

Ten Geleide

Dit nummer opent met een bijdrage van Huub van den Bergh. Hij stelt dat veel empirisch onderzoek uitspraken doet over gemiddelden, maar dat er meer recht wordt gedaan aan de onderzoeksgegevens door (ook) verschillen in variantie systematisch te bestuderen, onder meer door de verschillen tussen leerlingen onder de loep te nemen. Zo laat de auteur zien dat instructie in leesstrategieën een aantal leerlingen helpt beter te lezen, maar niet alle leerlingen. Dit verschil heeft te maken met de verschillen in leesvaardigheid. Alle redenen dus om als docent een breed didactisch repertoire tot de beschikking te hebben om dezelfde leesstrategieën op verschillende manieren te onderwijzen; met één manier van aanbieden zijn niet alle leerlingen in dezelfde mate gediend.

In het tweede artikel presenteren Ingrid Evers en Piet-Hein van de Ven hun onderzoek naar de kennis waarover leraren zelf dienen te beschikken. Het betreft leraren die lesgeven in de nationale taal en leraren van andere vakken die de nationale taal gebruiken in hun onderwijs. Voor het onderzoek zijn in een paar Europese landen bestaande documenten, vooral leerplannen, geanalyseerd. Op grond van de onderzoeksresultaten willen de auteurs pleiten voor meer discussie in Nederland over de rol van taal in het onderwijs. Wat moeten leraren nu precies kennen, wat is het belang van taal, het belang van het schoolvak, de positie van de leraar?

Het volgende artikel beschrijft het onderzoek van Mia Stokmans en Peter Broeder naar leesomvang in de vrije tijd van leerlingen in het voortgezet onderwijs, waarbij het leesgedrag van allochtone leerlingen vergeleken wordt met die van hun allochtone leeftijdsgenoten (Turkse en Marokkaanse middelbare scholieren). Een interessante bevinding is dat autochtone leerlingen in hun vrije tijd net zo veel blijken te lezen als hun allochtone leeftijdsgenoten. Ook is het opmerkelijk dat de 'leesnorm' van de ouders (wat ze zeggen én wat ze doen) in allochtone en autochtone gezinnen niet verschilt.

Verder in het Tijdschrift aandacht voor de leerpiramide van Bales, de piramide die stelt dat we het minst leren van wat we horen en het meest van wat we aan anderen uitleggen. Hoe zit het eigenlijk met het empirisch gehalte van deze in Nederland dikwijls geciteerde piramide? Leraren en materiaalontwikkelaars moderne vreemde talen kunnen hun vakdidactische kennis aanscherpen met behulp van het onlangs verschenen handboek van Erik Kwakernaak, *Didactiek van het vreemdetalenonderwijs*. De rubriek Gesignaleerd staat stil bij deze publicatie. Tot besluit brengt dit nummer het promotieonderzoek van Geertruida Maria Schijf onder de aandacht. Zij heeft onderzoek gedaan naar lees- en spellingvaardigheden van brugklassers.

Namens de redactie,
Carla Driessen

Gemiddelden en varianties

het onderschatte belang van varianties

geïllustreerd aan een studie naar het effect van leesstrategieën

HUUB VAN DEN BERGH

In veel empirisch onderzoek staan verschillen tussen gemiddelden centraal. Veelal wordt aan verschillen in variantie voorbijgegaan. Het gevolg is dat de informatie in de gegevens niet ten volle gebruikt wordt, en kansen op interpretaties gemist worden. In deze bijdrage wordt gedemonstreerd aan de hand van een analyse van een experiment naar effecten van leesstrategieën dat ook de variantie en variantieverschillen een bron van informatie zijn. Als we de verschillende bronnen van variantie niet in een analyse betrekken verliezen we mogelijkheden voor inhoudelijke interpretatie en nuancering van de resultaten.

