

Oratie 7 september 2006

Huub van den Bergh

Zeker weten
door zuiver meten?



Universiteit Utrecht
Faculteit Geesteswetenschappen
Letteren

Oratie

Uitgesproken bij aanvaarding van het ambt van bijzonder hoogleraar Didactiek en toetsing van het taalvaardigheidsonderwijs, vanwege de Vereniging van Leraren in Levende Talen op donderdag 7 september 2006.

Zeker weten door zuiver meten?

Meten is weten, tenminste dat wordt vaak gezegd. Hildo Wesdorp, van wie ik het vak geleerd heb, zei eens *'als de kwaliteit van het onderwijs net zo goed gecontroleerd werd als de kwaliteit van de kroketten, was de zaal nu leeg'*. Kortom, we moeten constant evalueren, niet alleen de kwaliteit van ons onderwijs, de kwaliteit van leerboeken, maar ook de kwaliteit in termen van leerlingprestaties. We moeten elke discussie over de kwaliteit van ons onderwijs een empirische basis geven. Pas dan kan de discussie vruchtbaar gevoerd worden, pas dan kunnen we gerichte maatregelen nemen om de kwaliteit van het onderwijs op micro-meso- en macroniveau te verbeteren. Dit staat buiten kijf. Als we dat niet doen, verzanden we in onderwijsvernieuwing op onderwijsvernieuwing, welke helaas lang niet altijd gerelateerd is aan een verbetering van ons onderwijs.

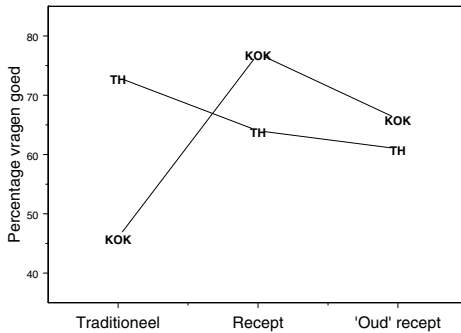
Maar, evalueren is niet gemakkelijk. Er zijn talloze valkuilen waar we vaak met open ogen intrappen. Aan de hand van één dictee weten we opeens dat eerstejaarsstudenten niet meer kunnen spellen. De vraag is dan 'kunnen we op grond van dat ene dictee generaliseren naar andere schrijfsituaties'? Weten we op grond van het aantal spelfouten in het Nationaal Dictee dat het met de spelling in Nederland en Vlaanderen beroerd gesteld is? Kortom, weten we zaken wel zo zeker als we ze gemeten hebben? Zit er in vrijwel elke vorm van 'meten' niet een enorme bron van onzekerheid? We hebben in

eerste instantie natuurlijk altijd toevalsfluctuaties; de meting zelf is nooit honderd procent betrouwbaar. Hier valt bij voorbeeld te denken aan de notoire onbetrouwbaarheid van de examens Nederlands in het VO; jaar in jaar uit is de betrouwbaarheid van de examens Nederlands beduidend lager dan die van andere talen. Is er iets inherent aan de Nederlandse taal dat van de examens zo'n onbetrouwbare meting maakt?

Maar, ik doel nu natuurlijk niet (alleen) op toevalsfluctuaties, maar vooral op meer conceptuele onduidelijkheden. Weten we echt hoe goed leerlingen kunnen lezen als ze een tekstbegriptoets gemaakt hebben? Of, weten we echt hoe goed leerlingen kunnen schrijven als zij net een zelfgeschreven tekst hebben ingeleverd? Dergelijke vragen worden veelal met 'nee' beantwoord. Deze onvrede wordt vaak vertaald in een afwijzing van het begrip 'vaardigheid', en een omarming van het begrip 'competentie'. Hiermee wordt veelal gedoeld op een toetsing in 'levensechte' situaties waarbij een multidimensionale vorm van toetsing centraal staat. Ik denk dat deze ruil van het begrip vaardigheid voor competentie niet zo verstandig is, en dat we daarmee weer eens het spreekwoordelijke kind met het badwater weggegooid zouden hebben. We hebben met deze ruil immers de hele theorie over het begrip vaardigheid buiten de deur gezet, zonder ons ook maar één moment af te vragen wat deze vaardigheidstheorie precies inhoudt, en zonder ons af te vragen wat het begrip lees- of schrijfvaardigheid nu precies omhelst.

Wat we nodig hebben is een theorie; een theorie waarin niet alleen gespecificeerd wordt wat we moeten meten, maar ook een theorie over evaluatie zelf. Ik zal proberen te laten zien dat dergelijke theorieën nauw met elkaar verbonden zijn, en dat de één niet kan zonder de ander.

Maar, laten we eerst eens kijken naar een voorbeeld. Twee groepen leerlingen, één van de theoretische leerweg in het VMBO en één van de basisberoepsgerichte leerweg van hetzelfde opleidingstype maakten elk drie tekstbegriptoetsen. Leerlingen in de theoretische leerweg, de oude MAVO, zouden het naar verwachting beter moeten doen dan de basisberoepsgerichte leerlingen. Voor de eerste toets klopt dit; gemiddeld hebben de theoretici meer dan 75% van de vragen goed beantwoord, en de beroepsgerichte leerlingen nog geen 50%.



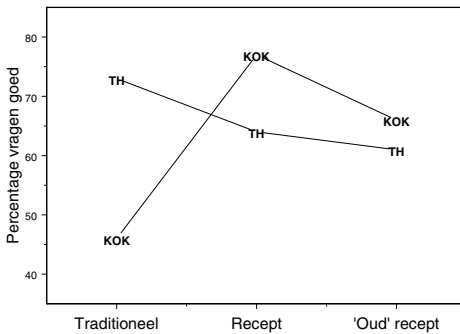
Figuur 1. Gemiddelde scores van leerlingen in de theoretische (T) en basisberoepsgerichte (B) leerweg van het VMBO op drie tekstbegriptoetsen.

Maar bij de tweede toets zijn de resultaten omgedraaid; leerlingen in de theoretische leerweg hebben (gemiddeld) beduidend minder vragen goed beantwoord dan hun beroepsgerichte collega's. Dus: volgens de eerste toets kunnen de theoretici het beste lezen, en volgens de tweede toets kunnen de beroepsgerichte leerlingen het beste lezen. Maar er is nog een derde toets gemaakt, die wellicht uitkomst kan bieden. Op deze tekstbegriptoets is het verschil tussen de beide groepen leerlingen veel kleiner en niet significant. Drie toetsen, drie keer een verschillend resultaat.

De eerste conclusie is: Wat gemeten is varieert! Of, wellicht worden drie verschillende constructen gemeten? Maar wat is er nu eigenlijk gemeten? Het is veel te gemakkelijk om op grond van dit voorbeeld meteen te concluderen dat deze meting van leesvaardigheid niet deugt. En dat we dus het construct leesvaardigheid zouden moeten vervangen door iets als 'leescompetentie'. Dames en heren hier begrijp ik niets van. We hebben een construct, 'leesvaardigheid', en om de één of andere reden zijn we er niet in geslaagd dat zo te meten dat de empirie overeenkomt met onze verwachtingen. Dan deugt het begrip dus niet!? Ik denk dat het veel vruchtbaarder is wanneer we ons afvragen waarom de metingen zo variëren. Zijn er wellicht externe factoren aan te wijzen die verklaren waarom de meting van leesvaardigheid zo varieert. En, daarvoor ligt nog de vraag wat

bedoelen we eigenlijk met leesvaardigheid? Welke cognitieve operaties moet een lezer verrichten om een tekst te begrijpen. Pas als we dat weten, en we weten dat de gebruikte tekstbegripstoetsen echt valide zijn dan heeft het wellicht zin om ...

Ik zal een tipje van de sluier oplichten. In het zoeven besproken voorbeeld waren de beroepsgerichte leerlingen koks in opleiding. En de tekstbegripstoetsen waren achtereenvolgens een traditionele toets, recepten, en uit het Oudnederlands 'vertaalde' recepten. Dan ziet de figuur er als volgt uit.



Figuur 2. Gemiddelde scores van leerlingen in de theoretische (TH) en basisberoeps-gerichte leerweg (KOK) van het VMBO op drie tekstbegripstoetsen.

Dus: theoretici hebben gemiddeld een hogere tekstbegripscore behaald bij een traditionele tekstbegripstoets met vierkeuze-vragen, en koks in opleiding bij het beantwoorden van vragen over recepten. Terwijl bij recepten in het Oudnederlands het verschil tussen beide groepen niet noemenswaardig is. Kennelijk hebben koks veel voordeel van hun grotere kennis van - de structuur - van recepten. We kunnen dus niet anders concluderen dan dat er een duidelijke relatie is tussen tekstbegrip, of leesvaardigheid-zoals-gemeten, en meer algemene (wereld)kennis.

Terug naar de vraag: wat meten we eigenlijk? Wat is tekstbegrip nu? Ik prijs mijzelf gelukkig dat ik op een afdeling werk waar de laatste jaren nogal wat werk gemaakt is van het meten van leesvaardig-

heid. En, dan doelen we met leesvaardigheid op het kunnen leggen van verbanden; verbanden tussen tekstinterne elementen, maar ook relaties met in het geheugen aanwezige kennisstructuren. Lezen is dus een actief proces waarbij de lezer actief relaties tussen tekstuele informatie legt én actief verbindingen maakt tussen de tekstuele informatie en de bij hem of haar aanwezige kennis. Beide vormen van relaties zijn essentieel voor een goed begrip van een tekst.

Zonder nu op alle 'ins' en 'outs' van deze omschrijving in te gaan, is het zonder meer duidelijk dat de kloof tussen leesvaardigheid en meer algemene kennis gedicht wordt. Zonder kennis over een onderwerp is een tekst gewoon veel moeilijker. Iets dat blijkt uit het zoeven gepresenteerde voorbeeldje, maar dat ook in talloze laboratoriumexperimenten aangetoond is (denk maar aan allerlei 'priming' experimenten). Leesvaardigheid is dus in de aard der zaak een begrip waarbij meer komt kijken; een leerling die volgens de één of andere toets goed kan lezen weet wellicht meer over het onderwerp dat in die tekst aangesneden wordt dan zijn medeleerlingen en heeft daarom die toets zo goed gemaakt. Omdat deze leerling zo veel weet over dit onderwerp is het voor hem of haar relatief eenvoudig om allerlei relaties te leggen tussen tekstuele informatie-eenheden, relaties die hij of zij wellicht al paraat had voordat de tekst gelezen werd. Bovendien is het relatief eenvoudig om de tekstuele informatie te koppelen aan al aanwezige kennis. Immers, de leerling leest zaken die hij deels al weet. Dus, uit de aard der zaak is leesvaardigheid in cognitief opzicht een multidimensioneel begrip. En, uiteraard hoeft dan de samenhang tussen de scores op verschillende tekstbegriptoetsen niet perfect zijn. En dus zegt een score op een algemene tekstbegriptoets niet alles over het tekstbegrip op een specifiek genre zoals het begrip van een recept. Maar, dat er wel een duidelijke consistentie zit in de scores op de traditionele toets, de receptentoets en de oude-receptentoets blijkt wel uit de substantiële correlatie per groep ($r > .55$, en geattenuëerd voor de onbetrouwbaarheid van de toetsen $r > 0.78$); een leerling die in een groep hoog scoort op de ene toets scoort ook relatief hoog op de andere toetsen.

Daarnaast biedt deze omschrijving van leesvaardigheid diverse handvaten om het begrip van een tekst te toetsen. Dat zijn dan toetsingen waarbij juist de relaties tussen informatie-eenheden centraal

staan. Het gaat dan om vragen waar de structuur van een tekst centraal staat. Hierbij kan gedacht worden aan vragen waarbij de leerling de elementen in een door de tekst bepaalde volgorde moet zetten. Dergelijke simpele vragen lijken op het eerste gezicht meer valide dan de gewone tekstbegripvragen; zo zijn de verschillen tussen leerlingen veel consistentier over teksten dan bij traditionele toetsen (Kamalski e.a., 2005). Maar, dit type vragen is niet toepasbaar in dezelfde situaties als de traditionele tekstbegripvragen, omdat ze beantwoord moeten worden in afwezigheid van de tekst.

Nu wordt, paradoxaal genoeg, de traditionele meerkeuzevraag vaak gebruikt als een criterium om de validiteit van volgordevragen tegen af te zetten. En, dat terwijl velen juist twijfelen aan de waarde van meerkeuzevragen voor het meten van leesvaardigheid. Met meerkeuzevragen zou het veel meer gaan om testwijsheid, gokvaardigheid, het wegstrepen van foute alternatieven, in plaats van om écht tekstbegrip. Onderzoek naar de relatie tussen open vragen en meerkeuzevragen is net zo oud als de meerkeuzevraag zelf. De uitkomsten zijn helaas wisselend. Deze verschillen in resultaten zijn deels toe te schrijven aan domweg slecht onderzoek; onderzoek waarbij eerst gedaan werd, en waar het denken er toen niet meer van kwam. Sommigen, zoals ene Van den Bergh, concentreerden zich bijvoorbeeld op de vraag of de kans om een item correct te beantwoorden hetzelfde is bij open als bij meerkeuzevragen (1988). Een dergelijke vraagstelling getuigt echter nauwelijks van enige conceptualisatie van de vraag naar de eventuele vergelijkbaarheid van open vragen en meerkeuzevragen. Het gaat er immers *niet* om of een leerling bij een tekstbegriptoets met open vragen exact dezelfde score behaalt als bij een toets met dezelfde vragen meerkeuze-vorm in termen van het aantal correct beantwoorde vragen. Het gaat om de vraag of de rangorde van leerlingen varieert ten gevolge van de vraagvorm. Natuurlijk is de score van een leerling op een toets met meerkeuzevragen hoger dan die van dezelfde leerling op een toets met open vragen, maar dat geldt voor alle leerlingen. Van belang is echter de vraag of goede lezers bij beide type vragen goede lezers zijn, en of slechte lezers bij beide vraagtypen slechte lezers zijn. Het gaat dus om de vraag of de rangorde van leerlingen beïnvloedt wordt door de vormgeving van de vraag. Als die rangorde verschilt dan doen open vragen en meerkeuzevragen een

beroep op verschillende cognitieve vaardigheden, maar als de rangorde niet (noemenswaardig) verschilt, dan meten we met beide vraagvormen dezelfde mentale vaardigheden.

Wanneer we dit criterium stellen, dan blijkt met name uit het onderzoek sinds de jaren tachtig dat voor leesvaardigheid deze samenhang niet perfect, maar wel erg hoog is. Dus, de rangorde van leerlingen op basis van scores op een tekstbegripstoets met open vragen en de rangorde van leerlingen op dezelfde toets met meerkeuzevragen is niet identiek, maar beide rangordes lijken wel heel erg op elkaar (zie: o.a. Van den Bergh, 1990). Dus, strikt genomen worden met open vragen en meerkeuzevragen tekstbegrip niet exact hetzelfde gemeten.

Maar, en daar schuilt het addertje onder het gras, het is een beetje vreemd om leerlingen eerst een tekst voor te leggen met open vragen, en om ze daarna precies dezelfde tekst voor te leggen met precies dezelfde vragen maar nu in meerkeuze-vorm (of vice versa). Je merkt dat meteen aan de reacties van de leerlingen bij de tweede afname, 'meneer, deze vragen hebben we al gehad!'. Bij een tweede meting met dezelfde tekst worden dan ook deels herinneringseffecten gemeten. In verschillende studies is dan ook een meer geavanceerde onderzoeksofzet gebruikt waar de leerlingen op verschillende momenten verschillende tekstbegripstoetsen maakten. Op grond van dergelijke studies komen we voor leesvaardigheid wederom tot de conclusie dat de correlatie tussen open vragen en meerkeuzevragen hoog maar niet perfect is. Echter, de verschillen ten gevolge van de tekst waarbij de vragen gesteld worden zijn vele malen groter dan de verschillen ten gevolge van vraagvorm. Het maakt voor een leerling dus veel meer uit bij welke tekst vragen gesteld worden dan of dit open vragen dan wel meerkeuzevragen zijn. De vraag naar de validiteit van open vragen en meerkeuzevragen wordt daarmee wel in een heel ander daglicht gesteld.

Als argument van heel andere aard wordt wel de dodelijke saaiheid van meerkeuzevragen genoemd. Dat dit argument niet bijster sterk is, wil ik aan de hand van enkele voorbeelden demonstreren. Dit zijn alle voorbeelden uit het Spaans als vreemde taal, die ik van de website van Peter Slagter heb geplukt. Dat het hier om vragen in het Spaans gaat, is -denk ik- geen enkel probleem.

Dus jij dacht dat je nog geen Spaans kent? Let maar eens op!

Uni, dos, tres, cuatro, cinco, seis, siete, octo, nueve, diez

1. Kies het lijstje met de woorden die allemaal bij Spanje horen

- Mañana, Costa Brava, Van Gogh, Dali
- SEAT, Iberia, Heineken, cerveza, Picasso
- Juan Carlos, fútbol, vino, tren, Sevilla
- California, Don Quijote, Majorca, tulipanes, Madrid

Ha contestado a 0 de 10 preguntas y hasta ahora has dado 0 respuestas correctas

Dus jij dacht dat je nog geen Spaans kent? Let maar eens op!

Uni, dos, tres, cuatro, cinco, seis, siete, octo, nueve, diez

2. Wat hoort NIET bij elkaar?

- Johan Cruyff y el fútbol
- Un gato y un perro
- Papá y mamá (padre y madre)
- La escuela y las vacaciones

Ha contestado a 1 de 10 preguntas y hasta ahora has dado 1 respuestas correctas

Let wel, dit zijn geen tekstbegripvragen, maar de toets heeft als doel om leerlingen te laten zien dat je met logisch redeneren, en toepassing van kennis een heel eind komt, ook als de vragen deels in een voor de lezer onbekende taal gesteld zijn. Dat werkt voor leerlingen die deze toets gemaakt hebben, voor zover ik ze gesproken heb althans, enorm motiverend. Zeker omdat direct aangegeven wordt dat een antwoord goed of fout is.

Maar, dit alles zijn alleen nog maar de ingrediënten van een theorie over tekstbegrip en dan nog alleen de evaluatie daarvan. We zijn natuurlijk geïnteresseerd in de mentale activiteiten die leerlingen tijdens het lezen uitvoeren. Want, die willen we in het onderwijs verbeteren, en vervolgens ook evalueren.

Ook daar wordt vanuit diverse gezichtspunten gewerkt aan een theorie. Van diverse kanten wil hier zeggen van manipulatie van kleine tekstelementen tot de verwerking van hele teksten. We zijn dan geïnteresseerd in de relatie tussen kenmerken van de tekst en de vervolgens noodzakelijke processen om deze tekst te verwerken.

Zo worden verschillen ten gevolge van tekstmanipulaties als de aan- of afwezigheid van connectieven onderzocht op het leesproces, maar ook op het begrip. Een mooi voorbeeld hiervan is onderzoek van Jentine Land. Zij vergelijkt het begrip van zogenaamde gefragmenteerde teksten, met dat van meer coherente teksten. Gefragmenteerde teksten zijn teksten in wat ik maar zal noemen: Jip en Janneke taal; korte zinnen, en elk verbindingswoord tussen zinnen is waar ook maar half mogelijk weggelaten. Dergelijke teksten zijn helaas usance in schoolboeken voor het VMBO. Het gaat dan om tekstjes als weergegeven in Figuur 3.

Figuur 3. Een voorbeeld van een gefragmenteerde tekst (overgenomen uit Land 2006).

Op ontdekkingsreis

In de 16^e eeuw maakten Hollandse handelaren verre reizen.
De handelaren wilden in Azië kruiden kopen.
Ze verkochten de kruiden weer in Nederland.
De handelaren maakten veel winst.
Deze kruiden waren in Nederland niet te koop.
De mensen in Nederland hadden de kruiden nodig.
De mensen gebruikten de kruiden om maaltijden meer smaak te geven.
Er werd voedsel bewaard in kruiden
.....

Gelukkig blijkt uit empirisch onderzoek dat dergelijke gefragmenteerde tekstjes minder goed begrepen worden dan dezelfde tekstjes waarbij de relaties tussen de zinnen wél expliciet gemarkeerd zijn (Land, 2006). Explicitering van de structuur in een tekst bevordert het begrip van die tekst.

Maar ook onderzoek naar leesprocessen wordt uitgevoerd, met bijzonder aansprekende resultaten. Zo liet Tanja Jansen in een studie naar literair lezen leerlingen hardop denken. Het aardige van deze studie is nu dat de leesprocessen van zwakke en sterke lezers onderzocht werden. Sterke lezers bleken hun proces tijdens het lezen aan te passen, terwijl zwakkere lezers hun leesproces veel minder aanpassen. Sterke lezers maken bij voorbeeld aan het einde van een tekst meer

inferenties dan aan het begin van die tekst. Een tweede vorm van aanpassing van het leesproces blijkt uit de mate waarin lezers zich aanpassen aan de te lezen tekst. Het leesproces van goede lezers wordt duidelijk beïnvloedt door de tekst die zij lezen, slechte lezers lezen alle teksten op dezelfde manier. Dus: goede lezers lijken hun leesproces aan te passen aan de hoeveelheid informatie die zij op een bepaald moment hebben, en zwakke lezers doen dat beduidend minder. Bovendien verschilt de aanpak van goede lezers van tekst tot tekst, terwijl die van zwakke lezers veel stabiel is, en daarmee minder afgestemd is op de tekst in kwestie.

Kortom, van alle kanten convergeert het bewijs: er zijn grote verschillen tussen teksten, en die komen tot uiting in proces en begrip. Opmerkelijk genoeg heeft dit tot op heden nauwelijks invloed gehad op de examens Nederlands in het voortgezet onderwijs. Daar kunnen we, gezien de examens in de afgelopen decennia, wel het tekstbegrip van onze leerlingen evalueren aan de hand van één enkele tekst. Dat de betrouwbaarheid een probleem is, nemen we haast voor lief. En, dan nog te bedenken dat de gerapporteerde betrouwbaarheden in conceptuele zin een overschatting van de werkelijke betrouwbaarheid van de meting zijn. Op de een of andere wijze zijn we niet in staat dit bij Nederlands op een betrouwbare, laat staan valide wijze te evalueren. Dit is wellicht toe te schrijven aan de grote variëteit in type vragen, de grote verschillen in achtergrondkennis bij de leerlingen en het geringe aantal teksten dat bij de examens Nederlands een rol speelt. Het klemt daardoor des te meer dat we bij Nederlands menen met slechts een enkele tekst te kunnen volstaan. Eigenlijk volstaan we bij HAVO/VWO maar met één tekst met vragen. Dat is wat weinig, zeker in vergelijking met de moderne vreemde talen, waar gemakkelijk vijftien verschillende teksten bevraagd kunnen worden. Bovendien kan het aantal teksten gemakkelijk opgeschroefd worden. Naast een traditionele tekstbegriptoets maken de HAVO/VWO-leerlingen voor Nederlands ook nog een zogenaamde geleide samenvatting. Wat weinigen vandaag de dag meer lijken te weten is dat de samenvatting destijds onderdeel uitmaakte van de meting van schrijfvaardigheid, maar nu plots tekstbegrip meet. Nu weet ik van geen onderzoek dat aantoonde dat met samenvatten feitelijk tekstbegrip

gemeten wordt, en dat een geleide samenvatting in het bijzonder garant staat voor een adequate meting van het tekstbegrip. Wat we wel weten is dat de beide metingen van tekstbegrip in het CE nauwelijks samenhangen. Dat betekent op zijn allerbest dat we twee totaal verschillende onderdelen van tekstbegrip toetsen in het CE. Maar het lijkt mij waarschijnlijker, dat we met de samenvatting nauwelijks tekstbegrip toetsen. Tenminste, niet met de manier waarop we scores aan de samenvattingen toekennen. Met de beste wil van de wereld kan spelling geen onderdeel van leesvaardigheid genoemd worden. Toch speelt spelling een rol bij het nakijken van de samenvattingen. Hetzelfde geldt de redelijk stringente eis met betrekking tot het aantal woorden, waarbij een leerling alleen het maximale aantal punten krijgt als zijn antwoord bondig genoeg geformuleerd is; ik geloof dat niemand de stelling zal onderschrijven dat goede lezers, want daar gaat het om, kortere antwoorden geven dan slechte lezers.

Dat de betrouwbaarheid van de examens Nederlands laag is, is algemeen bekend. Dat maakt de volgende tabel des te aardiger. In die tabel is becijferd hoeveel meer of minder leerlingen zouden zakken voor de opleiding waarvoor zij examen doen als we een halve punt meer of minder aan de examenprestatie zouden geven.

Tabel 1. Percentage leerlingen dat meer of minder slaagt bij verschuiving met een half punt en aantal leerlingen dat examen doet in het vak (gegevens centraal examen 2000).

Vak	VWO		HAVO	
	+0.5	-0.5	+0.5	-0.5
Nederlands	+0.8	-2.3	+0.9	-3.0

¹ Immers, bij tekstbegrip speelt kennis een belangrijke rol. De scores variëren derhalve ten gevolge van verschillen in achtergrondkennis. De betrouwbaarheid van de vragen van één tekst is derhalve een overschatting t.o.v. de betrouwbaarheid van de meting van leesvaardigheid in het algemeen.

Dus, als we voor Nederlands alle leerlingen een half puntje meer geven, dan stijgt het percentage leerlingen dat voor de VWO-opleiding geslaagd is met 0.8 procent. Voor het HAVO zouden 0.9% meer leerlingen geslaagd zijn als alléén het examencijfer voor Nederlands een half puntje hoger was geweest. Als het examencijfer een half puntje lager was dan zou tussen 2 tot 3% meer leerlingen gezakt zijn. Dus: uiterst kleine verschillen in cijfers voor het centrale examen hebben grote gevolgen. En, we moeten ons wel bedenken dat dergelijke kleine verschillen gezien de lage betrouwbaarheid van het examen feitelijk irrelevant zijn.

Nu is het gemakkelijk de examens Nederlands te kritiseren vanwege de beperkte betrouwbaarheid, of het beperkte aantal teksten dat in één examen aan de kandidaten voorgelegd wordt, ook in onderzoek zijn we vaak geen haar beter. Ook in onderzoek willen we vaak op grond van één enkele tekstbegriptoets met één enkele tekst de hele leeswereld verklaren. Of de ene leesdidactiek meer vruchten afwerpt dan de andere evalueren we maar al te vaak aan de hand van één tekstbegriptoets. En dat terwijl we allemaal weten dat de ene tekst de andere niet is (zie: o.a. Meuffels & Van den Bergh, 2005). In dit verband moet de onderwijsadviesraad, die vraagt om 'evidence based learning' zich welhaast een roepende in de woestijn voelen. Want een 'bewijs' op grond van één tekst kan nauwelijks een bewijs genoemd worden. Wat dit betreft moet er in onderwijsonderzoekland en de financiering van onderwijsonderzoek nog wel heel wat veranderen voor dat we echt van 'evidence based learning' kunnen spreken, of zelfs maar een stap in die richting kunnen maken.

We moeten daarbij wel beseffen dat we in Nederland wonen. Nederlanders zijn in één ding bijzonder sterk: eerst veranderen en daarna wellicht nog eens nagaan of de verandering wel zinvol is geweest. Het lijkt *mij* vanzelfsprekend dat we bijvoorbeeld eerst een aantal studies doen naar hoe we het begrip leesvaardigheid beter zouden kunnen meten, voordat we de examens veranderen. Zo zou ik eerst een aantal studies doen naar de relatie tussen tekstbegrip en samenvatten, voordat ik samenvatten 'lezen' zou willen noemen. Evenzo ook zou ik een aantal empirische studies willen uitvoeren voordat ik samenvatten uit het CE zou willen verwijderen. Kortom, we moeten de empirie veel meer laten spreken.

Oké, het kan beter, maar zet twee neerlandici, en twee meetspecialisten bij elkaar in een kamer, en je hebt zes meningen. Laat eerst de empirie spreken, in termen van echt onderzoek én in termen van een literatuurstudie en handel daarna naar bevind van zaken. Empirisch onderzoek is de basis van rationeel handelen. Dat hierbij meer aandacht nodig is voor verschillen tussen teksten dan nu vaak het geval is, spreekt voor zich. Maar al te vaak blijkt dat op grond van één meting van de leesvaardigheid niet de conclusies getrokken kunnen worden waar men voorshands op uit was.

Schrijven maakt geen onderdeel meer uit van het CE Nederlands. Daarvoor kunnen verschillende redenen naar voren gebracht worden. De eerste is de zogenaamde opdrachtspecificiteit; de ene opdracht is de andere niet. Dat wil zeggen: de ene leerling kan goed uit de voeten met de ene opdracht, maar niet met een andere opdracht en voor een andere leerling ligt dat juist andersom. Dat betekent dat de verschillen in beoordeling van teksten niet alleen verschillen in schrijfvaardigheid weerspiegelen, maar ook verschillen in bij voorbeeld achtergrondkennis over het onderwerp in kwestie. Om een betrouwbare indicatie van de schrijfvaardigheid van een leerling te verkrijgen moet deze dan ook meer schrijfoopdrachten maken. De schatting van het aantal benodigde opdrachten loopt uiteen van vijf tot wel veertig (Van den Bergh, Glopper & Schoonen, 1987). Dergelijke aantallen opdrachten om tot een betrouwbare meting van schrijfvaardigheid te komen, zijn voor diverse onderzoekers reden geweest om te verzuchten dat we schrijfvaardigheid voortaan maar met meerkeuzevragen zouden moeten meten; dan worden opdracht-, of onderwerpeffecten tenminste uitgesloten (zie: Van Schooten & De Glopper (1990) voor een overzicht).

Een tweede reden is gelegen in de beoordelingsproblematiek; verschillende beoordelaars kennen aan dezelfde tekst een ander cijfer toe. Het is zelfs zo sterk dat een grote groep docenten géén een tekst unaniem als voldoende of unaniem als onvoldoende beoordeelt (Wesdorp, 1974). Hoewel dat gold in onderzoek uit 1974 is er geen enkele reden om aan te nemen dat de zaak nu wezenlijk verbeterd is. Een instrument om de vrijheid van de beoordelaars kleiner te maken, en dus de overeenstemming in oordelen te vergroten is het Europese

referentie kader (ERK). Van dit ERK worden in dit opzicht wonderen verwacht. In het ERK worden (groveweg) zes verschillende niveaus onderscheiden, variërend van beginnenniveau (A1), tot het niveau van moedertaalsprekers (C2). Eén van de voordelen van het ERK is dat het een voor alle gebruikers duidelijke omschrijving bevat van de onderscheiden taalvaardigheidniveaus. Dit vergemakkelijkt de communicatie binnen de EU enorm, een sollicitant kan met referentie aan het ERK aangeven wat zijn taalvaardigheidniveau is. Echter, de goede bedoelingen en de hoge verwachtingen ten spijt blijkt de werkelijkheid weerbarstig. Weliswaar neemt de overeenstemming tussen getrainde en ongetrainde beoordelaars bij gebruik van het ERK toe; beoordelaars zijn het meer met elkaar eens wanneer zij gebruikmakend van het referentie kader teksten beoordelen, dan wanneer zij géén omschrijving van taalvaardigheidniveaus hebben. Maar, in het ERK wordt er voor de lager niveaus nauwelijks een onderscheid gemaakt tussen verschillende prestaties (Slagter, 2006). Dus, het ERK differentieert nauwelijks voor de lagere taalvaardigheidniveaus, veel minder althans dan alternatieven dat doen. Daarnaast blijkt het hoogste niveau bijkans onbereikbaar; weinig moedertaalsprekers zullen in hun moedertaal op het hoogste taalvaardigheidniveau ingeschaald worden – ik althans niet. Kortom, het Europees referentie kader is nog verre van ideaal. Vooral het gebrek aan onderscheidend vermogen voor de lagere taalvaardigheidsniveaus – en dan moet u denken aan bovenbouw VWO-leerlingen – is een groot probleem. Eén van de doelstellingen van het ERK om door adequate beschrijving het taalonderwijs en het bereikte eindniveau te verbeteren is dan ook vooralsnog onbereikbaar.

De tweede doelstelling is, zoals gezegd dat we allemaal hetzelfde verstaan onder een gespecificeerd niveau. We hebben dan als het ware een meetlat die voor iedereen geldt. Het belang van een dergelijke meetlat kan moeilijk overschat worden! De vraag is echter of deze pretentie met het ERK waar gemaakt kan worden. Hanteren Duitse docenten Duits dezelfde normen als Nederlandse docenten Duits wanneer zij de teksten van leerlingen beoordelen met het ERK? Of hanteren Engelse docenten Engels dezelfde normen als hun collega's in Nederland? Dit lijkt vooralsnog niet het geval. Moedertaal docenten lijken veel coulanter wanneer het gaat om de

beoordeling van teksten; grammaticale fouten, spelfouten e.d. worden schrijvers veel minder zwaar aangerekend dan begripsproblemen; oordelen over dezelfde teksten van moedertaal sprekers vallen dan ook hoger uit. Ongeacht of zij het ERK gebruiken of niet (Ubels 2006). Het ERK schiet dus overduidelijk zijn doel voorbij; aan de onderkant van de schaal wordt onvoldoende gedifferentieerd, de hoogste niveaus zijn ook voor moedertaalsprekers bijkans onhaalbaar, en native en niet-native beoordelaars hanteren ongelijke normen.

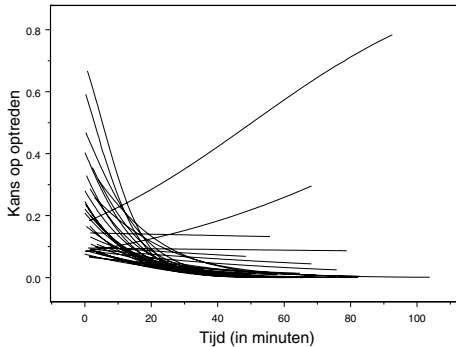
Dit alles zegt echter niets over het proces van schrijven zelf. En, het is toch wel handig als we inzicht hebben in de mentale activiteiten die een schrijver uitvoert tijdens het schrijven, omdat dat bepaalt wat er onderwezen moet worden en natuurlijk wat er gemeten wordt en wat we zouden moeten willen meten. Daarmee zijn we aangeland bij het schrijfprocesonderzoek. Schrijfprocesonderzoek is noodzakelijk voor zowel een valide meting van schrijfvaardigheid, als voor adequaat onderwijs.

Op grond van het oudere schrijfprocesonderzoek kan nauwelijks een onderscheid gemaakt worden tussen goede en minder goede schrijvers omdat de kwaliteit van de afgeleverde schrijfproducten vaak niet bepaald werd; het verschil tussen goede en minder goede schrijvers werd a priori bepaald op grond van de ervaring die de schrijver had; minder goede schrijvers hebben weinig ervaring en goede schrijvers hebben veel ervaring. Gezien de variabiliteit tussen opdrachten lijkt deze aanname uiterst kwetsief. Het is dan ook niet verwonderlijk dat in onderzoek de frequentie waarmee schrijvers processen als plannen, genereren, structureren, of reviseren uitvoeren niet of nauwelijks gerelateerd bleek aan deze operationalisatie van schrijfvaardigheid.

Tegenwoordig wordt in schrijfprocesonderzoek vaak onderscheid gemaakt naar het moment waarop een bepaalde activiteit uitgevoerd wordt (Rijlaarsdam & Van den Bergh, 2006). Het blijkt dat, om maar een open deur in te trappen, een activiteit als 'het lezen van de opdracht' vooral in het begin van het schrijfproces een bijzonder populaire activiteit is. Dat wil zeggen: 'het lezen van de opdracht' is een activiteit die in het begin van het schrijfproces veelvuldig ondernomen wordt. Dit is in de onderstaande figuur weergegeven. In deze figuur geeft elke lijn één schrijver weer. Op de x-as is het verloop in

de tijd weergegeven, en op de y-as is de kans dat een leerling bezig is met het lezen van de opdracht.

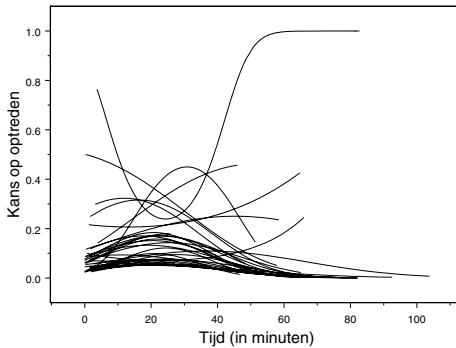
Figuur 4. Distributie van 'het lezen van de opdracht' over het schrijfproces.



Er zijn leerlingen die de opdracht als uitgangspunt nemen; aan het begin besteden zij veel tijd aan het lezen van de opdracht. Er zijn ook leerlingen die de opdracht als het ware links laten liggen en meteen beginnen met schrijven. Duidelijk is dat er grote verschillen zijn tussen leerlingen in de mate waarin zij de opdracht raadplegen. Over het algemeen wordt de opdracht aan het begin van het schrijfproces relatief vaak geraadpleegd, en minder naarmate het schrijfproces vordert, maar de mate van afname varieert van schrijver tot schrijver. Op dit algemene beeld zijn (natuurlijk) ook uitzonderingen; bij sommige leerlingen neemt de kans op het raadplegen van de opdracht gedurende het schrijfproces toe in plaats van af.

Voor andere processen ziet een dergelijk figuur, waarin de kans dat een proces optreedt is uitgezet tegen de tijd dat een leerling schrijft, er totaal anders uit. Neem nu genereren van informatie, het hardop uitspreken van stukjes geproduceerde tekst.

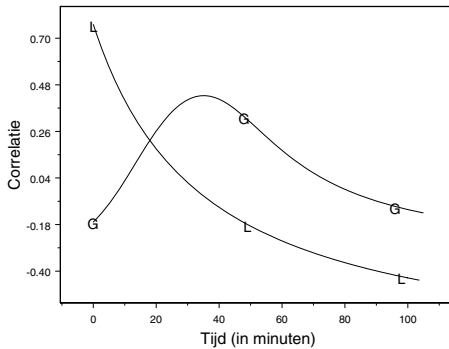
Figuur 5. Distributie van 'generereer-activiteiten' over het schrijfproces.



Wederom grote verschillen tussen leerlingen; de één begint direct met genereren, terwijl een ander eerst andere dingen doet, zoals het lezen van de opdracht. Tot nog toe zijn deze gegevens alleen beschrijvend van aard, en interpretatie anders dan er zijn grote verschillen is lastig.

Wat we nodig hebben is een interpretatiekader op grond waarvan we een conclusie over de kwaliteit van het schrijfproces zouden kunnen trekken. Een dergelijk interpretatiekader kunnen we (deels) ontleunen aan relaties tussen schrijfproces en schrijfproduct. Welke distributie van een activiteit is geassocieerd met (relatief) goede teksten, en welke distributie is geassocieerd met teksten van mindere kwaliteit. In de volgende figuur is de correlatie tussen schrijfproces en schrijfproduct weergegeven. Net zoals bij het proces het geval is, blijkt ook de correlatie tussen proces en product te variëren gedurende het schrijfproces. Zo hebben de leerlingen die in het begin van het schrijfproces veel aandacht aan de opdracht besteed hebben over het algemeen een beter product afgeleverd dan degenen die de opdracht op latere momenten in hun proces geraadpleegd hebben; in het begin is er een duidelijk positief verband tussen schrijfproces en schrijfproduct, maar al snel slaat dat om in een negatieve relatie, schrijvers die later in hun schrijfproces nog de opdracht raadplegen hebben een (relatief) slechte tekst geschreven.

Figuur 6. Correlatie tussen 'het lezen van de opdracht' (L) en tekstkwaliteit en 'genereren' (G) en tekstkwaliteit.



Voor genereren daarentegen is de samenhang tussen proces en product duidelijk anders. In het begin van het schrijfproces is de correlatie negatief; leerlingen die direct beginnen met genereren van informatie hebben een (verhoudingsgewijs) slechte tekst geschreven. Klaarblijkelijk is het beginnen met genereren van informatie niet zo'n erg handige strategie. Eerst oriënteren op de opdracht en dan pas informatie genereren lijkt veel beter, want na zo'n dertig minuten is de correlatie tussen genereren en tekstkwaliteit wel substantieel.

Wat we weten is dat bij deze beide processen de distributie redelijk constant is over opdrachten en over leeftijd; ook bij andere opdrachten raadplegen goede schrijvers vaak de tekst van de opdracht aan het begin, en stellen goede schrijvers genereren over het algemeen iets uit. Hetzelfde beeld zien we bij jongere schrijvers, ook hier raadplegen de betere schrijvers in het begin van het schrijfproces de opdracht (relatief) frequent, en stellen de betere schrijvers genereren uit tot een later stadium (zie: o.a. Van der Hoeven, 1997). En ook bij jongere schrijvers is het lezen van de opdracht in het begin positief gecorreleerd aan tekstkwaliteit, en blijkt dat het uitstellen van genereren van informatie positief gerelateerd is aan tekstkwaliteit.

Hoewel we het één en ander weten over schrijfprocessen blijven er vele interessante vragen over. Bij voorbeeld: in hoeverre zijn vaardigheden als leesvaardigheid, of revisievaardigheid, woordkennis, maar

ook achtergrondkennis bepalend voor de distributie van schrijfprocessen en aldus gerelateerd aan tekstkwaliteit? Het aardige is dat voor sommige vaardigheden gebleken is dat zij van invloed zijn op de distributie van processen, en alleen via het proces gerelateerd zijn aan kwaliteit van de geproduceerde tekst. Bij voorbeeld: leerlingen die goed kunnen reviseren, reviseren later in het schrijfproces. En, alleen op latere momenten is er een verband tussen revisies en tekstkwaliteit (Van der Hoeven, 1997), of de relatie tussen tekstkwaliteit en woordkennis in een vreemde taal. In verschillende onderzoeken is de correlatie tussen woordkennis en tekstkwaliteit afwezig, maar woordkennis is wel van invloed op de verdeling van cognitieve processen, en die zijn weer gerelateerd aan tekstkwaliteit. Dus: via de distributie van schrijfprocessen lijken dergelijke vaardigheden invloed op de kwaliteit van de geproduceerde tekst uit te oefenen (Tillema, 2005).

Een tweede, interessante vraag is die naar de verschillen tussen taken. Zoals al gememoreerd, is maar al te vaak aangetoond dat oordelen over de kwaliteit van verschillende teksten uiterst laag samen hangen. Elke schrijfpdracht meet voor het grootste deel iets unieks. Maar hoe deze verschillen terug te zien zijn in het schrijfproces is een terra incognita. Veel schrijfprocesonderzoekers beperken zich in hun analyse tot één opdracht, en dan spelen verschillen tussen opdrachten natuurlijk géén rol meer. Maar we weten ook niet in hoeverre de door de onderzoekers gerapporteerde resultaten generaliseerbaar zijn naar andere opdrachten, of in hoeverre het een resultaat betreft dat alleen geldt voor deze opdracht (Rijlaarsdam & Van den Bergh, 2006; Van Weijen e.a. 2006).

Een derde vraag is die naar schrijven in een vreemde taal. Gek genoeg kunnen kinderen vaak wel schrijven, argumenteren en redeneren in het Nederlands, maar zodra er geschreven moet worden in een vreemde taal gaat het mis. Nu zijn de verschillen tussen teksten in het Nederlands en bij voorbeeld teksten van de zelfde leerlingen in het Engels veel groter dan je zou mogen verwachten op grond van de geringere woordenschat in het Engels, of de geringere grammaticale kennis van het Engels. Nee, het is eigenlijk nog erger. Wat dacht u van de volgende voorbeelden: 'A word like stress means like succes. Right, because when you to press to stress, than you do better your best. And that's good'. Het lijkt wel alsof elke vorm van argumentatie, en elke

vorm van redeneren ontbreekt zodra geschreven moet worden in een vreemde taal.

Wetenschap is traag, langzaam en laat zien wat iedereen al weet. Dat klopt, tenminste tot op zekere hoogte. Maar om na te gaan of een schrijfdidactiek werkt moeten we eerst weten hoe leerlingen schrijven; welke mentale activiteiten leerlingen moet ondernemen, opdat een goede tekst geschreven wordt. Vervolgens moet aangetoond worden dat de nieuwe, experimentele didactiek juist de succesvolle cognitieve activiteiten stimuleert, en de mindere effectieve inhijbeert. Maar om het allemaal nog ingewikkelder te maken, er moet gedemonstreerd worden dat door de nieuwe didactiek de leerlingen op het juiste moment de juiste dingen gaan doen. Bij voorbeeld dat zij eerst de opdracht lezen, informatie genereren, evalueren en structureren voordat zij gaan schrijven.

Het gegeven dat lezen en schrijven veel ingewikkelder in elkaar zitten dan op het eerste gezicht vaak gedacht wordt, is wellicht de reden dat we vaak achter elke nieuwe hype aanlopen. Iemand lanceert een idee, en voordat dat ook maar serieus beproefd is, voordat ook maar nagegaan kan zijn of dat idee ook werkt, in termen van cognitieve processen en vaardigheden, lopen we er al mee weg. Er zijn in dit opzicht slechts buiten gewoon weinig leermiddelen voor de talen waarvoor onderzocht is welke verandering in cognitieve activiteiten zij bewerkstelligen, en welke effecten dat heeft op de kwaliteit van de taakuitvoering door de leerlingen. Het advies van de onderwijsraad voor 'evidence based learning' heeft wat dat betreft nog weinig gehoor gevonden; didactisch onderzoek waarin de gehele causale keten van didactiek via schrijfproces naar tekstkwaliteit doorlopen wordt, is uiterst maar dan ook uiterst schaars. Ik ken eigenlijk maar één onderzoek waarin dit systematisch aan de orde gesteld is. Martine Braaksma is in een aantal studies het effect van observerend leren schrijven nagegaan. Haar uitgangspunt was dat een schrijver feitelijk twee zaken tegelijk moet doen: Hij moet niet alleen een goede tekst schrijven, hij moet ook leren van zijn schrijfproces. Als het schrijfproces nu veel cognitieve capaciteit kost, dan komt het leren er vaak niet meer van. Het is dus verstandig het leren schrijven los te koppelen van het schrijven zelf. Dat kan door bij voorbeeld leerlingen te

laten kijken naar een andere leerling die hardop denkend schrijft. Doordat over dit schrijfproces vragen gesteld worden, kan een leerling, zonder dat hij echt aan het schrijven is toch een oordeel geven over de efficiëntie van het proces dat hij zoeven heeft gezien. Het aardige is nu dat Braaksma niet alleen een effect heeft laten zien in termen van tekstkwaliteit (observerend leren werkt!), maar ook in termen van proces. Leerlingen die naar een schrijfmodel gekeken hadden, vertoonde een efficiëntere distributie van activiteiten over het schrijfproces. Let wel, dat gold niet voor alle activiteiten, maar wel voor de activiteiten waarvan Braaksma kon aantonen dat ze gerelateerd waren aan tekstkwaliteit (zie: Braaksma et al 2004).

Samenvattend kom ik tot de conclusie dat zowel voor schrijfprocesonderzoek, onderzoek naar schrijfvaardigheid, en schrijfdidactisch onderzoek, we op zijn best de beschikking hebben over de rudimenten van een werkbare theorie. Veel vragen zijn tot dusver blijven liggen. Eén ding is echter zeker, het al oude adagium 'laat leerlingen schrijven als je wilt weten hoe goed ze kunnen schrijven' (De Groot & Van Naersen 1973, p.7) is niet volledig, want de variabiliteit ten gevolge van kenmerken van de opdracht is veel te groot. Als we de schrijfvaardigheid of het schrijfproces van een leerling aan de hand van één opdracht evalueren, dan weten we één ding zeker: we kunnen slechts uiterst beperkt generaliseren naar een andere opdracht. Kortom, we kunnen in zo'n geval nauwelijks iets zeggen over schrijfvaardigheid in het algemeen. We hebben een inhoudelijke theorie nodig die aangeeft hoe schrijfvaardigheid in elkaar zit, wat we moeten onderwijzen en wat en hoe we dat moeten evalueren.

Tot nog toe heb ik alleen maar gesproken over de leerling; welke mentale processen worden uitgevoerd, en welke zouden uitgevoerd moeten worden voor een betere taakuitvoering. Hoe kunnen we verschillen tussen leerlingen op een adequate manier blootleggen, hoe kunnen we een goede manier van toetsen ontwikkelen, hoe komen we er achter welke cognitieve activiteiten een leerling moet ontplooiën om een taak succesvol uit te voeren. En, wat zegt dit over de te leren vaardigheden. Dit type onderzoek is noodzakelijk omdat daarmee indicaties verkregen worden wat een leerling moet leren.

Aan de andere kant is er natuurlijk de docent. Hoe kan de

docent leerlingen het beste leren om goed te leren lezen en goed te schrijven in het Nederlands en in de vreemde talen. En, natuurlijk ook hoe goed te spreken en te luisteren. Dames en heren, ik weet het niet! Wat ik wél weet is dat dit eigenlijk een valse vraag is. Er is niet één goede manier! Geen enkele wijze waarop een docent zijn stof aan kan bieden, is de meest effectieve voor alle leerlingen! Sommige zaken werken beter voor de ene leerling en andere aanbiedingsvormen werken voor een andere leerling (zie: o.a. De Mayer & Rijmenans, 2004, of meer specifiek voor schrijven: Kieft, 2006). Wel zijn er naast een aantal randvoorwaarden enkele constanten in te ontdekken. De randvoorwaarden zijn op zich simpel. Er moet voldoende lestijd zijn om het vak te onderwijzen. Ik schaam me haast vandaag de dag dit te moeten zeggen, maar dit is niet het geval. Op bij voorbeeld het VMBO staat het vak Nederlands onder druk; de lestijd Nederlands lijkt steeds meer onderdeel uit te maken van projecten. Waar blijft dan het onderwijs in de moedertaal. Géén enkel land dat zo slordig met het onderwijs in de moedertaal om lijkt te gaan als Nederland. Maar ook het onderwijs in de vreemde talen staat onder druk. De uren nemen af, want, zo wordt vaak gezegd, 'wij zijn toch zo goed in talen', het kan dus wel een beetje minder. Dit staat op gespannen voet met de Europese doelstellingen waarin van elke burger verwacht wordt dat hij of zij twee vreemde talen min of meer vloeiend spreekt. Nu zijn er weinig harde gegevens bekend over een vergelijking in de tijd (was het vroeger echt wel beter?) of een inhoudelijke typering van het niveau van het onderwijs. Maar dat het niveau achteruit zal gaan als er minder onderwijs gegeven wordt, lijkt een waarheid als koe.

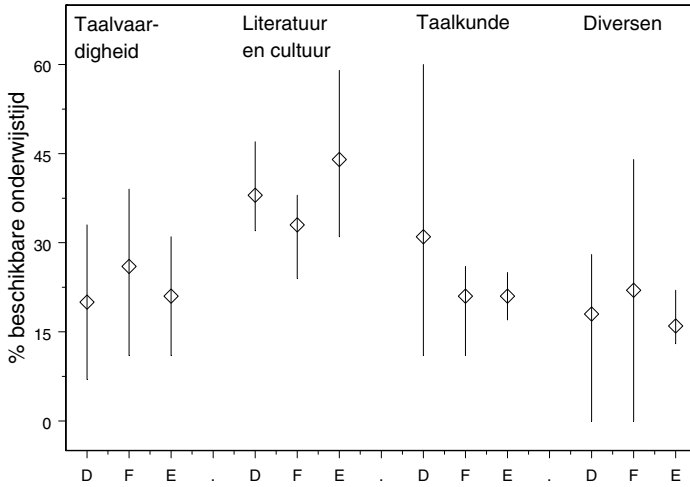
Wanneer we meer inhoudelijk naar ons vreemde-taalonderwijs kijken, dan valt op dat het vreemde-taalonderwijs voornamelijk in het Nederlands gegeven wordt. Slechts in een bijzonder kleine minderheid van de lessen Frans, Duits en Engels spreekt de docent de doeltaal. Opmerkelijk toch dat je Duits, Frans of Engels in het Nederlands leert! Dat is ongeveer hetzelfde als wanneer een wiskunde docent geen cijfers meer gebruikt; het hart wordt uit het vak gesneden. Gek dat in de ons omringende landen de doeltaal wel gebruikt wordt in het vreemde-taalonderwijs en in Nederland nauwelijks. Nu wordt wel gezegd dit een kwestie van traditie is, en dat afgaand op correcte

intonatie e.d., docenten in elk geval de vreemde taal wel beheersen. Het schaarse gebruik van de vreemde taal kan dan in elk geval niet liggen aan een gebrek aan beheersing van die taal, zo is de redenering. Het is veel meer een kwestie van gewoonte, voortvloeiend uit het traditionele grammatica-vertaalonderwijs, in plaats van een bewuste beslissing van de docent.

Ik denk dat deze conclusie wellicht te kort door de bocht is; ik weet niet zo zeker of alle docenten de vreemde taal op een voldoende niveau beheersen. Alle eerstegraads docenten zouden hun taal minimaal op 'near native' niveau moeten beheersen, althans dat is de doelstelling van onze universiteiten. Dit blijkt echter niet zondermeer uit het universitaire curriculum. In de onderstaande figuur geef ik een overzicht van de beschikbare cursustijd die in de bachelor besteed wordt aan taalvaardigheid, literatuur en cultuur, taalkunde en diversen (zoals vertalen). In deze figuur representeert de ruit het gemiddelde, en de lengte van de lijn de verschillen tussen universiteiten.

Voor Duits wordt gemiddeld bijna twintig procent van de onderwijstijd in de bachelor verplicht aan taalvaardigheid besteed. Afhankelijk van de universiteit varieert dit echter van circa tien tot ruim dertig procent. Voor Frans varieert de onderwijstijd van ruim tien tot ruim veertig procent, en voor Engels van tien tot dertig procent, met een gemiddelde van rond de twintig procent van de beschikbare cursustijd in de bachelor. Dat betekent dat een student die educatieve master binnentreedt, de student die op het punt staat te beginnen aan zijn lerarenopleiding, in (minimaal) drie maanden taalvaardigheidsonderwijs op 'near native' niveau gebracht wordt. Ik ben bijzonder nieuwsgierig naar de wijze waarop dit onderwijs vormgegeven wordt. Want om studenten in drie maanden onderwijs, met tussen de vier à zes contacturen per week, op 'near native' niveau te brengen is werkelijk een topprestatie. De Nederlandse universiteiten beseffen nog niet op welk een goudmijn zij zitten; zij moeten deze kennis exploiteren, want zelfs de 'nonnen van Vught' kunnen nog wat van hen leren!

Figuur 7. Gemiddelde tijdbesteding aan vier onderdelen op de universiteiten voor drie schooltalen (? : gemiddelde over opleidingen; | : variatie tussen opleidingen; D: Duits; F: Frans; E: Engels).



Natuurlijk wordt nu direct tegengeworpen dat niet alle taalvaardigheid opgedaan wordt in het expliciete taalvaardigheidsonderwijs; taalvaardigheid is ook tijdens andere vakken in de bachelor een belangrijk aandachtspunt. Natuurlijk, dat klopt! Maar het is toch wel opmerkelijk dat bij voorbeeld lang niet alle literatuur op de universiteiten gelezen en besproken wordt in de doeltaal, maar in het Nederlands. Of dat, gezien het studieprogramma, vertalen gezien wordt als de optimale voorbereiding op het leraarschap. Kortom, enige vernieuwing en aanpassing aan de huidige tijd en de huidige eisen zou het universitaire onderwijsprogramma niet misstaan.

In dit verband zou het buitengewoon aardig zijn de taalvaardigheid van de studenten eens echt te meten, en na te gaan in hoeverre de doelstelling om studenten op 'near native' niveau af te leveren, meer is dan een papieren tijger. Ik ken maar één opleiding in een vreemde taal waar dit ook daadwerkelijk gebeurt, en waar studenten ook daadwerkelijk uit de opleiding verwijderd worden bij een gebreken gebrekkige taalvaardigheid.

Het is een goede Nederlandse gewoonte om eerst op grote schaal veranderingen door te voeren, en pas daarna wellicht een evaluatie uit te voeren. Dat geldt voor de basisvorming, voor de tweede fase, maar ook voor het bachelor-master systeem. Wanneer we ons beperken tot het voortgezet onderwijs (zie: Buelens (2006) voor een karakterisering van ons universitair onderwijs) dan valt op dat de basisvorming door de inspectie van het onderwijs niet al te positief beoordeeld is, en dat al aan de tweede fase werd gesleuteld voordat deze serieus geëvalueerd was. Voorts rept vrijwel iedereen over veranderende leerlingprestaties, waarbij de voorstanders van de verandering vanzelfsprekend 'concluderen' dat de leerling nu meer kan, en de tegenstanders trekken even vanzelfsprekend de tegenovergestelde 'conclusie' dat het allemaal alleen minder wordt. De constante in deze discussies is echter het gebrek aan een empirische basis. Voor verreweg het grootste deel worden meningen en vooroordelen geventileerd die op zijn best een karikatuur van de situatie in het onderwijs zijn. Nu lijkt dit wat vreemd, want ook ik geef hier zo maar een mening.

Het aardige is echter dat er wel enige empirische referentie is. In dit verband verwijs ik naar de evaluatie van de basisvorming. Met deze evaluatie verschaft de inspectie ons een enorme bron van informatie en kennis over wat er daadwerkelijk in ons onderwijs gebeurt. In het kader van deze evaluatie hebben inspecteurs op grote schaal lessen bezocht en oordelen over diverse aspecten van deze lessen geveld. Deze evaluatie heeft om verschillende redenen een schat aan gegevens opgeleverd. Ten eerste is het voor het eerst –in Nederland althans– dat op zo'n grote schaal onderwijs geobserveerd is. Ten tweede is, en dat mag ook wel eens gezegd worden, zelden gebruik gemaakt van zulke kwalitatief goede beoordelaars, en ten derde zijn zelden zoveel lessen van zoveel docenten per school geobserveerd. En, dan was het niet, zoals zo vaak, een gelegenheidssteekproef, maar een echte random steekproef, van zowel scholen, als van docenten binnen scholen; met deze gegevens kunnen we dus wat!

De beoordelingsprocedure was als volgt: de inspectie onderscheidt een aantal standaarden, zoals: pedagogisch handelen, didactisch handelen, actief leren, differentiatie, etc. Bij elke standaard horen een aantal indicatoren. De inspecteur beoordeelde de lessen aan de hand van de indicatoren. Het uiteindelijke oordeel over een standaard is

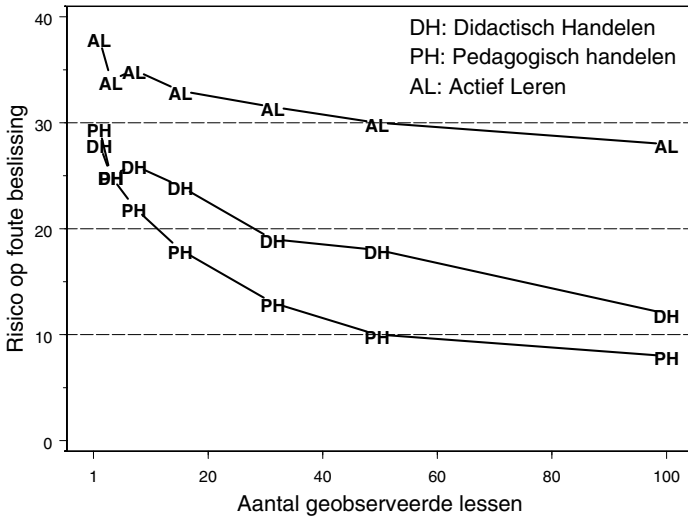
gegeven aan de hand van deze indicatoren. Om tot een voldoende oordeel over een les te komen moeten de zogenaamde essentiële indicatoren voldoende beoordeeld zijn. Het oordeel over het onderwijs op een school tenslotte is bepaald door het aantal als voldoende beoordeelde lessen. In feite heeft er nu een dubbele normering plaatsgevonden. Eerst is nagegaan welke indicatoren in een les als voldoende beschouwd werden, wat leidde tot een oordeel over de kwaliteit van de geobserveerde les. En, daarna is men aan de hand van het aantal lessen voldoende gekomen tot een oordeel over de kwaliteit van het onderwijs op de school in kwestie.

Deze dubbele normering heeft meer consequenties dan men op het eerste gezicht zou zeggen. Informatief in dit opzicht is het risico op een foute beslissing, of: de kans dat een school ten onrechte als slecht in de hoek gezet wordt, dan wel ten onrechte als goed beschouwd wordt. Beide type foute beslissingen zijn natuurlijk niet wenselijk. Als het onderwijs op een school onterecht als inadequaot gekenschetst wordt, dan wordt deze school onterecht in het verdomhoekje geplaatst, wat zeker met de huidige openheid grote consequenties kan hebben. Ook het onterecht niet opmerken van slecht onderwijs is niet chique, om het maar voorzichtig te zeggen. Dan krijgen immers hele groepen leerlingen niet het onderwijs waar zij recht op hebben. Nu is het, gewoon als tijdverdrijf aardig om eens te kijken hoe groot dit risico nu eigenlijk is. Dat kan door uit te gaan van de gegevens die de basisvorming heeft opgeleverd en aan de hand hiervan diverse simulaties uit te voeren. Hierbij spelen natuurlijk diverse factoren een rol, zoals de verschillen tussen docenten en scholen, het aantal geobserveerde lessen etc. Maar, de grap is nu dat al deze gegevens ofwel bekend verondersteld kunnen worden op grond van de door de Inspectie verzamelde gegevens, ofwel gesimuleerd kunnen worden.

In Figuur 8 is het risico op een foute beslissing weergegeven. Het blijkt dat de kans op een foute beslissing varieert afhankelijk van het aantal observaties, i.e. het aantal geobserveerde lessen op een school. Maar voor twee van de drie aspecten is de kans op een foute beslissing zelfs wanneer honderd lessen geobserveerd worden groter dan tien procent. Dus: het onderwijs op meer dan tien procent van de scholen wordt onterecht als adequaat of onterecht als inadequaot gekenschetst. Uit de figuur blijkt dat voor 'Didactisch handelen', er bij

observatie van acht lessen een risico ongeveer vijftwintig procent is. Bij acht observaties is het risico bij 'Pedagogisch handelen', iets meer dan twintig procent en bij Actief leren vijfendertig procent.

Figuur 8. Risico op een onjuist oordeel over drie aspecten van het onderwijsleerproces.



Als we een risico van tien procent acceptabel vinden, dan zijn 'Actief leren' en 'Didactisch handelen' niet te beoordelen. 'Pedagogisch handelen' zou dan alleen bij vijftig of meer waarnemingen adequaat beoordeeld kunnen worden. Maar we kunnen ook een groter risico op een foute beslissing acceptabel achten. Stel, we accepteren een kans van twintig procent op incorrecte beslissing, dan nog blijkt 'Actief leren' niet te beoordelen, en vraagt 'Didactisch handelen' meer dan dertig en 'Pedagogisch handelen' zelfs meer dan vijftig observaties (zie: o.a. Van den Bergh & Zwarts, 2004).

De risico's zijn hoog, wellicht te hoog. Dit ligt aan de dubbele dichotomisering die is toegepast. Beter is om de scores op de individuele indicatoren op te vatten als de scores op een test, en elke les te zien als een onderdeel van deze test. Pas achteraf worden dan de indi-

catoren gesommeerd tot één geheel. En, pas dan wordt een evaluatieve conclusie getrokken over de kwaliteit van het onderwijs op een school; er wordt pas gesproken over goed onderwijs als een bepaald percentage indicatoren voldoende is beoordeeld. Dan zouden bij voorbeeld voor 'Pedagogisch handelen' tussen de acht à twaalf lessen nodig zijn voor een maximaal risico van tien procent. Voor de goede orde, in de oude situatie met de dubbele normering zijn vijftig lessen nodig voor een zelfde risico.

Maar terug naar de inhoud. Iedereen kan kritiek geven op de inhoud van beoordelingformulieren, of op het feit dat de overeenstemming tussen de beoordelende inspecteurs niet 100% was, of, zoals ik net gedaan heb, op de normeringsprocedure. Dit alles neemt niet weg dat deze studie een schat aan gegevens blootgelegd heeft.

Eén van de conclusies die consequent, bij alle vakken, naar voren kwam is dat er enorme verschillen tussen docenten (binnen scholen) zijn, én dat de docent de spil van goed onderwijs is. Ook binnen één school zijn de verschillen tussen docenten die dezelfde taal onderwijzen enorm. In elke school zijn wel docenten te vinden, waarvan vrijwel elke les gebruikt kan worden als promo-filmpje voor het vak docent; lessen waarin elke opdracht zo ingericht is dat deze een doel voor de leerling heeft, in plaats van enkel een doel voor de docent of gewoon een opdracht is die maar gedaan moet worden. Deze docenten weten hun opdrachten zo in te kleden dat het lijkt alsof de leerling zelf aan het roer staat. Maar aan de andere kant zijn er ook op elke school docenten die er niet in slagen leerlingen aan het werk te zetten en/of te houden. In alle nadere analyses bleek de docent de spil waarom het draait; het hoeft wellicht géén betoog, maar goed onderwijs is veel meer afhankelijk van de docent dan van de school.

Aan deze conclusie kunnen verschillende consequenties verbonden worden. De eerste conclusie is: goed onderwijs wordt gemaakt door de docent. Het is daarom van het grootste belang dat geïnvesteerd wordt in de kwaliteit van docenten. Systeemveranderingen, zoals bij voorbeeld de invoering van het VMBO, leiden niet tot beter onderwijs, alleen investeringen in docenten –in termen van opleiding, nascholing en taakomvang– kunnen dat doen.

De tweede conclusie is dat goed onderwijs afhankelijk is van de docent. Die bepaalt *hoe* welke leerstof wanneer wordt aangeboden. Daarbij wordt hij of zij ondersteund door de methode, maar daarvan lijkt in de onderbouw de invloed veel minder groot. De docent draagt er zorg voor dat leerlingen gemotiveerd aan het werk zijn, en aan het werk blijven.

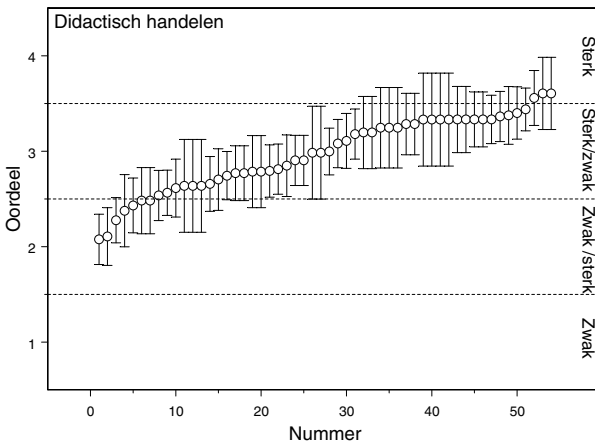
De derde conclusie is dat goed onderwijs alleen gerealiseerd kan worden met goede docenten, maar niet alle docenten geven in alle klassen even goed les. Er is een duidelijk samenspel tussen docent en klas. Wat een docent in de ene klas, als het ware met de vingers in de neus realiseert, lukt nauwelijks in de andere klas.

De vierde conclusie: onderwijsvernieuwing vindt in de klas plaats. Docenten bedenken nieuwe aanbiedingsvormen, voeren nieuwe evaluatievormen uit, en leggen verbanden binnen en tussen vakken. Erkend moet worden dat docenten de spil zijn van onderwijsvernieuwing en onderwijsverandering. Goed onderwijs begint en eindigt bij de docent voor de klas, en hier moet de broodnodige erkenning voor komen.

Het is spijtig dat de Inspectie haar werkwijze vrijwel direct na uitvoering van de evaluatie van de basisvorming verandert heeft. Weliswaar worden globaal nog dezelfde indicatoren gebruikt, dat wil zeggen de labels zijn hetzelfde, maar de inhoud is veranderd. Dit is om verschillende redenen spijtig. Als belangrijkste wil ik documentatie in de tijd noemen. We willen immers vrijwel altijd weten hoe goed het gaat met ons onderwijs. We willen bij voorkeur kunnen aantonen dat het onderwijs van nu beter is dan het onderwijs van tien jaar geleden. Maar, als we willen laten zien dat het onderwijs heden ten dagen beter is, dan is het noodzakelijk dat we de wijze waarop we gegevens verzamelen, noch de gegevens die we verzamelen veranderen. Doen we dat wel dan ontvalt elke basis aan een vergelijking in de tijd. Dan vervallen we, zoals maar al te vaak het geval is, tot onderbuik gevoelens als 'vroeguh, toen we nog de gulden hadden was alles beter!'. Men zou dus wel drie keer na moeten denken, en wel heel zwaarwegende argumenten moeten hebben om grote veranderingen in de inhoud en procedures te wijzigen, want dan ontvalt elke empirische basis aan de discussie en zijn er als het ware terug bij af.

Het beoordelingsformulier van de inspectie kan ook gebruikt worden als instrument in een vorm van zelfevaluatie. Dit is, voor zover ik weet, op een klein aantal scholen gebeurt. Enerzijds hebben docenten hun eigen lessen aan de hand van een schema met indicatoren beoordeeld, anderzijds zijn op verschillende scholen lessen door studenten geobserveerd. Dit heeft voor betrokkenen een aardig doorkijkje in het eigen onderwijs opgeleverd, al moet daarbij gezegd worden dat dit niet door alle docenten altijd even prettig ervaren is. Een concreet voorbeeld. Op twee scholen zijn in drie weken alle lessen geobserveerd door een team van twaalf onderzoekers (studenten en medewerkers). Van elke docent zijn meerdere lessen geobserveerd, zodat ook inzicht verkregen kan worden in de variatie tussen lessen. In de onderstaande figuur wordt hiervan een overzicht gegeven voor één aspect: Didactisch handelen. In deze figuur representeert elk bolletje één docent, en de horizontale lijnen de verschillen tussen de lessen van de betreffende docent.

Figuur 9. Gemiddelde (o) en variatie (I) in oordelen over het didactisch handelen van alle docenten van één school.



Duidelijk is in Figuur 9 dat de oordelen over het 'Didactisch handelen' van de docenten op deze school verschillen; gemiddeld variëren de oordelen van 'meer zwak dan sterk' tot zonder meer 'sterk'. De

oordelen over de taallessen nemen de hele bandbreedte in beslag; voor alle vier de schooltalen zijn er op deze school sterke maar ook zwakkere docenten geobserveerd. Docenten die klassen zonder enige moeite aan het werk zetten, waar de leerlingen vanzelfsprekend in de vreemde taal spraken, maar ook lessen waarin de docent op zijn of haar kop kan gaan staan en de leerlingen van alles deden maar niet dat wat ze geacht werden te doen.

Hoewel er duidelijke verschillen tussen docenten geconstateerd zijn, is ook in één oogopslag duidelijk dat er een enorme variatie is binnen een docent. Zo kan het gemiddelde oordeel wel 'meer sterk dan zwak' zijn, toch kan afhankelijk van de klas waaraan lesgegeven wordt, er een zwakke les geobserveerd zijn omdat de klas zo lastig is. Dat dergelijke informatie van belang is voor de evaluatie van het onderwijs op deze school spreekt voor zich. Want, wat er in een les gebeurt, wat een docent werkelijk doet, dat weten we vaak niet precies. Te vaak wordt de deur van het klaslokaal gesloten, waardoor intervisie weer onderaan het prioriteitenlijstje beland.

Zelfevaluatie van scholen was aanleiding voor een klein experiment, waarbij docenten na elke les een lesbeoordelingsformulier invulden, en een andere docent achter in de klas eenzelfde formulier invulde. In een andere les waren de rollen omgekeerd. Discrepancies tussen de lesgevende en de beoordelende docent waren er natuurlijk te over, waarbij de beoordelende docent over het algemeen strenger bleek dan de lesgevende docent. Maar, deze verschillen waren aan de onderkant van de schaal veel groter dan aan de bovenkant; als een lesgevende docent een les als matig beschouwde dan was het oordeel van de beoordelende docent daar over het algemeen ver onder. Op grond van deze verschillen concludeer ik dat we vaak met een te roze bril naar onze eigen lessen kijken, en strenger zijn voor anderen dan voor onszelf. Oftewel, een docent is ook maar een mens. Echter, dat geeft wel stof tot gesprek! En, is dus een begin voor verbetering van ons onderwijs!

Dames en heren, de conclusie van deze rede is dat aan elke evaluatie theorie ten grondslag ligt: we kunnen niet valide evalueren zonder een adequate theorie over hetgeen we evalueren én een theorie over evalueren zelf. Helaas blijkt maar al te vaak dat tenminste één van

beide theorieën vergeten is op het moment dat we aan het evalueren slaan.

Aan het einde van deze redevoering wil ik van de gelegenheid gebruik maken om de Vereniging van Leraren in Levende Talen en het College van Bestuur van de Universiteit Utrecht te bedanken voor het in mij gestelde vertrouwen. Ik hoop dat ook waar te maken.

Op de tweede plaats wil ik alle promovendi bedanken, ik heb van hen meer geleerd dan zij ooit van mij hebben kunnen doen!

Ook al mijn collega's en oud-collega's wil ik niet ongenoemd voorbij laten gaan; veel problemen in onderwijs en onderzoek ontstijgen de mogelijkheden van een individuele docent en onderzoeker; zonder hun collegiale adviezen en hulp zou ik vaak de mist ingegaan zijn.

Drie collega's wil ik met naam noemen: de altijd positieve Ted Sanders die altijd ewel een positieve constructieve opmerking over weet te maken, Gert Rijlaarsdam met wie ik nu al bijna een kwart eeuw onderzoek mag doen, en Michel Zwarts, die altijd en overal een manier weet om de gegevens op een inventieve manier te analyseren.

Onderwijsonderzoek is onmogelijk zonder de belangeloze medewerking van docenten, en leerlingen die vaak niets meer terugkrijgen dan een voor hen nutteloos onderzoeksverslag en de laatsten zelfs dat nog niet eens. Mijn grootste dank gaat dan ook naar hen uit!

En, natuurlijk mijn lief!

Literatuur

- Bergh, H. van den (1988). Open en meerkeuze-vragen tekstbegrip: Een onderzoek naar de relatie tussen prestaties op open en meerkeuze-vragen. *Tijdschrift voor Onderwijsresearch*, 12, 304-312.
- Bergh, H. van den, Gloppe, K. de & Schoonen, R. (1987). Directe metingen van schrijfvaardigheid: Validiteit en taakeffecten. In, F.H. van Eemeren & R. Grootendorst (Red.), *Taalbeheersing in ontwikkeling*. Dordrecht: Foris Publications.
- Bergh, H. van den (1990b). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement*, 14, 1-12.
- Bergh, H. van den & Zwarts, M. (2004). Aantallen observaties bij de evaluatie van het onderwijsleerproces. Inspectie van het Onderwijs: De Meern.
- Braaksma, M. Rijlaarsdam, G., Bergh, H. van den & Hout Wolters, B.A.M. (2004). Observational learning and its effects on the orchestration of the writing process. *Cognition & Instruction*, 22, 1-36.
- Buelens, G. (2006). Bom onder hoger onderwijs. Rendementsdenken en internationaliseren in het Bolognatijdperk. *Ons Erfdeel. Vlaams-Nederlands cultureel tijdschrift*, 49, 1, pp. 3-17.
- Groot, A.D. de & Van Naerssen, R.F. (1973). *Studietoetsen construeren, afnemen, analyseren* (2e druk). Den Haag: Mouton.
- Hoeven, J. Van der (Pedagogiek UU, 1997). *Children's composing: a study into the relationship between writing processes, text quality, and cognitive and linguistic skills*. Rodopi: Amsterdam.
- Inspectie van het onderwijs (1999). *Onderwijsleerproces in de basisvorming: Evaluatie rapport*. Utrecht: Inspectie van het Onderwijs.
- Janssen, T., Braaksma, M., Rijlaarsdam, G. & Bergh, H. van den (2006). Flexibility in reading literature; Differences between weak and strong adolescent readers. *Journal of Educational Psychology* (submitted for publication).
- Kamalski, J., Sanders, T., Lentz, L. & Bergh, H. van den (2005). Hoe kun je het beste meten of een leerling een tekst begrijpt? Een vergelijkend onderzoek naar vier methoden. *Levende Talen*, 4, 3-9.
- Kieft, M. (2006). *The effects of adapting writing instruction to students' writing strategies*. Amsterdam: Uva.
- Land, J., Sanders, T. & Bergh, H. van den (2006) Wat maakt een studietekst geschikt voor vmbo-leerlingen? Stichting Lezen: Amsterdam (<http://www.lezen.nl/index.html>).
- Mayer, S. de & Rijmenans, R. (2004). *Onderzoek naar kenmerken van effectieve scholen: Kritische factoren in een onderzoek naar de schooleffectiviteit in het technisch en beroepssecundair onderwijs in Vlaanderen*. Academia Press: Antwerpen.
- Rijlaarsdam, G. & Bergh, H. van den (2006). Writing process theory: A functional dynamic approach. In, C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. The Guilford Press: London, New York.
- Schooten, E. Van & Gloppe, K. De (1990). *De validiteit van meerkeuze-instrumenten voor*

- het meten van schrijfvaardigheid*. Tijdschrift voor Taalbeheersing, 12, 93-110.
- Slagter, P.J. (2006). How do I know that my Level B1 is your Intermediate Mid? The Common European Framework of Reference for Languages, the ACTFL Guidelines for Writing, and Assessment of Student Writing. *Modern Language Journal* (Submitted for publication).
- Slagter, P.J. (2006). www.let.uu.nl/spaans/games/contar1_10/game1a.html.
- Tillema, M. (2005). *De relatie tussen woordenschat en taakuitvoering bij het schrijven in een vreemde taal*. Paper gepresenteerd op 7 September op het VIOT-congres te Nijmegen.
- Tweede fase adviespunt. (2005). Zeven jaar tweede fase, een balans. <http://www.tweede-fase-loket.nl/evaluatie>.
- Ubels, P. (2006). *Beoordeling van schrijfproducten door Nederlandse en Engelse docenten met én zonder het Europees Referentie Kader*. Doctoral scriptie: Universiteit Utrecht.
- Weijen, D. van, Bergh, H. van den, Sanders, T. & Rijlaarsdam, G. (2006) *Schrijven: Plannen en de eheid van analyse (in voorbereiding)*.
- Wesdorp, H. (1974). *Het meten van productief-schriftelijke taalvaardigheid*. Muusses: Purmerend.

Curriculum Vitae

Huub van den Bergh ('s Gravenhage 1957) studeerde onderwijspsychologie aan de Universiteit van Amsterdam. Vanaf 1983 was hij werkzaam bij de Stichting Centrum voor Onderwijsonderzoek (nu bekend als SCO-Kohnstamm-instituut). Daar deed hij onder leiding van Hildo Wesdorp onderzoek naar het niveau van het taalonderwijs aan het einde van de basisschool, en naar de validiteit van de examens Nederlands voor MAVO en VBO. Op dat laatste onderwerp promoveerde hij in 1989 aan de Universiteit van Amsterdam.

Na zijn promotie maakte hij de overstap van Amsterdam naar Utrecht. Hij was betrokken bij diverse grootschalige projecten als periodieke peiling van het onderwijsniveau in het voortgezet onderwijs (Cito), en de evaluatie van de basisvorming (Inspectie van het Onderwijs) en relaties tussen schoolexamen en centraal examen.

Bij de UU was hij als medesupervisor betrokken bij diverse promotieonderzoeken op veel terreinen, variërend van schrijfdidactisch onderzoek, schrijfprocesonderzoek, onderzoek naar de ontwikkeling van kritische periodes in het leren van het Duits, onderzoek naar leesgedrag, onderzoek naar effectief onderwijs, onderzoek naar effecten van onderwijspolitiek op het hoger onderwijs, tot meettechnische en meer methodische studies over taalonderwijs.

De laatste uitgaven in deze reeks zijn:

- Leen Dorsman, *Een hippopotamus in toga en andere hooggeleerden Utrechtse professoren 1815-1940* (2003)
- Ton Naaijkens, *De wegen van de vertaling* (2003)
- Jan Odijk, *Herbruikbare woorden en regels* (2003)
- Renger de Bruin, *Ingelijste regenten. Een museale collectie als onderzoeksterrein voor bestuurs-geschiedenis* (2003)
- Frits van Oostrom, *Academische kwesties. Van middeleeuwse literatuur naar universiteit en maatschappij* (2003)
- Henk Verkuyl, *Woorden, woorden, woorden* (2003)
- Wiljan van den Akker & Gillis Dorleijn, *De Muze: een vrouw met den blik van een man. Opvattingen over 'mannelijkheid' en 'vrouwelijkheid' in de Nederlandse poëzie tussen 1880 en 1940* (2003)
- Sonja de Leeuw, *Hoe komen wij in beeld? Cultuurhistorische aspecten van de Nederlandse televisie* (2003)
- Johannes J.G. Jansen, *De radicaal-islamitische ideologie: van Ibn Taymiyya tot Osama ben Laden* (2004)
- Hermine Joldersma, *The International Dimension of Middle-Netherlandic Song* (2004)
- Ido de Haan, *Politieke reconstructie. Een nieuw begin in de politieke geschiedenis* (2004)
- Pieta van Beek, *'Poeta laureata': Anna Maria van Schurman, de eerste studente in 1636* (2004)
- Sieb Nooteboom, *Waar komen de letters van het alfabet vandaan?* (2004)
- Jan Luiten van Zanden, *De timmerman, de boekdrukker en het ontstaan van de Europese kenniseconomie. Over de prijs en het aanbod van de Industriële Revolutie* (2004)
- Thijs Pollmann, *Aftellen* (2004)
- Ted Sanders, *Tekst doordenken. Taalbeheersing als de studie van taalgebruik en tekstkwaliteit* (2004)
- Keetie E. Sluyterman, *Gedeelde zorg. Maatschappelijke verantwoordelijkheid van ondernemingen in historisch perspectief* (2004)
- Joost Kloek, *Een scheiding van tafel en bed (met verweerde kinderen)* (2004)
- Monique Moser-Verrey, *Isabelle de Charrière and the Novel in the 18th Century* (2005)
- Paul Op de Coul, *De opmars van de operaregisseur. Een encensering van Mozarts Zauberflöte uit 1909* (2005)
- Peter de Voogd, *Laurence Sterne's Maria uitgebeeld: boekillustratie en receptiegeschiedenis* (2005)
- Nicole Pellegrin, *Entre inutilité et agrément. Remarques sur les femmes et l'écriture de l'Histoire à l'époque d'Isabelle de Charrière (1740-1806)* (2005)
- Berteke Waaldijk, *Talen naar cultuur. Burgerschap en de letterenstudies* (2005)
- Orlanda Soei Han Lie, *Wat bezielt een mediëvist? Mastering the Middle Ages* (2005)
- Sjef Barbiers, *Er zijn grenzen aan wat je kunt zeggen* (2006)

Colofon

Copyright: Huub van den Bergh
Vormgeving en druk: Labor Grafimedia BV, Utrecht
Deze oplage is gedrukt in een oplage van 300
genummerde exemplaren, waarvan dit nummer is.
Gezet in de PBembo en gedrukt op 120 grams papier Biotop.
ISBN 90-76912-69-6
Uitgave: Faculteit Geesteswetenschappen/Letteren,
Universiteit Utrecht, 2006.
Het ontwerp van de reeks waarin deze uitgave verschijnt is
beschermd.

