

De T2-specificiteit van het Centraal Schriftelijk Eindexamen Engels onderzocht

MARION TILLEMA-KORTMAN

In dit artikel wordt verslag gedaan van een experiment waarin de T2-specificiteit van het Centraal Schriftelijk Eindexamen Engels voor vwo wordt onderzocht. Dit wordt gedaan door ijking van de examens aan in het Nederlands vertaalde examenteksten. Aan het experiment namen 61 leerlingen uit 5-vwo deel. De centrale vraag luidt: *Meten de examens voornamelijk de Engelse taalbekwaamheid of spelen andere factoren ook een rol?* Uit de resultaten blijkt dat de examens Engels voor vwo niet optimaal T2-specifiek zijn. Toch kan het huidige examen een geschikt toetsingsinstrument zijn, mits aan enkele voorwaarden wordt voldaan.

In Nederland bestaat het Centraal Schriftelijk Eindexamen voor de moderne vreemde talen voor alle schooltypen uit een leesvaardigheids-toets. Deze toets heeft een aandeel van vijftig procent in het totale eindcijfer voor het schoolvak. De slagingskansen van scholieren in het voortgezet onderwijs hangen sterk af van hun eindexamencijfer. Zo demonstreren Van den Bergh, Zwarts en Rohde (2003) dat als enkel het Engelse eindexamencijfer voor havo-leerlingen een halve punt opgewaardeerd wordt, het totale slagingspercentage voor het havo

landelijk al stijgt met ruim 1%. Het is dan ook uitermate belangrijk dat de examens datgene meten wat ze behoren te meten.

In het geval van de moderne vreemde talen (hier aangeduid met T2) is dit bekwaamheid in, en kennis van de taal in kwestie. We kunnen ook zeggen: de examens moeten voldoende T2-specifiek zijn. Andere, niet ter zake doende factoren mogen geen rol van betekenis spelen.

Hieronder wordt verslag gedaan van een onderzoek naar de T2-specificiteit van de examens. Eerst wordt, door middel van literatuurstudie, bepaald welke factoren een rol spelen bij het lezen. Van elk van deze factoren zal worden bepaald in hoeverre deze invloed zouden mogen hebben op de examenprestaties. Vervolgens zal verslag worden gedaan van een experimenteel onderzoek naar de T2-specificiteit van de examens. Vanwege temporele en onderzoekstechnische beperkingen concentreert dit onderzoek zich op het Centraal Schriftelijk Eindexamen Engels voor het vwo.

Het construct leesvaardigheid

Binnen de huidige theorieën op het gebied van leesvaardigheid wordt aangenomen dat

lezen voor tekstbegrip¹ twee soorten cognitieve processen omvat, die elkaar onderling beïnvloeden. In de eerste plaats onderscheidt men zogenaamde ‘lagere’ of linguïstische leesprocessen. Men denke dan bijvoorbeeld aan lexicale, grammaticale en orthografische kennis. Daarnaast onderscheidt men de ‘hogere’ leesprocessen, waartoe ook de zogenaamde leesstrategieën worden gerekend. Voorbeelden hiervan zijn het zoeken naar sleutelfragmenten en -woorden (Bimmel 1999, 20), het monitoren van het eigen begrip van de tekst tijdens het lezen (Hulsker 2002, 16-17), het aanwenden van kennis van de wereld (Hulsker 2002, 16), et cetera.