Inleiding

In vrijwel elk empirisch onderzoek gaat de aandacht uit naar gemiddelden. Bijvoorbeeld: 'Nederlander gelukkigste mens van Europa', 'de gemiddelde Nederlander heeft een IQ van 100'. Zelden worden de verschillen tussen, in dit geval, Nederlanders benadrukt. Wanneer we een experiment doen, gaat alle aandacht uit naar verschillen in gemiddelden; in de experimentele conditie wordt gemiddeld genomen hoger/lager gescoord dan in

de controle conditie. De verschillen tussen leerlingen – of meer in het algemeen: proefpersonen – worden veelal als minder relevant bijproduct beschouwd.

Een nieuwe didactiek werkt, als de leerlingen gemiddeld een hogere score behalen. Immers, dan zijn de leerlingen door de nieuwe, experimentele didactiek gemiddeld genomen meer vooruit gegaan dan in de controle conditie. Er is dan een causaal verband aangetoond tussen de experimentele didactiek en de prestaties van de leerlingen. Echter, aan de variantie die ook vaak toeneemt als gevolg van een experimentele didactiek, wordt op zijn best lippendienst bewezen.

Zoals in elk methodologieboek staat, is het verstandig leerlingen at random aan condities toe te wijzen. Dan zijn immers a priori verschillen in gemiddelde tussen condities verwaarloosbaar. Bovendien kan dan een indicatie verkregen worden van de gemiddelde groei in prestaties; het gemiddelde verschil tussen voor- en nameting, of wel de gemiddelde leerwinst, dan wel het gemiddelde verschil tussen de onderscheiden condities.

Kortom, het is een en al gemiddelde wat de klok slaat. De spreiding in scores is in dit verband niet meer dan een noodzakelijk kwaad. Immers, de vraag of twee gemiddelden sig-

nificant van elkaar verschillen is afhankelijk van de variantie (en de daarvan afgeleide standaarddeviatie en standaardfout). Zo is hetzelfde verschil tussen twee gemiddelden wel significant bij een kleine spreiding, maar niet bij een grotere spreiding.

In het verleden zijn er dan ook tal van manieren bedacht om de variantie te verkleinen, en aldus toch significante resultaten te krijgen. Denk maar aan covariantie-analyse. In feite 'bindt' de covariaat dan een deel van de geobserveerde variantie, waardoor 'scherper' getoetst kan worden.

In deze bijdrage wil ik nader ingaan op deze verwaarlozing van de variantie. Ik wil met drie voorbeelden aantonen hoe belangrijk variantie wel niet is, en dat we zowel in analyse van concrete onderzoeksresultaten als in theorievorming enorme kansen laten liggen.

Een 'simpel' experiment

In een experiment naar de effectiviteit van leesstrategieën lezen leerlingen (uit vier vwo-3 klassen; $N = 106^1$) voor en na het experiment drie teksten met vragen. De teksten zijn afkomstig uit 'Open Forum', een opiniëpagina voor langere artikelen (circa twee A4) in de Volkskrant. Hierbij is gekozen voor teksten met relatief dezelfde opbouw; in de

eerste alinea wordt 'het' probleem geschetst en wordt al een standpunt ingenomen. In het middendeel worden voor- en tegenargumenten gewogen, en in de laatste alinea wordt de oplossing van de schrijver(s) gerecapituleerd.

In de experimentele conditie wordt tussen de beide metingen aandacht besteed aan leesstrategieën (skimming, het lezen van begin en einde van een paragraaf, onderstrepen van sleutelfragmenten, en gebruik van structuurmarkeringen; zie bijvoorbeeld: Bimmel & Van Schooten, 2004). In één les staat één strategie centraal, die wordt uitgelegd, (hardop-denkend) voorgedaan en geoefend. In de twee laatste lessen van de lessenreeks staat de keuze voor een strategie centraal; leerlingen moeten kiezen wanneer een strategie goed gebruikt kan worden en wanneer een strategie minder relevant is. In de andere (controle) conditie lezen de leerlingen teksten en beantwoorden hier vragen over. De tijd die besteed wordt aan lezen is in beide condities gelijk (zes lessen van vijftig minuten). Vanzelfsprekend zijn de leerlingen at random aan de beide condities toegewezen, zodat verschillen in beginniveau géén rol spelen. De gegevens worden gerapporteerd in percentage correct beantwoorde tekstbegripvragen (zie Tabel 1).

Uit de gegevens in Tabel 1 blijkt dat de leerlingen in de experimentele conditie 45.7% van de vragen correct had op de beginmeting en 54.1% van de tekstbegripvragen op de

eindmeting goed had. Voor de leerlingen in de controleconditie zijn dat respectievelijk 45.8% en 47.4%. Het verschil in gemiddelden tussen condities is op de voormeting klein, en op de nameting duidelijk groter.

Standaard variantie-analyse (voor herhaalde waarnemingen) leert dat het interactie-effect tussen conditie en meetmoment significant is ($F(1, 98) = 4.91; p = 0.03$). Dus: het verschil tussen de voor- en nameting is afhankelijk van de conditie. In de experimentele conditie zijn de leerlingen meer vooruit gegaan dan in de controle conditie (zie Tabel 1). Ook het minder relevante hoofdeffect van meetmoment ($F(1, 98) = 5.09; p = 0.03$) is significant. De conclusie ligt dan ook voor de hand: in de experimentele conditie is de vooruitgang tussen voor- en nameting veel groter dan in de controleconditie. Kortom, leesstrategieën helpen leerlingen beter te lezen.

Nu is niet alleen de vraag of een verschil significant is van belang, ook de grootte van het effect is relevant. Het interactie-effect verklaart 5% van de geobserveerde verschillen. Hiermee kan dit effect, in termen van Cohen (1988), als een middelmatig effect geïdentificeerd worden.

Dit is echter niet hele verhaal dat bij een dergelijke analyse hoort. We hebben immers de varianties tot dusver verwaarloosd. Dit geldt zowel de verschillen tussen de drie tests

per meetmoment, als de verschillen tussen leerlingen. Om met dat laatste te beginnen. Uit Tabel 1 blijkt dat de verschillen tussen leerlingen met name in de experimentele conditie toegenomen zijn. In de experimentele conditie zijn de verschillen in tekstbegripscores tussen leerlingen bij de nameting bijna twee keer zo groot als bij de voormeting (10.8 versus 19.7). En, bij de controle conditie is er wat dit betreft nauwelijks een verschil tussen beide standaarddeviaties (11 versus 12). Dat de variantie op de nameting groter is in de experimentele conditie wordt ook door toetsing bevestigd. Op de nameting in de experimentele conditie is de variantie groter dan de variantie op enig ander moment of in enig andere conditie ($\chi^2 > 10.96; df = 1; p < 0.01$).

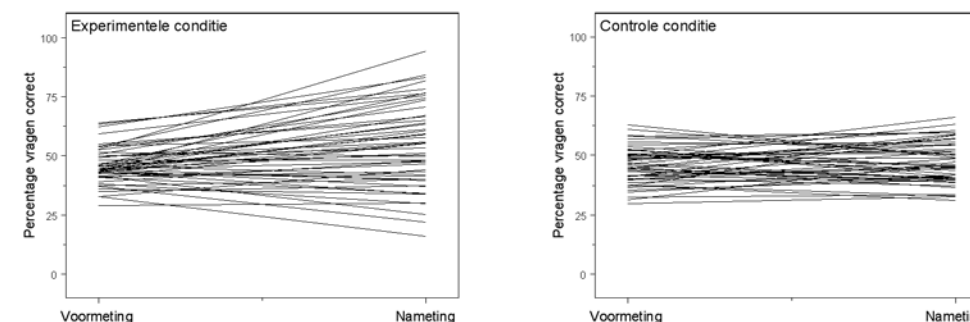
Om inzicht te geven wat dit betekent voor de (som)scores van de individuele leerlingen, en dus inzicht in de verschillen tussen leerlingen is Figuur 1 geconstrueerd. In deze figuur is voor elke leerling in de experimentele en in de controle conditie één lijn getrokken, die de score op de voormeting met die op de nameting verbindt.

In de experimentele conditie gaan leerlingen gemiddeld genomen vooruit. Echter, dat geldt niet voor alle leerlingen. We zien heel duidelijk dat de verschillen tussen leerlingen toegenomen zijn; op de nameting zijn de verschillen tussen leerlingen veel groter dan

Conditie	VOORMETING		NAMETING	
	Gemiddelde	(sd)	Gemiddelde	(sd)
Experimenteel	45.7	(11.1)	54.1	(19.7)
Controle	45.8	(11.0)	47.4	(12.0)

Tabel 1: Gemiddelden (in percentages correcte antwoorden) en standaarddeviaties (sd) voor twee meetmomenten in experimentele ($N = 51$) en controle conditie ($N = 49$)^a.

a. De prestaties van zeven leerlingen zijn niet meegenomen omdat zij in twee of meer lessen afwezig waren (3 leerlingen), de voor- of nameting niet gemaakt hadden (3 leerlingen), of de voormeting dusdanig slecht gemaakt had dat de score niet afweek van een kansscore (1 leerling).



Figuur 1: Individuele scores op voor- en nameting in experimentele en controle conditie.

op de voormeting. In de controle conditie zijn de verschillen tussen leerlingen op beide meetmomenten ongeveer gelijk.

Er mag dan gemiddeld genomen wel sprake zijn van een middelmatig effect, dat geldt zeker niet voor alle leerlingen; sommige leerlingen in de experimentele conditie scoren op de nameting duidelijk lager dan op de voormeting. Deze leerlingen zijn achteruit gegaan! Hier dringt de vraag naar een verklaring zich op. Een dergelijke verklaring zou gezocht kunnen worden in termen van *aptitude treatment interacties* (Cronbach, 1957). Aan dergelijke *aptitude treatment interacties* (ATI's) wordt te vaak te weinig aandacht besteed. Toch blijkt uit onderzoek naar bijvoorbeeld schooleffectiviteit dat een bepaalde (pedagogisch-)didactische setting zelden effectief is voor alle leerlingen (zie bijvoorbeeld: Inspectie van het Onderwijs, 1999; De Maeyer & Rymenans, 2004). We moeten ons realiseren dat uit studies naar taalvaardigheid zelden is gebleken dat één didactische ingreep één type opdracht, of één manier van uitleggen effectief is voor alle leerlingen; dat iets 'werkt' bij de ene leerling is geen garantie dat het ook 'werkt' bij een andere leerling.

Eén van de mogelijkheden zou kunnen zijn dat het effect van leesstrategieën afhankelijk is van de leesvaardigheid van de leerlingen. Daarom is in een variantie-analytisch model het effect van conditie én het effect van de voormeting geschat. Het verschil met

de voorgaande analyse is dat hier de gegevens geanalyseerd zijn in een regressiemodel. In dit model zijn we geïnteresseerd in het gemiddelde (de constante, want de voormeting is gecentreerd), de variantie rond het gemiddelde, de regressie van de nameting op de voormeting en de variantie van dit regressiegewicht (zie Tabel 2).

Uit Tabel 2 blijkt dat voor de gemiddelde leerling (i.e. een leerling met een gemiddelde score op de voormeting) er (gemiddelde) een duidelijk verschil is tussen de beide condities (56.2 in de experimentele conditie en 46.2 in de controle conditie). We zien ook dat de variantie rond de gemiddelde score nauwelijks varieert tussen beide condities (79.1 versus 75.1). Dat betekent dat in de experimentele conditie 80% van de gemiddelde leerlingen een score heeft tussen 47.2 en 65.1; voor de controle conditie loopt het 80% waarschijnlijkheidsinterval van 30.1 tot 52.3

Belangrijker echter voor de vraag naar het effect van leesstrategieën zijn de regressiegewichten en de variantie van deze regressiegewichten. In de experimentele conditie is er een sterker verband tussen voor- en nameting dan in de controle conditie; in de experimentele conditie neemt de score op de nameting 1.1% correct beantwoorde vragen toe per procent correct beantwoorde vragen op de voormeting. In de controle conditie is de score van een leerling met 1% meer vragen correct op de voormeting, 0.3% hoger op de

	EXPERIMENTELE CONDITIE		CONTROLE CONDITIE	
	Gemiddelde	Variantie	Gemiddelde	Variantie
Constante	56.2 (2.7)	79.1 (18.3)	46.2 (1.6)	75.1 (15.0)
β^* Voormeting ^a	1.1 (0.3)	2.3 (1.1) ^b	0.3 (0.1)	0.0 (0.0)

a. Gecentreerd rond het gemiddelde (i.e. voormeting - 45); b Covariantie tussen voor- en nameting 9.7 ($r = 0.60$).

Tabel 2: Parameterschattingen voor een alternatief model voor het effect van leesstrategieën.

nameting. Dit verschil is duidelijk significant ($\chi^2 = 17.4$; $df = 1$; $p < 0.001$). Ook zien we dat de variantie van de beide regressiegewichten verschilt ($\chi^2 = 8.6$; $df = 1$; $p < 0.001$). Dit geeft aan dat de regressiecoëfficiënt in de experimentele conditie wèl afhankelijk is van de score op de voormeting, maar in de controle conditie niet². De regressiecoëfficiënt is hoger voor leerlingen met een hoge score op de voormeting, en kan zelfs negatief worden voor leerlingen met een lage score op de voormeting (het 80%-waarschijnlijkheidsinterval voor de regressie van de score op de nameting op die van de voormeting varieert van -0.8 tot 3.0). Klaarblijkelijk zijn (relatief) leeszwakke leerlingen niet geholpen met het aangeboden onderwijs in leesstrategieën.

Maar we zijn er nog niet. Tot dus ver hebben we de verschillen tussen teksten nog niet in onze analyse betrokken. En, dat is onterecht! Immers, door het maken van somscores is de variantie tussen teksten ten onrechte verwaarloosd. Doordat we de verschillen tussen teksten verwaarlozen, onderschatten we

de variantie, waardoor we te snel geneigd zijn om van een significant verschil tussen condities te spreken. Daar komt bij dat we eigenlijk gelijktijdig willen generaliseren naar twee populaties: de populatie leerlingen én de populatie teksten. Immers, we willen niet alleen aantonen dat er één tekst gevonden kan worden waarvoor geldt dat de experimentele manipulatie (i.e. het aanleren van leesstrategieën) een positief effect heeft, we willen aantonen dat dat effect onafhankelijk is van de tekst. Pas als we kunnen aantonen dat het voor leerlingen wat uitmaakt in welke conditie zij zitten én dat het effect generaliseerbaar is over teksten dan hebben we een echt resultaat (vergelijk: Clark, 1973).

Om een dergelijke conclusie te kunnen onderbouwen moeten we naast de gemiddelden (per meetmoment en conditie) verschillende variatiecomponenten onderscheiden: de variantie tussen teksten, de variantie tussen leerlingen op de voor- en nameting, en een errorvariantie. In totaal moeten dus minimaal negen variatiecomponenten onder-

Parameter	CONDITIE				
	Experimentele		Controle		
	voor	na	Voor	Na	
Gemiddeld	45.7	54.1	45.8	47.4	
S ² (...)	VARIANTIES				
	S ² (error)	15.1	20.0	16.4	12.2
	S ² (personen)	60.3	318.7	62.4	79.6
	S ² (teksten) ^a	46.9		46.7	

a. Eigenlijk zou de variantie tussen teksten ook uitgesplitst kunnen worden naar voor- en nameting. Echter, gezien het gering aantal teksten op elk van de meetmomenten is dit in dit specifieke geval weinig zinvol.

Tabel 3: Gemiddelden (% correct beantwoorde vragen) en varianties uitgesplitst naar conditie en tekst.

scheiden worden. Minimaal negen, want de variantie tussen teksten kan natuurlijk ook verschillen tussen condities en tussen voor- en nameting (al is het eerste hier niet het geval, en het laatste hier weinig zonzvol gezien het geringe aantal teksten).

In Tabel 3 zijn de gemiddelden per conditie, alsmede de verschillende variantiecomponenten weergegeven. We zien (wederom) dat op de voormeting de gemiddelden tussen de twee condities nauwelijks verschillen. Dat geldt ook voor de varianties maar is wat minder duidelijk zichtbaar. Immers, de totale variantie is opgesplitst in drie delen: de variantie binnen personen, de variantie tussen personen en de variantie tussen teksten. In bijvoorbeeld de voormeting in de experimentele conditie is de totale variantie ($15.1 + 60.3 + 46.9 = 122.3$). Dit komt overeen met een standaarddeviatie van 11.1, wat gelijk is aan de standaarddeviatie in deze conditie in Tabel 1 (11.1)³.

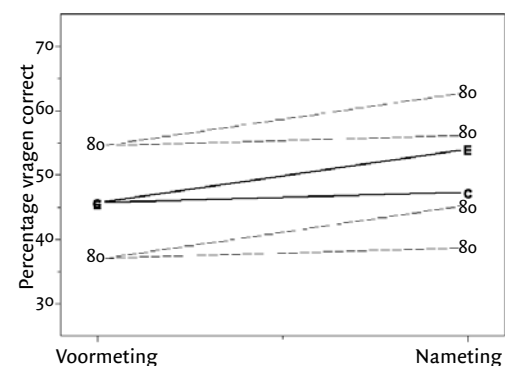
Ook bij deze analyse is het interactie-effect tussen meetmoment en conditie significant ($\chi^2 = 4.77$; $df = 1$; $p = 0.04$). Hetgeen wederom wil zeggen dat de gemiddelden op de nameting wel verschillen ($\chi^2 = 5.44$; $df = 1$; $p = 0.02$), maar op de voormeting niet ($v^2 = 0.07$; $df = 1$; $p = 0.79$).

Het belangrijkste is echter dat er in deze analyse verschillende variantiecomponenten geschat worden. De verschillen tussen teksten zijn relatief groot. In de experimentele conditie is de variantie 46.9 ($sd = 6.8$), en in de controle conditie is deze niet significant kleiner 46.7 ($sd = 6.8$). Dus, afhankelijk van de tekst varieert het gemiddelde enorm. In Figuur 2 is dit grafisch weergegeven door 80% betrouwbaarheidsintervallen te construeren voor de mate waarin de gemiddelde score varieert van tekst tot tekst.

Vanzelfsprekend varieert de gemiddelde score van derdeklassers van tekst tot tekst. Doordat de variantie tussen teksten geschat is, kan een indicatie gegeven worden van de

verschillen in scores louter en alleen door tekstkenmerken. Als we een willekeurige tekst uit de U-pagina nemen (en daar vergelijkbare opdrachten bij geven), dan zullen derdeklassers gemiddeld bij 80% van de teksten tussen de 37% en 54% van de vragen correct beantwoorden; bij 10% van de teksten halen zij een hogere score, en bij 10% een lagere score (zie Figuur 2). Op de nameting halen leerlingen bij 80% van de teksten een score tussen 45% en 63% in de experimentele conditie en tussen 38% en 56% correct beantwoorde vragen in de controle conditie. Het is duidelijk dat de tekst er veel meer toe doet dan conditie. Het is voor de score van een leerling belangrijker welke tekst hij voorgelegd krijgt, dan of hij al dan niet onderwijs in leesstrategieën gevolgd heeft.

Ten tweede blijkt dat de correlatie tussen voor- en nameting op tekstniveau laag is ($r = 0.31$ in de experimentele en $r = 0.34$ in de controle conditie). We moeten dan ook concluderen dat er op de ene tekst wél een verschil is tussen de beide condities, maar dat dit op de andere tekst niet het geval hoeft te zijn. We kunnen ons dan ook niet aan de indruk onttrekken dat het effect van leesstrategieën



Figuur 2: Gemiddelde per conditie voor de voor- en nameting en 80% betrouwbaarheidsintervallen voor de verschillen ten gevolge van tekst (C: Controle conditie; E: Experimentele conditie).

in deze studie niet voor alle teksten aantoonbaar is.

Conclusie

In veel onderzoek blijkt dat de verschillen tussen leerlingen toenemen ten gevolge van onderwijs (zie bij voorbeeld: Anasatasi, 1958; of voor specifieke voorbeelden: Van den Bergh & Hoeksma, 1993 (spelling en lezen); Braaksma, 2002 (schrijfvaardigheid); Overmaat, 1996 (schriftelijke taalvaardigheid); Moonen, 2008 (leren van woorden)). Toch worden deze variantieverschillen nauwelijks gemodelleerd laat staan geproblematiseerd. Hoeveel theorieën hebben bijvoorbeeld expliciet betrekking op de variantie? Omdat variantie, de verschillen tussen leerlingen, in theoretisch opzicht stiefmoederlijk behandeld wordt, hebben veel verwachtingen in, of rationales achter onderzoek ook geen betrekking op de verschillen tussen leerlingen; men focust op verschillen in gemiddelden.

Aan de hand van een eenvoudig experiment is gedemonstreerd dat de verschillen tussen leerlingen door experimenteel onderwijs kunnen toenemen. Het is aannemelijk gemaakt dat een dergelijke heteroscedasticiteit (verschillen in varianties tussen condities/meetmomenten) verklaringen behoeft. Hier kunnen *aptitude treatment* interacties een welkome aanvulling van de theorie zijn. Onderzoek waarin ATT's centraal staan is echter schaars (zie bijvoorbeeld: Braaksma, Rijlaarsdam & Van den Bergh, 2002; Kieft, Rijlaarsdam, Galbraith & Van den Bergh, 2007). De enkele voorbeelden die er zijn, geven wel aanleiding voor verdere theorievorming. In het onderhavige voorbeeld is aannemelijk gemaakt dat het effect van het aanleren van leesstrategieën afhankelijk is van de leesvaardigheid van de leerlingen. Leerlingen met een (relatief) hoge leesvaardigheid hebben meer geprofiteerd van het aanleren van deze leesstrate-

gieën op deze manier. Leerlingen met een (relatief) lage leesvaardigheid hadden geen baat bij het aanleren van deze leesstrategieën op deze manier. In tegendeel, hun score op de nameting is lager dan die op de voormeting. Daarmee is gedemonstreerd dat in dit experiment de *aptitude treatment* interactie – in elk geval deels – verklaard kan worden vanuit het leesvaardigheidniveau van de leerlingen; slechte lezers hebben minder voordeel van het leren van leesstrategieën dan goede lezers. Wederom gedemonstreerd dat zelfs bij een relatief homogene groep (derdeklassers vwo) hetzelfde onderwijs niet even effectief is voor alle leerlingen.

De vraag die resteert, is waarom de relatief slechte lezers geen voordeel hebben van leesstrategieën. Is de keuze voor deze leesstrategieën niet de meest handige voor deze doelgroep, of is de didactiek niet optimaal voor zwakkere lezers? Is wellicht het overschakelen op meer globale begripstrategieën juist een probleem voor (relatief) zwakke lezers, die wellicht meer gericht zijn op het stapsgewijs construeren van een cognitieve representatie van de tekst, dan de opbouw van een situatiemodel waarin relaties tussen tekstelementen én relaties met de al aanwezige kennis gelegd worden (vergelijk: Land, 2008)? Ongeacht de antwoorden op deze vragen, lijkt het praktisch gezien noodzakelijk dat docenten een breed didactisch repertoire tot hun beschikking hebben om dezelfde strategieën op verschillende manieren te onderwijzen; met één manier van aanbieden, zoals in deze studie, zijn niet alle leerlingen in dezelfde mate gediend. