

### 3 Auditieve en audiovisuele teksten in luistertoetsen Nederlands

K. Schreuder

#### 1. Inleiding

Een luistertoets bestaat uit: gesproken tekst en daarbij behorende toetsvragen. Uit de antwoorden die kandidaten op deze toetsvragen geven, moet blijken in hoeverre zij de gesproken tekst hebben begrepen. Dat begrip kan begrip van de integrale tekst zijn of begrip op bepaalde aspecten ervan. De gesproken tekst wordt gewoonlijk vastgelegd op een audioband of een audiocassette. De verspreiding van videorecorders op de Nederlandse middelbare scholen maakt het nu, in 1990, mogelijk die teksten audiovisueel aan te bieden.

Hoe essentieel is dat verschil: teksten op video of teksten op audio? Enerzijds is hier sprake van een technische kwestie. Toen het Cito ruim 15 jaar geleden begon met de productie van luistertoetsen voor het vak Nederlands, is voor teksten op audioband of -cassette gekozen; niet omdat men audiovisuele teksten om redenen van validiteit afwees, maar eenvoudig omdat videoteksten nog niet tot de technische mogelijkheden behoorden. Scholen beschikten niet over de nodige afspelerapparatuur. De keuze was in het begin van de jaren 70 nog eerder: teksten op audioband aanbieden of teksten in de klas laten voorlezen door de docent. Maar er zit ook een theoretische kant aan de zaak. Interpersoonlijke communicatie is vaak audiovisueel. Puur auditieve communicatie komt voornamelijk voor bij de radio en de telefoon. Omdat men met een taalvaardigheidstoets afspraken over de vaardigheid van leerlingen in de gewone werkelijkheid buiten de toetsituatie wil doen, kan men tot op zekere hoogte proberen in die toets taal aan te bieden, die zoveel mogelijk lijkt op dat gewone taalgebruik. Tot op welke hoogte is omstrede, omdat er ook psychometrische eisen een rol spelen. Een volstrekt realistische aanbieding van tekstmateriaal in een taaltoets is bovendien vrijwel onmogelijk en toetstechnisch ongewenst. De meest realistische toets is niet automatisch ook de meest valide toets.

Dit artikel is gebaseerd op het uitgangspunt: het visuele aspect aan mondelinge taalsituaties is zo belangrijk, dat audiovisueel tekstmateriaal in een luistertoets de voorkeur verdient boven auditieve teksten. De keuze van auditieve teksten of audiovisuele teksten is een kwestie van begrips- en inhoudsvaliditeit en niet, zoals sommigen beweren, alleen van indrukvaliditeit. Dat geldt zowel voor tekstsoorten waarin het beeld aan de gesproken tekst eigen informatie lijkt toe te voegen als voor teksten met uitsluitend de spreker in beeld, zoals de lezing, waarin dat misschien in mindere mate gebeurt. Men moet bij gebruik van audiovisuele teksten het visuele een kans geven en een "videotoets" niet zien als een "geïllustreerde audiotoets". Het gaat bij gebruik van videoteksten niet om een traditionele luistertoets met plaatjes, maar om een audiovisuele toets met eigen kenmerken en mogelijkheden.

#### 2. Onderzoeksvraag

In ons onderzoek staat de vraag centraal:

Wat voegt video aan audio toe wat betreft tekstsoorten en vraagmogelijkheden in luistertoetsen Nederlands?

Het onderzoek wordt uitgevoerd op het Instituut voor Toetsontwikkeling (Cito) te Arnhem in het project Luistertoetsen Nederlands. In dat project worden schoolonderzoektoetsen voor luistervaardigheid voor met name LBO en MAVO en verder ook voor HAVO en VWO gemaakt. De scholen kunnen die toetsen kopen en gebruiken in hun schoolonderzoek. In het onderzoek gaan wij na of het voor het Cito zinvol is in de bestaande luistertoetsen van audioteksten over te gaan op videoteksten, dus of de nu gebruikte lezingen, interviews, radioprogramma's en rollenspelen vervangen kunnen worden door film- en videoprogramma's. Het onderzoek betreft in hoofdzaak: het maken van enkele experimentele luistertoetsen met teksten op video, het afnemen van die toetsen op relevante steekproeven van leerlingen, het analyseren van de resultaten en het doen van aanbevelingen voor de nabije toekomst.