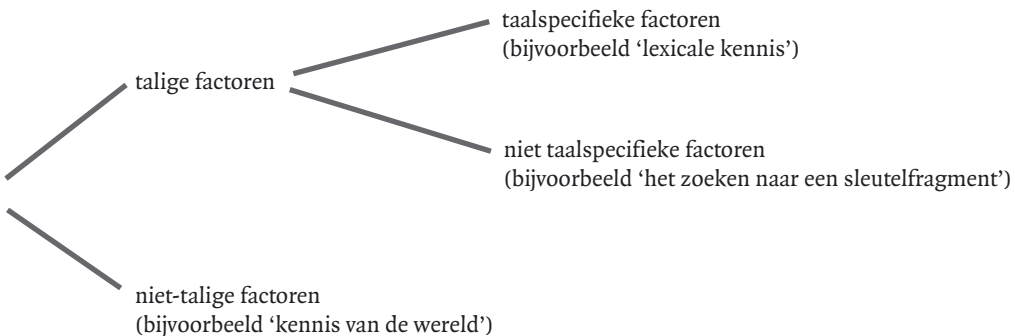
Een factor die ook vaak van belang wordt geacht voor leesvaardigheid is intelligentie. Groot en Bekkers (1987, 612) leggen uit dat het onmogelijk is om de factor intelligentie, en met name verbale intelligentie, uit te sluiten bij het toetsen van leesvaardigheid op hogere niveaus. Volgens hen is taalmateriaal op hogere niveaus, zoals het vwo-niveau, conceptueel complex, zowel op zins- als op tekstniveau. Om dit complexe materiaal te verwerken is een redelijke mate van (verbale) intelligentie nodig.

Overigens is het waarschijnlijk dat veel van de genoemde factoren met elkaar zullen

samenhangen. ‘Kennis van de wereld’ zal bijvoorbeeld samengaan met intelligentie.

Binnen de groep van bovengenoemde factoren valt een onderscheid te maken tussen talige en niet-talige factoren. Binnen de talige categorie vallen alle ‘lagere’ processen en een aantal ‘hogere’ processen, zoals het zoeken naar sleutelfragmenten en -woorden. Het aanwenden van kennis van de wereld valt in de categorie van niet-talige factoren. Ook de factor intelligentie is in beginsel niet-talig, hoewel enige samenhang tussen intelligentie en taalbekwaamheid niet ondenkbaar is (Groot en Bekkers 1987, 12). Binnen de talige categorie kan opnieuw onderscheid worden aangebracht, namelijk tussen taalspecifieke en niet-taalspecifieke factoren. De ‘lagere’ processen zijn taalspecifiek. Dat wil zeggen dat ze verschillen per taal. De meeste talige ‘hogere’ processen zijn niet taalspecifiek. Afgezien van kleine verschillen tussen taalgebieden omtrent de opbouw van teksten, is het zoeken naar sleutelfragmenten in Nederlandse teksten dezelfde vaardigheid als het zoeken naar sleutelfragmenten in Engelse teksten.

Zie figuur 1 voor een schematisch overzicht van de hierboven besproken indeling van factoren.



Figuur1: Factoren die van belang zijn voor de leesvaardigheid

Strikt genomen houdt de definitie van T2-specificiteit in, dat examens enkel een beroep zouden mogen doen op de beheersing van taalspecifieke vaardigheden. We kunnen ons echter afvragen of dat realistisch en wenselijk is. Zo er al een toets bestaat die zuiver kennis van de vreemde taal meet, dan is het de vraag in hoeverre de uitslag van een dergelijke toets voorspelt hoe bekwaam de getoetste leerling is in de T2 en hoe vaardig hij of zij zal zijn in realistische communicatieve situaties in de T2.

Een bovenmatig grote rol mogen niet-taalspecifieke talige factoren echter niet spelen. Als taalafhankelijk factoren de overhand hadden in de totstandkoming van de examen-score, dan zou er één gezamenlijk examen voor alle moderne vreemde talen kunnen worden ingesteld; ze meten immers toch dezelfde vaardigheid!

Het moge tenslotte duidelijk zijn dat een te grote rol voor niet-talige factoren onwenselijk is. Het doel van het examen is immers te bepalen of het taalvaardigheidsniveau in de betreffende taal toereikend is.

Kortom, enige invloed van factoren die niet-taalspecifiek, maar wel degelijk talig zijn is niet bezwaarlijk. Primair moeten de scores voor de examens, in dit geval de examens Engels voor het vwo, echter bepaald worden door taalspecifieke factoren. Het is dus van belang om vast te stellen hoe groot de invloed van niet-taalspecifieke factoren is ten opzichte van de invloed van taalspecifieke factoren. Hulsker (2002) onderzocht de samenhang tussen een aantal gespecificeerde factoren, waaronder niet-taalspecifieke factoren, en de scores op leesvaardigheidsexamens Engels voor het havo. Uit haar analyse bleek dat, van alle onderzochte factoren, de taalspecifieke 'lagere' factoren 'woordenschat' en 'grammaticale kennis' het sterkst correleerden met tekstbegrip. Van de 'hogere' factoren die niet-taalspecifiek waren, correleerde 'begrip van verbanden en structuur' het sterkst met tekstbegrip.

Hieronder wordt verslag gedaan van een experiment waarmee de invloed van alle niet-taalspecifieke factoren bepaald kan worden. Op deze manier is het mogelijk om vast te stellen hoe niet-taalspecifieke componenten bijdragen aan de scores voor de examens Engels voor het vwo.

Onderzoeksopzet

De hoofdvraag van het onderhavige onderzoek luidt: Zijn de Cito-examens Engels voor het vwo voldoende T2-specifiek of is er meer voor nodig dan slechts bekwaamheid in de T2 om goed te scoren op deze examens?

Groot (1996, 37) geeft uitleg over één van de manieren waarop deze vraag kan worden beantwoord: moedertaaliking. Een vanuit theoretisch oogpunt ideaal onderzoeksprocedé zou als volgt verlopen. Eerst zou men een groep proefpersonen de oorspronkelijke examenteksten laten maken. Vervolgens zou men willen weten hoe dezelfde proefpersonen op dezelfde teksten in het Nederlands zouden scoren. Wanneer men echter dezelfde teksten, weliswaar vertaald, maar met dezelfde inhoud en structuur, nog eens presenteerde aan dezelfde leerlingen, dan zouden de uitkomsten worden beïnvloed door een leer- of geheugeneffect. Eigenlijk zou men op dit punt dus het geheugen van de proefpersonen willen wissen om vervolgens de vertaalde examenteksten af te nemen.

Groot (1996, 37) stelt een realistischer onderzoeksopzet voor, waarin men niet tweemaal dezelfde groep proefpersonen hoeft aan te spreken. Hij legt uit dat een moedertaalikingsexperiment op twee manieren ingericht kan worden. De eerste manier is om een bepaalde examentekst af te nemen op zowel een steekproef uit de doelgroep als op een vergelijkbare groep moedertaalsprekers van de vreemde taal in kwestie. De tweede manier is om een bepaalde examentekst, waarvan

de helft is vertaald in de moedertaal van de doelgroep, af te nemen op (een steekproef uit) de doelgroep. Vanwege praktische overwegingen is het hieronder beschreven onderzoek ingericht volgens het laatstgenoemde stramien.

Indien de Cito-examens Engels voor het vwo voldoende T2-specifiek zijn, dan is het te verwachten dat de scores behaald op de Nederlandstalige tekstgedeelten significant hoger liggen dan de scores behaald op de originele Engelstalige tekstgedeelten. De moeilijkheid van een T2-specifieke toets is immers gelegen in de vreemde taal. Deze moeilijkheid wordt opgeheven door vertaling in het Nederlands. Verder valt, bij optimale T2-specificiteit, te verwachten dat de Nederlandstalige scores niet verschillen van de maximum haalbare score².

PROEFPERSONEN

Aan het experiment namen 61 leerlingen uit 5-vwo deel, afkomstig van twee scholen. Het experiment werd aan het einde van de maand juni uitgevoerd. De leerlingen bevonden zich dus in de laatste fase van hun vijfde leerjaar.

