gerritsen het best de brief kunnen richten?
- A* aan de burgemeester en wethouders van de gemeente
 - B aan de koningin
 - C aan de korpschef van de politie in de gemeente
 - D aan de minister van onderwijs

Toelichting

Dit is de enige keer dat een werkende vrouw wordt opgevoerd in de toets waar dit item deel van uitmaakt. Het is daarom geen goed idee haar als incompetent af te schilderen. Mogelijke verbetering:

Al vier weken ligt de straat rond de winkel van mevrouw Gerritsen nu al opengebroken, omdat er een nieuwe riolering aangelegd moet worden. Mevrouw Gerritsen vindt dat veel te lang. Ze wil een protestbrief schrijven. Aan wie kan mevrouw Gerritsen de brief het beste richten?

Eerlijke verdeling

Ook volgorde- en numerieke aspecten komen zo aan het licht. Worden mannelijke personages altijd vóór vrouwelijke genoemd? Hoe is de verhouding mannen-vrouwen, allochtonen-autochtonen, blanken versus niet-blanken in het toetsmateriaal? Staan er evenveel onderwerpen in die meisjes interessant vinden als onderwerpen die jongens aanspreken? Als dit niet zo is, wordt deze keuze dan gerechtvaardigd door de leerstof?

5.5 Toetsafname

Biasdetectie in de pretest

Wanneer men een proefafname uitvoert, kunnen aan de hand van de uitkomsten nog items worden verwijderd of bijgesteld. Daarom dienen er procedures toegepast te worden om vast te stellen dat verschillen in gemiddelde groepsprestaties veroorzaakt worden door verschillen in de vaardigheid die getoetst wordt en niet door irrelevante factoren.

Maatregelen

Wanneer items met bias gedetecteerd worden, dient de vraag beantwoord te worden of deze items uit de toets verwijderd moeten worden of dat de effecten geneutraliseerd moeten worden door middel van items die tegenovergestelde bias vertonen. Het antwoord op deze vraag is niet eenvoudig te geven. Het hangt af van de toetsspecificatie en van de leerstof. Bij tekstbegrip kan men de component 'achtergrondkennis' nooit uitsluiten. Het is in dat geval vaak beter te compenseren. Bij de keuze van de teksten moet men rekening houden met de interesses en verschillen in kennis van verschillende groepen kandidaten. Bij andere toetsen of vakken kan die keuze anders zijn en verdient het wellicht de voorkeur om items met bias te vervangen door neutrale items.

Geen pretest

Indien een proefafname niet mogelijk is, dient er extra veel aandacht te worden besteed aan het constructieproces. De statistische gegevens na de definitieve afname kunnen dan misschien helpen om missers bij volgende toetsen te voorkomen. In sommige gevallen kunnen bepaalde items niet meegerekend worden bij de eindscore. Signalering van items die itembias vertonen zou tot de standaardprocedures moeten behoren bij de analyse van examens en toetsen. In ieder geval dienen toetsscores opgesplitst te worden in groepsscores voor kandidaten van verschillende etnische achtergrond en van verschillende sekse, wanneer steekproeven van voldoende omvang beschikbaar zijn.