#### 3. Tekstkeuze

Van belang is allereerst het inventariseren van mogelijke tekstsoorten en mogelijke vraaginhouden als eerste aanzet voor wat uiteindelijk een toetsmatrix zou moeten worden voor audiovisuele toetsen. Beide zaken, het kiezen van tekstsoorten en van vragen, hangen natuurlijk samen.

Wat de teksten betreft zijn in dit geval de voor toetsen gebruikelijke criteria als moeilijkheidsgraad, belangwekkendheid, representativiteit, bevrraagbaarheid van teksten en dergelijke niet de belangrijkste. Van meer belang zijn de volgende criteria, omdat die met name het verschil tussen auditieve en audiovisuele teksten betreffen.

Teksten moeten

1. verschillen in de relatie tussen beeld en geluid;
2. wisselende aantallen sprekende personen bevatten;
3. in werkelijkheidsgehalte variëren van documentaire tot nederlandsstalig drama.

Het belangrijkste criterium is het eerste. Dit eerste criterium is gebaseerd op literatuur over verwerking van auditieve en visuele informatie en in deze vorm afkomstig uit Heuvelman 1989. In die publicatie doet Heuvelman verslag van onderzoek naar optimale aanbiedingsvormen voor audiovisuele Teleac-programma's. Bij het zoeken naar de optimale programmvorm baseert Heuvelman zich onder meer op een herziene versie van de zogenaamde sommatiehypothese. Die hypothese heeft betrekking op meerkanalige informatieverwerking, bij voorbeeld via geluid en beeld.

Sommatiehypothese meerkanalige aanbieding van informatie

Twee kanalen kunnen aanbieden:

- a. redundante informatie
- b. aan elkaar gerelateerde informatie

- c. niet-gerelateerde informatie
- d. tegenstrijdige informatie

De experimenten rond deze hypothese hebben voornamelijk betrekking op aanbieding van woorden en afbeeldingen van de woordbetekenissen (foto's bij voorbeeld) in laboratoriumsituaties. Resultaat is onder meer, dat meerkanalige aanbieding van aan elkaar gerelateerde informatie (woorden en plaatjes van de woordbetekenissen) leidt tot informatiewinst: ppn. weten in een dergelijke conditie later meer items te herkennen dan bij enkelkanalige aanbieding. Dat resultaat is weergegeven in een herziene sommatiehypothese.

Herziene sommatiehypothese meerkanalige aanbieding van informatie

- a. bij meerkanalige aanbieding van woord en afbeelding: informatiewinst
- b. bij meerkanalige aanbieding van woord (gesproken en gedrukt): geen informatiewinst
- c. bij meerkanalige aanbieding van woord en niet-gerelateerde afbeelding: informatieverlies
- d. bij enkelkanalige aanbieding van een woord of een afbeelding: betere informatieoverdracht dan bij c, even veel informatieoverdracht als bij b, slechtere informatieoverdracht dan bij a.

Centraal staan hier de begrippen "gerelateerde" en "redundante" informatie. Nu is bij woorden en plaatjes nog wel vast te stellen of twee vormen van informatie aan elkaar gerelateerd zijn, redundant zijn of tegenstrijdig, maar bij levende film- en televisiebeelden is dat veel moeilijker. Bij film en televisiebeelden kan men volgens Heuvelman die begrippen beter vervangen door een begrip als "afstemming". Beeld en geluid moeten op elkaar zijn afgestemd om informatiewinst te bewerkstelligen boven aanbieding alleen in beeld of alleen in geluid.

Afstemmingshypothese audiovisuele informatie.

- a. als beeld en geluid op elkaar zijn afgestemd, treedt op: informatiewinst
- b. als een spreker in beeld is, is er geen sprake van winst of verlies ten opzichte van enkelkanalige (auditiële) informatieaanbieding
- c. als beeld en geluid niet op elkaar zijn afgestemd of elkaar zelfs tegenspreken, treedt informatieverlies op.

Volgens Heuvelmans hypothese is bij film en video dus sprake van een neutrale afstemming tussen beeld en geluid, wanneer een spreker in beeld is. De meerkanalige aanbieding van informatie leidt in dit geval niet tot meer begrip bij kijkers dan bij enkelkanalige aanbieding van het geluid alleen. Wanneer er sprake is van voice over, dus van een of meer stemmen achter beelden, dan kan de meerkanalige aanbieding informatiewinst of verlies ten opzichte van enkelkanalige aanbieding betekenen. Dat hangt af van de afstemming tussen beeld en geluid. Bij hoge afstemming treedt winst op, bij ontbrekende afstemming treedt informatieverlies op, bij voorbeeld omdat beeld en geluid elkaar tegenspreken.

Deze afstemmingshypothese lijkt belangrijk genoeg om er bij de tekstkeuze voor de experimentele toetsen rekening mee te houden. De te gebruiken teksten moeten een gevarieerde verhouding tussen beeld en geluid te zien geven, om alle mogelijkheden die uit de hypothese voortvloeien te kunnen onderzoeken. Vooral op grond van deze overweging en voorts ook op grond van de twee andere criteria zijn vier teksten als uitgangsmateriaal gekozen:

- een televisiediscussieprogramma (omdat daarin twaalf verschillende personen aan het woord komen, dus veel meer dan op een audioband van elkaar te onderscheiden zijn door leerlingen)
- een lezing (met een half uur lang de spreker in beeld)
- een speelfilm (met 10 sprekende personen in een fictieve context)
- een documentaire (omdat daarin de relatie tussen beeld en geluid allerlei varianten kent).

De onderzoeksresultaten die verder in dit artikel ter sprake komen, hebben betrekking op de toets bij de laatste tekst, de documentaire.

#### 4. De videotoets bij 'Een zaak van niveau'

De gekozen documentaire is *Een zaak van niveau* van Louis van Gasteren, uit 1990. Deze documentaire gaat over 1000 jaar strijd van de Nederlanders tegen het water. Hij is op 2 januari 1990 door de Nederlandse televisie uitgezonden. In de loop van 1990 zijn aan de film drie kunstrijzen toegekend. Het gaat hierbij derhalve om een film van uitzonderlijke klasse, een reden te meer (naast de boven genoemde) om de film als uitgangspunt voor een toets te nemen.

Van de film zijn vier fragmenten van in totaal 29 minuten lengte geselecteerd voor de toets. De hele film duurt 57 minuten. Bij deze fragmenten zijn 37 toetsvragen gemaakt, gewone tekstbegripvragen naar hoofdedachten van onderdelen, tekstrelaties, conclusies uit combinaties van gegevens en belangrijke details. De vragen zijn afgestemd op het niveau vierde klas MAVO en LBO, D-niveau. Die vragen zijn als volgt in te delen naar de plaats waar het goede antwoord is te vinden.

Vraagcategoriën Een zaak van niveau

Het goede antwoord staat

1. expliciet in het geluidsspoor met de spreker in beeld
2. expliciet in het geluidsspoor achter direct ondersteunende beelden
3. expliciet in het geluidsspoor achter indirect ondersteunende beelden
4. expliciet in beelden en impliciet in het geluidsspoor
5. impliciet in beelden en geluidsspoor.

In de eerste drie categorieën is Heuvelmans afstemmingshypothese te herkennen. De categorieën 4 en 5, die in Heuvelmans onderzoek niet voorkwamen, vergen wellicht enige toelichting. In categorie 4 zitten de vragen die gesteld zijn naar aanleiding van filmpassages zonder gesproken tekst; daarbij geven alleen de beelden dus de antwoorden aan de leerlingen die de toets afleggen. Het blijkt echter steeds zo te zijn, dat beelden op zichzelf genomen poly-interpretabel zijn. De ene interpretatie waarop gemikt wordt in het goede antwoord op een vraag komt alleen tot stand door

gesproken tekst uit de rest van de film erbij te betrekken. Dat wil zeggen: vragen die uitsluitend in het beeld worden beantwoord, komen niet voor: zij zijn niet van een objectief vaststelbaar goed antwoord te voorzien. (Omstreden is of dergelijke vragen in een luistertoets thuishoren; dat is een kwestie die hier buiten beschouwing wordt gelaten.)

Categorie 5 bevat vragen waarbij het goede antwoord uit de film moet worden afgeleid. Dat zijn o.a. vragen naar (evidente maar onuitgesproken) bedoelingen van de filmmaker.

In de constructieperiode zijn eerste versies van de toets gemaakt door 230 leerlingen van vierde klassen MAVO en LBO op C- en D-niveau. In die pretest bleken veel van de vragen veel te moeilijk. Op grond van de pretestresultaten zijn de vragen gerevisieerd en klaar gemaakt voor de experimentele toetsafneming. (De 37 vragen golden samen als "ruwe toets". Op grond van de afnemingsresultaten is uiteindelijk een definitieve toets van 26 items samengesteld. Zie par. 6)

### 5. Het afnemen van de toets

De toets is in het kader van het onderzoek als volgt afgenomen. Opzet experimentele toetsafneming

	MAVO 4	HAVO 4	VWO 4
Video	264	294	271
Audio	176		200
Controle	289		269

De cellen van de opzet zijn aangegeven met de aantallen uiteindelijk verwerkbare (en feitelijk ook verwerkte) antwoordbladen.

Per deelnemende school mocht in principe één klas de toets afleggen. Per cel hebben tussen 9 en 13 klassen meegedaan. Wij nemen aan dat door het vrij grote aantal ppn. de groepen als gelijkwaardig te beschouwen zijn. Zeker is dat niet. Voor de matching van de groepen is het bij een design als dit gewoonlijk optimaal, om deelnemende klassen at random te verdelen in evenveel groepen leerlingen als er profgroepen of condities zijn, hier drie: video, audio en controle. Maar bij een klassikaal af te nemen toets als deze ( met een collectief te bekijken of te beluisteren band) is dat praktisch onmogelijk: het valt niet te controleren of deelnemende scholen zich dan precies aan de gecompliceerde instructies daarvoor houden. Alle scholen zelf bezoeken is te arbeidsintensief.

De toets was weliswaar bestemd voor vierde klas MAVO en LBO, maar om zo precies mogelijk de moeilijkheidsgraad te kunnen vaststellen is de toets ook afgenomen op vierde klassen HAVO en VWO. LBO is als pretestgroep afgefallen, vanwege de onduidelijkheid die daar tijdens het schooljaar, voor het examen dus, bestaat over het niveau van de leerlingen.

"Audio" wil het volgende zeggen. Omdat wij met name geïnteresseerd waren in het verschil tussen video- en audiotolisen, is van de toets ook een audioversie gemaakt. Dat wil zeggen dat het geluidspoor van de videoband integraal overgenomen is op audiocassettes. De leerlingen in deze groep maakten dezelfde vragen als die in de video-groep.

Datzelfde geldt ook voor de controlegroep: dat waren leerlingen die alleen de vragen maakten, zonder de film te zien of de audioband te beluisteren. Deze laatste groep is ingesteld om bij de analyses de verschillen tussen de scores in de video- en de audioconditie beter te kunnen beoordelen. Bovendien kan men zo bij sommige vragen inzicht krijgen in de voorkennis die leerlingen van het onderwerp hebben. Gezien het onderwerp van de film (Nederland en het water) kan voorkennis een flinke rol spelen.

Deze proefopzet is gemaakt mede naar het voorbeeld van de experimenten van Heuvelman. Hij vergeleek behalve video-, audio- en controleversies nog veel meer versies van nieuwe Teleac-programma's met elkaar zoals: versies met en zonder presentator in beeld, met en zonder op het scherm geprojecteerde titels, met realistische beelden en met schematische beelden en met of zonder zogenaamde visuele analogieën. Voor wat betreft het verschil tussen video- en audioversies vond hij als resultaat: bij open vragen scoorden ppn. in de audioconditie hoger dan in de videoconditie. Bij meerkeuzevragen scoorden zij in de audioconditie iets lager dan in de videoconditie: de gemiddelde p-waarden waren .66 in de audio- en .68 in de videoconditie. Wij verwachten met de door ons gestelde vierkeuzevragen in ieder geval voor de totale toets een groter verschil tussen audio en video te verkrijgen, met in de videoconditie steeds de hoogste score.

### 6. Resultaten

In hoeverre genoemde verwachting betreffende het scoreverschil in video- en audioconditie uitkwam, laat het volgende overzicht zien.

Overzicht P-waarden en KR-20-waarden bij 26 items

	P-waarde	KR20-waarde
MAVO-VIDEO	.69	.56
MAVO-AUDIO	.61	.58
MAVO-CONTROLE	.40	.35
HAVO-VIDEO	.77	.62
VWO-VIDEO	.81	.56
VWO-AUDIO	.71	.61
VWO-CONTROLE	.56	.25

Toelichting. Uit de ruwe toets van 37 vragen zijn er 26 geselecteerd voor de definitieve toets. Bij de selectie zijn de resultaten voor MAVO aangehouden. De volgende selectiecriteria zijn gehanteerd. Verwijderd zijn vragen

- met hogere p-waarden in de controlegroep dan in de videoconditie (6 vragen)
- die in de video-groep ongeveer even hoog scoorden als in de controlegroep (2 vragen)

- met extreme p-waarden in de videoconditie, dus te moeilijke of te gemakkelijke vragen (2 vragen)
- met lage rir-waarden in de videoconditie (1 vraag)

Uit de resultaten over deze 26 items blijkt het volgende.

- De videotoets is wat moeilijkheidsgraad betreft het meest geschikt voor MAVO 4.
- Er is verschil tussen video- en audioconditie, zowel bij MAVO als bij VWO en in beide gevallen ook steeds in dezelfde richting: video scoort hoger dan audio.
- De betrouwbaarheid bij de MAVO-leerlingen komt, wanneer de toets met de Spearman-Brown-formule verlengd wordt tot 50 items (voor Cito-toetsen een gangbare lengte) net boven de .70; dat is acceptabel.
- De score van de controlegroep is .40 bij MAVO. Dat is iets hoger dan op grond van eerdere ervaringen verwacht kon worden. Bij VWO is de score in de controlegroep .56 en dat is veel te hoog.
- De betrouwbaarheid is zowel bij MAVO als bij VWO laag in beide controlegroepen (resp. .35 en .25); dat is volgens de verwachting. Vragen beantwoorden zonder tekst is blijikbaar een weinig consistente vaardigheid.

De verschillen tussen de cellen zijn alle significant. Voorzover dat de verschillen met de controlegroep betreft is dat uiteraard niet verwonderlijk: de 26 items zijn mede naar die verschillen toe geselecteerd. Maar belangrijker zijn de verschillen tussen de schooltypen en die tussen de video- en de audioconditie. Er is variantie-analyse uitgevoerd met het BMDP-pakket. De resultaten zijn de volgende.

#### Variantie-analyse

##### 1 Verschillen MAVO-HAVO-VWO Video

Ppn.: 829

Schooltype	DF	F-waarden	P
Error	2	74.17	0.000

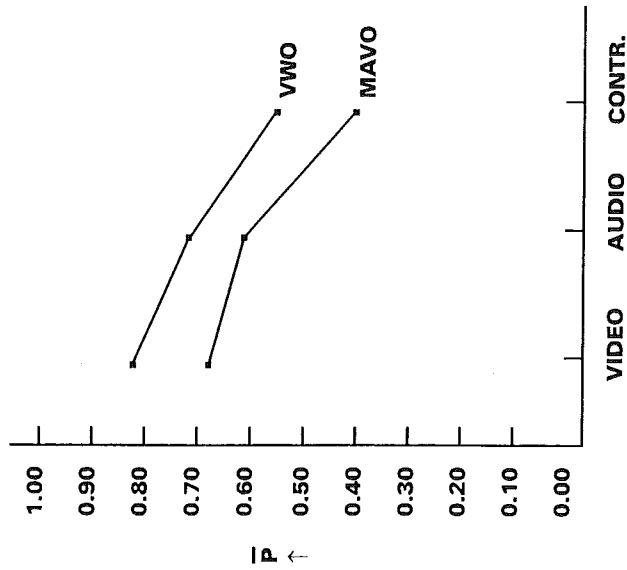
##### 2 Verschillen MAVO-VWO in drie condities

Ppn.: 1469

Schooltype	DF	F-waarden	P
Conditie	1	395.67	0.000
Interactie	2	699.64	0.000
Error	2	6.99	0.009
	1463		

De geringe interactie tussen schooltype en conditie is duidelijker zichtbaar in onderstaande grafiek. De twee lijnen lopen niet parallel. Het scoreverschil tussen MAVO en VWO is in de audioconditie iets kleiner dan in de video- en controleconditie.

Grafiek. Interactie schooltype - conditie



Wij waren met name geïnteresseerd in het verschil tussen video- en audioconditie: voor de toets als geheel en voor de vijf vraagcategorieën afzonderlijk. De resultaten per categorie waren de volgende.

P-waarden vijf vraag categorieën MAVO	MAVO-VIDEO	MAVO-AUDIO
Aantal vragen		
Categorie		
1	4	.66
2	5	.81
3	7	.65
4	6	.62
5	4	.69

Zoals verwacht mocht worden is het verschil het grootst in categorie 4: de categorie vragen met het goede antwoord voornamelijk in de beelden. Het verschil is in categorie 1 (spreker in beeld) kleiner dan in categorie 2 (voice over bij direct ondersteunende beelden), maar even groot als in categorie 3 (idem bij indirect ondersteunende beelden). Een duidelijke positie van categorie 1 geeft dit resultaat niet.

Het onderzoek naar het verschil tussen video- en audioversies van luistertoetsen zal worden herhaald bij de overige drie toetsen. Naar verwachting zal het verschil bij de lezing (spreker in beeld) klein zijn. Bij de twee andere toetsen (met als teksten het discussieprogramma met 12 sprekers in beeld en de speelfilm) is de grootte van het verschil moeilijk te voorspellen.

Men kan zich afvragen of grote verschillen nodig zijn om overgang van audioteksten naar videoteksten in luistertoetsen te verantwoorden. Ook bij geringe verschillen kan gebruik van audiovisueel materiaal verdedigd worden tegenover gebruik van enkelkanalige audiotieve teksten en wel om redenen van validiteit. Dramatische verschillen tussen video- en audioversies zouden er wel toe kunnen bijdragen dat de overstap ook inderdaad gemaakt wordt: zij zouden de beslissende stimulans ervoor kunnen zijn om het geld dat voor videoversies nodig is ook feitelijk uit te geven. Want dat onderzoeksresultaat is in ieder geval onomstreden: video is vele malen duurder dan audio.

#### Bibliografie

Heuvelman, A.  
1989 *Buiten beeld*. Lisse.

## 4 Signalering van leesproblemen d.m.v. tekstbegrip-toetsing?

H.I. Hacquebord

Veel leerlingen in het voortgezet onderwijs hebben problemen met het (bestude-rend) lezen van hun schoolboeken. Het achterliggende leesproces, de wijze waarop de tekstuele informatie wordt verwerkt en geïntegreerd in bestaande kennisstruc-turen, onttrekt zich echter grotendeels aan de waarneming. Instrumenten voor het meten van tekstbegrip betreffen immers veelal produktmetingen.

Het verwerkingsproces van de lezer op basis van een propositionele analyse van de tekststructuur (Van Dijk & Kintsch, 1983) is uitgangspunt geweest bij de ontwikkeling en constructie van toetsen voor tekstbegrip (Hacquebord, 1989; Galema, 1989), die behalve produktmeting ook een signalering van leesproblemen willen zijn. De vraag die in deze bijdrage zal worden gesteld is of en in hoeverre dit instrument een dergelijk diagnostische functie kan waarmaken. Toetsgegevens uit onderzoek zullen worden gelegd naast observatiegegevens m.b.t. leesproblemen van verschillende groepen 'probleemlezers'.

### 1. Kader en vraagstelling

In ons inmiddels afgesloten onderzoek naar tekstbegripproblemen van anderstalige en Nederlandstalige leerlingen in het voortgezet onderwijs (Hacquebord, 1989) zijn toetsen voor tekstbegrip ontwikkeld. Deze bleken betrouwbare instrumenten voor tekstbegripmeting bij Nederlandstalige én anderstalige leerlingen, representatief wat betreft tekstsoort en moeilijkheidsgraad en bovendien objectief en efficiënt af te nemen. Niet alleen in onderzoek, maar ook in de onderwijspraktijk zouden deze toetsen kunnen worden gebruikt, en wel voor signalering en mogelijk ook diagnos-tisering van taal- en leesproblemen bij leerlingen in het algemeen en bij andersta-ligen in het bijzonder. Mijns inziens is er bij scholen voor voortgezet onderwijs een duidelijke behoefte aan zo'n instrument. Veel scholen proberen in de brugklas leerlingen te selecteren die extra ondersteuning nodig hebben, onder andere op het gebied van de taalbeheersing. Dit gebeurt meestal in het kader van "Remedial Teaching".

In hoeverre leerlingen met specifieke taalproblemen, bijvoorbeeld anders-taligen, er bij een dergelijke screening ook uit rollen, is sterk afhankelijk van het gekozen instrumentarium. Een voorbeeld: op een school waar wij onderzoek wilden doen bij verschillende groepen leerlingen in ondersteuningslessen, waaronder anderstaligen, bleken deze laatsten nauwelijks geselecteerd te worden met de toetsinstrumenten die gebruikt werden door de school. Toen wij aan dit toetsinstru-mentarium van de school een eenvoudige receptieve woordenschattoets mochten toevoegen, bleken ineens vele anderstalige leerlingen hierop uit te vallen. Op toetsen voor technisch en begrijpend lezen of voor spelling deden de anderstaligen het niet opmerkelijk slechter dan hun Nederlandstalige medeleerlingen. De conclusie dat het