MATERIALEN

De deelnemende proefpersonen werd gevraagd om een leesvaardigheidstoets te doen. Deze toets bestond uit vier teksten uit vwo-examens Engels uit voorgaande jaren³, samen met 32 vragen, waarvan 29 in multiple-choicevorm en drie in de vorm van een open vraag. De complete toets zou twee lessen beslaan.

Alleen relatief gemakkelijke⁴ teksten werden gebruikt, aangezien in het onderhavige experiment met leerlingen uit 5-vwo werd gewerkt.

Bij de betrokken docenten werd gecontroleerd of het toetsmateriaal bij de leerlingen niet bekend was.

De Engelstalige teksten en vragen werden vertaald in het Nederlands. Vanwege de mogelijkheid dat de ene teksthelft gemakkelijker of moeilijker is dan de andere teksthelft, werden twee toetsversies geconstrueerd. In toets A werd elk van de vier teksten aangeboden in de volgorde Nederlands-Engels. In toets B werd voor elke tekst de volgorde Engels-Nederlands aangehouden. In tabel 1 is een schematische weergave van de onderzoeksopzet te vinden.

Bij het vertalen is erop gelet dat het niveau van abstractie, de redeneertrant en de complexiteit van de originele Engelse teksten behouden bleef. Hierdoor werd equivalentie zo veel mogelijk behouden en waren het origineel en de vertaling dus optimaal vergelijkbaar. Een voorbeeld van een tekstfragment met vraag en de bijbehorende vertaling is weergegeven in kader 1.

PROCEDURES

Afname van de toetsen vond plaats tijdens reguliere lessen. De potentiële invloed van verschillen in de moeilijkheidsgraad tussen tekstgedeelten werd als volgt geneutraliseerd. De ene helft van de groep proefpersonen maakte de eerste helft van elke tekst in het Nederlands en de tweede helft van elke tekst in het Engels. Voor de andere helft van de groep was de volgorde precies omgekeerd: de eerste helft van elke tekst werd in het Engels gemaakt en de tweede in het Nederlands (hiermee Groot 1996, 37 volgend). Ook de vragen, behorend bij de teksten, waren op een dergelijke manier opgesplitst: de ene helft in het Nederlands, de andere helft in het Engels. Omdat elke examenvraag doorgaans betrekking heeft op één enkele alinea, was het mogelijk om de toetsen zo te construeren, dat de antwoorden op vragen die in het Nederlands werden gesteld te vinden waren in het Nederlandstalige tekstgedeelte. Hetzelfde gold voor Engelstalige vragen.

Een alinea uit de oorspronkelijke examentekst met bijbehorende vraag.

Elizabeth Haiken is assistant professor of history at the University of Tennessee and a high level of scholarly and thorough research is everywhere evident. This is not a populist book. It reads like a very well-written PhD thesis. The problem with such an approach, though, is that cosmetic surgery is a populist subject. Much as I rue the triumph of opinion over knowledge which characterises so much contemporary writing, this book errs so far in the opposite direction. The fascination of cosmetic surgery lies not in how it developed, but in why: what Venus Envy cries out for is conjecture.

6. What is the reviewer's opinion of Haiken's book, according to paragraph 6?

- A. Even though the book is aimed at an educated public, its academic style makes it inaccessible.
- B. The book abounds in information but fails to look into causes.
- C. The book's thoroughness makes up for its comparative lack of readability.
- D. Though intended as an objective study, the book suffers from unacademic bias.

Elizabeth Haiken is universitair docent geschiedenis aan de Universiteit van Tennessee en een hoog niveau van wetenschappelijk en grondig onderzoekswerk is overal duidelijk. Dit is geen populistisch boek. Je leest het als een goed geschreven proefschrift. Het probleem met een dergelijke aanpak is echter dat cosmetische chirurgie een populistisch onderwerp is. Hoe zeer ik de triomf van de mening over kennis, die zoveel recent werk karakteriseert, ook betreur, dit boek maakt de fout het tegenovergestelde te doen. Het fascinerende van cosmetische chirurgie zit hem niet in h oe het zich ontwikkelde, maar in waar om: wat Venusnijd oproept is giswerk.

6. Wat is de mening van de recensent over Haikens boek volgens alinea 6?

- A. Hoewel het boek is bedoeld voor een hoger opgeleid publiek, maakt de academische stijl ervan het ontoegankelijk.
- B. Het boek loopt over van informatie, maar laat na om naar oorzaken te kijken.
- C. De grondigheid van het boek maakt het relatieve gebrek aan leesbaarheid goed.
- D. Hoewel bedoeld als een objectieve studie, lijdt het boek aan een onwetenschappelijke eenzijdigheid.

Kader 1: Fragment uit de tekst 'I'm not sure it's quite me...' van Cressida Connolly, examen vwo Engels 2001, tijdvak 1, met daaronder de Nederlandse vertaling.

Tekst	Teksthelft	Toetsversie A	Toetsversie B
'I'm not sure it's quite me...'	teksthelft 1 (16 vragen)	Nederlands	Engels
	teksthelft 2 (16 vragen)	Engels	Nederlands
'Schools on alert as former pupils sue'	teksthelft 1 (16 vragen)	Nederlands	Engels
	teksthelft 2 (16 vragen)	Engels	Nederlands
'When one man's meat is another man's invasive surgery...'	teksthelft 1 (16 vragen)	Nederlands	Engels
	teksthelft 2 (16 vragen)	Engels	Nederlands
'The Fat of the Land'	teksthelft 1 (16 vragen)	Nederlands	Engels
	teksthelft 2 (16 vragen)	Engels	Nederlands

Tabel 1: Schematische weergave van de onderzoeksopzet

Resultaten

Alle gemaakte toetsen werden nagekeken met behulp van de officiële correctievoorschriften van het Cito. Voor elk goed antwoord werd één punt toegekend. De maximumscore per persoon per toetsversie bedroeg dus zestien punten.

Voor alle toetsversies werd de teststabiliteit, uitgedrukt in betrouwbaarheidscoëfficiënten, berekend. Betrouwbaarheidscoëfficiënten hebben altijd een waarde die ergens tussen nul en één ligt. Hoe dichter bij één, hoe hoger

de betrouwbaarheid. Bij een perfecte betrouwbaarheid zou een leerling, als hij of zij dezelfde toets nogmaals zou maken, in theorie (geheugen- en vermoeidheidseffecten uitgeschakeld) exact dezelfde score behalen.

Zoals uit tabel 2 kan worden opgemaakt, is er geen sprake van een hoge mate van betrouwbaarheid. De meest voor de hand liggende verklaring hiervoor is het relatief kleine aantal items per toetsversie (16). Vanwege het beperkte aantal beschikbare uren was het echter niet mogelijk om het aantal items per toets uit te breiden. Een groter aantal items per

	Ned. versie A	Ned. versie B	Eng. versie A	Eng. versie B
r (betrouwbaarheid)	.31	.54	.53	.41

Tabel 2: Betrouwbaarheidscoëfficiënten voor de Nederlands- en Engelstalige toetsen (versie A: volgorde Nederlands-Engels; versie B: volgorde Engels-Nederlands; 16 items per toetsversie)

Tekst	Nederlands		Engels		N	K
	gem.	s.d.	gem.	s.d.		
'I'm not sure it's quite me...'	64.2	23.2	55.0	27.2	61	100
'Schools on alert as former pupils sue'	76.0	25.8	67.3	29.5	59	100
'When one man's meat is another man's invasive surgery...'	76.6	27.4	73.3	30.3	61	100
'The Fat of the Land'	71.5	22.9	55.4	29.1	61	100
totaal	72.1		62.7			100

Tabel 3: Resultaten: gemiddelde score (gem.) als percentage van de maximum haalbare score K, standaarddeviatie (s.d.), aantal proefpersonen (N) en maximum haalbare score (K)

toets had hoogstwaarschijnlijk geresulteerd in hogere betrouwbaarheidscoëfficiënten⁵.

Tabel 3 geeft informatie over de gemiddelde scores (in de tabel weergegeven als percentages van de maximum haalbare score) en standaarddeviaties.

Voor de Nederlandstalige gedeelten behaalde men gemiddeld steeds een hogere score dan voor de Engelstalige gedeelten.

Om inzicht te krijgen in de verschillen tussen de scores, werd een variantieanalyse uitgevoerd. De Nederlandstalige scores bleken significant ($\alpha < .005$) hoger te liggen dan de Engelstalige scores. Bovendien bleek de variantie tussen teksten in het Engels groter dan in het Nederlands. In het Engels zijn de teksten dus moeilijker en liggen ze qua moeilijkheidsgraad verder van elkaar. Dit betekent dat de moeilijkheidsgraad van sommige Engelse teksten sterker toeneemt ten opzichte van het Nederlands dan die van andere⁶.

Discussie

Uit de resultaten komt een dubbelzinnig beeld naar voren. Enerzijds bleken de scores op het Nederlandstalige gedeelte significant hoger te liggen dan de scores op het Engelstalige gedeelte. Dit geeft aan dat kennis en beheersing van het Engels wel degelijk één van de beslissende factoren is.

Anderzijds blijkt uit de resultaten een groot verschil tussen de gemiddelde Nederlandstalige score en de maximum haalbare score. Dit suggereert dat behalve taalvaardigheid in de vreemde taal ook andere, niet-taalspecifieke factoren de scores beïnvloeden. Deze bevinding strookt met die van Gosse (1993). Zij stuitte in haar leesvaardigheidsonderzoek ook op problemen met de T2-specificiteit van de eindexamens. Nu zou men, terecht, kunnen tegenwerpen dat de perfecte score van 100% niet haalbaar is. Hier zijn verschillende rede-

nen voor aan te voeren. Zo is het denkbaar dat men bij het lezen in de moedertaal de toets onbewust enigszins onderschat. Men leest in dit geval zogezegd ‘met een half oog’, zoals men ook met een half oor kan luisteren, hetgeen de concentratie negatief zal beïnvloeden met alle mogelijke consequenties van dien, zoals het slordig lezen van vragen. Echter, zelfs als een maximum haalbare score van 90% wordt aangehouden is er sprake van een gat van 17,0%. Dit is bijna het dubbele van het verschil tussen de gemiddelde Engelstalige score en de gemiddelde Nederlandstalige score.

Tenslotte bleek dat de moeilijkheidsgraad van sommige teksten in het Engels sterker toeneemt ten opzichte van hun Nederlandse equivalenten dan die van andere Engelse teksten. Het scoreverschil tussen de Engelse en de Nederlandse versie van dezelfde tekst is voor sommige teksten dus (significant) groter dan voor andere teksten. In eerste instantie is men wellicht geneigd om deze toename in de verschillen tussen teksten toe te schrijven aan taalspecificiteit. Echter, het gebruik van bijvoorbeeld moeilijke, infrequente woorden in het Engels noopt tot het gebruik van strategieën als het afleiden van de woordbetekenis uit de context. Deze afleidstrategieën zijn nu weer niet taalspecifiek. We moeten derhalve concluderen dat de toename van verschillen in moeilijkheid zowel taalspecifieke als minder taalspecifieke kenmerken heeft.

Concluderend lijkt het gerechtvaardigd om, hoewel beheersing van het Engels zeker een factor van belang blijkt, te stellen dat de Cito-examens Engels voor het vwo niet optimaal T2-specifiek zijn. Naast taalbekwaamheid in het Engels spelen andere factoren, die niet T2-specifiek zijn, een rol.

Hoe nu verder?

Laten we aan de hand van één van de vragen uit het gebruikte toetsmateriaal eens trach-

ten na te gaan wat men moet kunnen om de vraag correct te beantwoorden. Als voorbeeld nemen we vraag 6, horend bij de tekst ‘I’m not sure it’s quite me’ van Cressida Connolly (kader 1). Het juiste antwoord is antwoord B.

Om het antwoord op deze vraag te vinden, moet de lezer inzien dat er een verband is tussen de uitspraak ‘this book errs so far in the opposite direction’ en de uitspraak ‘the fascination of cosmetic surgery lies not in how it developed, but in why: what Venus Envy cries out for is conjecture’, namelijk dat de tweede uitspraak een verduidelijking en concreetisering is van de eerste uitspraak. Hiermee stuiten we op een essentieel onderdeel van het karakter van lezen. Woorden op schrift verwijzen per definitie naar een bredere notie in de werkelijkheid. In een geschreven tekst worden gedachten per definitie op gecondenseerde wijze weergegeven. Veel stappen in die gedachtegang worden, vrijwel altijd onbewust, impliciet gelaten, omdat wordt aangenomen dat de lezer dezelfde zaken vanzelfsprekend acht als de schrijver.

Deze beperking van geschreven teksten is niet taalspecifiek. Men zou dus kunnen stellen dat lezen in elke taal meer vraagt dan enkel kennis van die taal. Voor de vwo-examens geldt bovendien dat de teksten van een hoog niveau zijn en een hoge mate van abstractie kennen. Algemene intelligentie zal daarom voor een deel het begrip van de tekst bepalen. Het resultaat van dit onderzoek, dat naast bekwaamheid in de T2 nog andere factoren een rol spelen, is met deze gedachte in het achterhoofd dan ook niet verwonderlijk.

Zolang het Centraal Schriftelijk Eindexamen bestaat uit een leesvaardigheidstoets, is het onvermijdelijk dat in de toetsing ook andere factoren dan T2-bekwaamheid, namelijk vaardigheden die voor lezen in alle talen nodig zijn, een rol spelen. De vraag is echter: hoe groot is deze rol? Het hierboven besproken onderzoek suggereert dat de rol van niet-taalspecifieke factoren nogal groot

is ten opzichte van de rol van taalspecifieke factoren. Het zou goed zijn de invloed van niet-taalspecifieke factoren in de toekomst te beperken tot een minimaal niveau. De rol van deze factoren mag nooit groter zijn dan die van taalspecifieke factoren. Verder dient men ervoor te waken dat een factor als 'kennis van de wereld' niet invloedrijker is dan, bijvoorbeeld, leesstrategieën. Niet-talige factoren mogen niet belangrijker zijn dan factoren die niet taalspecifiek, maar wel degelijk talig zijn.

Vragen wij van onze scholieren taalvaardigheid (zodat ze naast taalspecifieke kennis ook niet-taalspecifieke en niet-talige kennis kunnen aanwenden om 'in de echte wereld' communicatief uit de voeten te kunnen) of enkel taalkennis (zodat hun vaardigheid in reële communicatieve situaties ernstig wordt beperkt, doordat er slechts taalspecifieke kennis, maar geen niet-taalspecifieke en niet-talige kennis voorhanden is)? Indien we vinden dat van examenkandidaten taalvaardigheid mag worden verwacht, dan kan de huidige examenvorm een geschikt meetinstrument zijn. Hierboven is reeds uitgelegd dat enige invloed van niet-taalspecifieke factoren bij leestoetsen altijd onvermijdelijk is. Het terugdringen van de rol die deze factoren spelen ten opzichte van taalspecifieke factoren zal daarom een moeilijk karwei zijn. Slaagt men hier echter in, dan zijn de leestoetsen van het Cito een goed eindexameninstrument.

NOTEN

1. Met de term 'lezen' wordt in dit artikel steeds 'lezen voor tekstbegrip' bedoeld.
2. Deze voorspelling is niet statistisch te bewijzen of te ontcrachten. Aangezien de maximum haalbare score voor ieder proefpersoon even hoog is, is er geen sprake van spreiding. Dit betekent dat er geen variantieanalyse mag worden uitgevoerd om het verschil tussen de Nederlandstalige scores

en de maximum haalbare score te berekenen. Het woord 'significant' ontbreekt dan ook in de laatste voorspelling.

3. De gebruikte teksten waren:
Tekst 1: 'I'm not sure it's quite me...'/Ik weet niet zeker of het iets voor mij is...', examen vwo 2001, tijdvak 1.

Tekst 2: 'Schools on alert as former pupils sue'/Scholen alert: dagvaarding door oud-leerlingen', examen vwo 1999, tijdvak 1.

Tekst 3: 'When one man's meat is another man's invasive surgery...'/Wat vlees is voor de één, is voor de ander een inwendige operatie...', examen vwo 2000, tijdvak 1.

Tekst 4: 'The Fat of the Land'/De Vette Jaren', examen vwo 1999, tijdvak 1.

4. De moeilijkheidsgraad van de teksten werd vastgesteld aan de hand van door de Citogroep berekende gemiddelde scores per vraag.

5. De Citogroep berekent zelf ook de betrouwbaarheid van elke tekst die in een examen voorkomt. De resultaten hiervan zijn als volgt:

Tekst 1 ('I'm not sure it's quite me...'): $r = .48$ (9 testitems, 942 deelnemers).

Tekst 2 ('Schools on alert as former pupils sue'): $r = .39$ (7 testitems, 4000 deelnemers).

Tekst 3 ('When one man's meat is another man's invasive surgery...'): $r = .37$ (6 testitems, 4000 deelnemers).

Tekst 4 ('The Fat of the Land'): $r = .41$ (10 testitems, 4000 deelnemers).

Ook hier is dus geen sprake van hoge betrouwbaarheidscoëfficiënten. Het aantal items per toets is hier nog kleiner dan in het onderhavige onderzoek. Wederom had een groter aantal items waarschijnlijk geresulteerd in hogere betrouwbaarheidscoëfficiënten. (Met dank aan dhr. N. van Zuijlen van de Citogroep voor het ter beschikking stellen van deze gegevens.)

6. Ik dank Huub van den Bergh voor zijn hulp bij het uitvoeren van de analyses.

LITERATUUR

Bergh, H. van den, Rohde, E. & Zwarts, M. (2003). Is het ene examen het andere? Over de stabiliteit van schoolonderzoek en centraal examen. *Pedagogische Studiën*, 80, 176-191.

Bimmel, P. (1999). *Training en transfer van leesstrategieën. Training in de moedertaal en transfer naar een vreemde taal, een effectstudie bij leerlingen uit het voortgezet onderwijs*. 's Hertogenbosch: Uitgeverij Malmberg.

Gosse, L. (1993). *Foreign language specificity of reading comprehension tests*. Doctoraalscriptie Universiteit Utrecht. Utrecht: English Department.

Groot, P.J.M. & Bekkers, C. (1987). Tekstbegriptoetsing moderne vreemde talen en intelligentie. *Levende Talen Tijdschrift*, 484, 515-519.

Groot, P.J.M. (1996). T1- en T2-leesvaardigheid. Een contrastief/linguïstische benadering. In: Groot, P.J.M. & Nienhuis, L.J.A. (red.) *Leesvaardigheid in de eerste en in een tweede taal. Een vergelijking* (pp. 35-45). Utrecht: Centre for Language and Communication.

Hulsker, A. (2002). *Exploring FL reading problems. An experimental diagnostic test of English reading comprehension*. Doctoraalscriptie Rijksuniversiteit Groningen.

Tillema, M. (2003). *Cito leestoetsen voor vwo. T2-specificiteit en scharnierwoorden*. Doctoraalscriptie Universiteit Utrecht.