6 Literatuur

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association (1988). Guidelines for nonsexist language in APA journals. In: *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.
- Asher, S.R. (1980). Topic interest and children's reading comprehension. In R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 525-534), Hillsdale, NJ: Erlbaum.
- Bossers, B.H. (1992). *Reading in two languages. A study of reading comprehension in Dutch as a second language and in Turkish as a first language*. Proefschrift. Rotterdam: Van Driel.
- Bügel K. (1993). *Sekseverschillen in tekstbegrip bij moderne vreemde talen*. Proefschrift. Arnhem: Cito.
- Coenen, M. en Vallen, T. (1991). Itembias in de eindtoets basisonderwijs. *Pedagogische Studiën*, 68, 15-26.
- Dais, T.A. (1993). An analysis of transition assessment practices: do they recognize cultural differences? In: T. Dais; and others. *Selected readings in transition: cultural differences, chronic illness and job matching*. Vol.2. Illinois University.
- Educational Testing Service (1987). *ETS sensitivity review process. An overview*. Princeton NJ: Educational Testing Service.
- Educational Testing Service (1987). *ETS standards for quality and fairness*. Princeton NJ: Educational Testing Service.
- Eggen, T.J.H.M. & Sanders. P.F. (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Fincher, C. (1990). *Issues in public examinations. Social justice, public interest and the SAT*, paper werd gepresenteerd op de conferentie van de International Association for Educational Assessment, Maastricht.
- Fremer, J., Diamond, E.E., & Camara, W.J. (1989). Developing a code of fair testing. Practices in education. *American Psychologist*, 44, 1062-1067.
- Green, B.F., Crone, C.R., & Folk, V.G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement*, 26, 147-160.
- Groen, H., & Kreeft, H. (1992). *Een voorstel voor kwaliteitscriteria voor centrale examens in het voortgezet onderwijs*. Arnhem: Cito.
- Hellekant, J. (1994). Are multiple-choice tests unfair to girls? *System*, 22, 349-353.
- Holland, P.W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale NJ: Lawrence Erlbaum.
- House, E.R., & Care, N.S. (1979). Fair evaluation agreement. *Educational Theory*, 29, 159-169.
- Johnston, P. (1984). Prior knowledge and reading comprehension test bias, *Reading Research Quarterly*, 19, 219-239.
- Joint Committee on Testing Practices (1988). *Code of fair testing practices in education*. Washington D.C.: Joint Committee on Testing Practices.
- Jong, M. de, & Vallen, T. (1989). Linguïstische en culturele bronnen van itembias in de Eindtoets Basisonderwijs voor leerlingen uit etnische minderheidsgroepen. *Pedagogische Studiën*, 66, 390-402.
- Kerkhoff, A., & Vallen, T. (1984). Cultural Biases in Second Language Testing of Children. In G. Extra en T. Vallen (Eds.), *Ethnic minorities and Dutch as a second language* (pp. 133-147). Dordrecht: Foris.
- Kleijn, P. de, & Nieuwborg, E. (1987). *Basiswoordenboek Nederlands*. Groningen: Wolters-Noordhoff.
- Langer, J.A. (1984). Examining background knowledge and text comprehension. *Reading Research Quarterly*, 19, 468-481.

- Ministry of Education, Western Australia, Social Justice Branch (1991). *A fair day. Guidelines for the use of bias-free language*. Ministry of Education, Western Australia.
- Mok, I. (1990). *Racisme en schoolboeken*. Amsterdam: Parel.
- Murphy, R.J.L. (1980). Sex differences in GCE examinations entry statistics and success rates. *Educational Studies*, 6, 169-178.
- Nederlands Instituut van Psychologen (1988). *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studieboeken*. Assen: Van Gorcum.
- Ouborg, M.J. (1987). *Sexeverschillen en antwoordvormen*. Onderzoek naar verschillen in resultaten tussen meisjes en jongens op de eindexamenvragen van 1984 en 1986. Specialistsich bulletin Nr. 57. Arnhem: Cito.
- Steffensen, M.S., Joag-Dev, C., & Anderson, R.C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 15, 10-29.
- Porreca, K.L. (1984). Sexisme in current ESL textbooks. *TESOL Quarterly*, 18, 705-724.
- Office for Minority Education (1980). *An approach for identifying and minimizing bias in standardized tests: a set of guidelines*. Princeton, NJ: Educational Testing Service.
- Scheuneman, J.D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics, *Journal of Educational Measurement*, 27, 109-131.
- SLO (1995). *Meertaligheid en schoolse taalvaardigheden. Tips voor auteurs en redacteurs van schoolboeken voor het voortgezet onderwijs*. Enschede: SLO.
- Testscreeningscommissie (1990). *Toepasbaarheid van psychologische tests bij allochtonen*. Utrecht: Landelijk Bureau voor Racismebestrijding.
- Tittle, C.K. (1982). Use of Judgemental Methods in Item Bias Studies. In R.A. Berk (Ed.), *Handbook for detecting test bias*, Baltimore, London: The John Hopkins University Press.
- Uiterwijk H. (1994) *De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen*. Proefschrift. Arnhem: Cito.
- United States Commission on Civil Rights (1993). *The validity of testing in education and employment*. Washington, DC.
- Verhallen-van Ling, M. (1994). *Lexicale vaardigheid van Turkse en Nederlandse kinderen: een vergelijkend onderzoek naar betekenis-toekenning*. Proefschrift Universiteit Amsterdam.
- Women in EFL materials (1991). On balance. Guidelines for the representation of women and men in english language teaching materials. *IATEFL Newsletter*, 113, 8-10.

Checklist

Deze checklist bevat een aantal criteria waarmee een toets als geheel beoordeeld kan worden. Een aantal criteria geldt ook voor afzonderlijke items. Er worden een aantal voorbeelden ge-given van onaanvaardbare stereotyperingen van groepen mensen.

Voor de toets als geheel

- 1 Is bij de onderwerpen die in de toets voorkomen rekening gehouden met verschillen in interesse en achtergrondkennis van meisjes en jongens, allochtonen en autochtonen? Meisjes hebben gemiddeld meer achtergrondkennis ten aanzien van menselijke relaties, gevoelens, opvoeding en verzorging, huishouden, literatuur, kunst en filosofie. Jongens weten meer over economische en politieke onderwerpen, over misdaad, sport, techniek en apparaten. Leerlingen verschillen op het gebied van culturele kennis, normen, waarden, gewoontes, godsdienst en ervaringen.
- 2 Komen er in het contextmateriaal en de situatieschetsen even vaak vrouwelijke als mannelijke personages voor? Wordt er ook verwezen naar leden van etnische minderheden? Worden in de items ook de ervaringen weergegeven van etnische minderheden en van vrouwen? Zijn deze ervaringen overwegend positief?
- 3 Worden vrouwelijke en mannelijke, allochtone en autochtone personages op dezelfde manier behandeld? Worden allochtonen voornamelijk in een positieve context gepresenteerd net als autochtonen? Verwijst men zowel naar vrouwelijke als naar mannelijke personages als men in het algemeen spreekt ('de verzekerde', 'de waarnemer', 'de kandidaat'). Worden vrouwelijke personages of leden van etnische minderheden niet als bijzondere gevallen of uitzonderingen voorgesteld?
- 4 Is de manier waarop leden van bepaalde bevolkingsgroepen worden weergegeven niet stereotyperend? Worden vrouwen ook met andere activiteiten in verband gebracht dan met huishoudelijke en verzorgende? Wordt er in de toets van uitgegaan dat vrouwen en leden van etnische minderheden in een breed scala van beroepen werkzaam kunnen zijn?
- 5 Als kinderen worden afgebeeld, is de jongen dan niet steeds ouder en/of groter dan het meisje?

Voor de afzonderlijke items

- 1 Staan er geen overbodige zinnen en gegevens in de vraagstelling? Is het contextmateriaal of de situatieschets functioneel met het oog op de te toetsen vaardigheid?
- 2 Is de vraagstelling duidelijk, ondubbelzinnig en expliciet? Weten de leerlingen wat er van ze verwacht wordt?
- 3 Is de taal die in het item en het bijbehorende contextmateriaal wordt gebruikt niet te moeilijk voor de kandidaten?
Is het item gescreend op de volgende aspecten:
 - Moeilijkheidsgraad van het gebruikte vocabulaire
 - Overbodig gebruik en/of moeilijkheidsgraad van functiewoorden, zoals voegwoorden

- Overbodige of moeilijke verwijzingen door middel van verwijzwoorden. Als verwijzwoorden gebruikt worden moeten ze duidelijk en ondubbelzinnig zijn, ook voor iemand die minder taalvaardig is
 - Uitdrukkingen of werkwoorden met voorzetsels die moeilijk te begrijpen zijn voor minder taalvaardige leerlingen
 - Moeilijke samengestelde woorden in het item
 - Worden er geen lange, samengestelde zinnen gebruikt in het item?
- 4 Gaat het item over een onderwerp waarvan kan worden verwacht dat een bepaalde groep leerlingen er meer over weet dan een andere? Bevat het item contextmateriaal over een onderwerp dat specifiek is voor één bepaalde cultuur, of over een bezigheid van één bepaalde groep (bijvoorbeeld voetbal)? Als dit het geval is, dient er een afweging plaats te vinden bij het samenstellen van de toets: het item kan verwijderd of gecompenseerd worden door een item met een tegenovergesteld effect.
 - 5 Bevat het item geen negatieve opmerkingen of conclusies over bepaalde bevolkingsgroepen in de stam of de alternatieven? Bijvoorbeeld dat een bepaalde groep niet voldoet aan de normen van de Nederlandse maatschappij of 'achter' is ten opzichte van andere.
 - 6 Bevat het item geen stereotiepe beschrijving van (leden van) bepaalde bevolkingsgroepen?
 - 7 Is het tekst- of contextmateriaal niet controversieel? Verwoordt de tekst bijvoorbeeld het standpunt dat bepaalde 'rassen' inferieur zijn aan andere of dat vrouwen in bepaalde opzichten minder competent zijn dan mannen? Indien dergelijk materiaal wel deel uitmaakt van de toets, moet dit een direct gevolg zijn van de toetsspecificatie.
 - 8 Indien de toets over onderwerpen gaat die betrekking hebben op minderheden of vrouwen, zijn de uitspraken die in de items voorkomen dan correct en overwegend positief? Wordt er impliciet of expliciet van een bepaalde 'norm' uitgegaan? Worden vrouwelijke prestaties of gedragingen vergeleken met die van mannen? Worden ontwikkelingen in andere landen beoordeeld volgens Nederlandse maatstaven (mate van welvaart, aantal analfabeten)?

Stereotypen en beschrijvingen die onacceptabel kunnen zijn (dit moet beoordeeld worden in de context waarin deze voorkomen)

Geen bevolkingsgroep dient te worden afgeschilderd als superieur of inferieur met betrekking tot

- bijdrage aan de maatschappij
- emotionele stabiliteit
- eerlijkheid
- ijver
- intelligentie
- leidinggevende capaciteiten
- moraal
- uiterlijke kenmerken
- lichamelijke eigenschappen

Voorbeelden van onacceptabele stereotypering van allochtonen

- vooral geschikt voor bepaalde beroepen
- minder geneigd tot werken
- leven vooral van uitkeringen
- spreken met een grappig accent
- hebben allemaal dezelfde cultuur en opvattingen (in tegenstelling tot leden van de dominante cultuur)
- beter in fysieke dan in intellectuele taken
- primitief
- onderontwikkeld

Voorbeelden van mogelijk onacceptabele benamingen en stereotyperingen van vrouwen

- alleen geschikt voor bepaalde beroepen
- emotioneel
- gauw in paniek
- zwak
- passief
- ijdel
- vrouwtje
- manwijf
- het zwakke geslacht
- beschrijving van uiterlijk, terwijl bij mannen eigenschappen en intellectuele capaciteiten worden beschreven.
- uitzonderingen op de (mannelijke) norm
- ondergeschikt aan mannen
- 'hij' of andere mannelijke woorden als ook vrouwen bedoeld worden. (In plaats daarvan kan men een ander woord kiezen, in plaats van mankracht bijvoorbeeld menskracht, het meervoud gebruiken 'zij' of 'je', 'men', of hij en zij afwisselen. Algemene termen, zoals 'arts', 'toeschouwer', 'verzekerde' worden geacht zowel mannen als vrouwen in te sluiten.)
- vrouwen altijd als tweede noemen in opsommingen.

Voorbeelden van mogelijk onacceptabele stereotyperingen van mannen

- alleen geschikt voor bepaalde beroepen
- ongeschikt om bepaalde huishoudelijke werkzaamheden te verrichten
- afhankelijk van vrouwen voor eten en verzorging
- kunnen niet met vrouwen samenwerken
- het sterke geslacht
- agressief
- dominant
- onderdrukkend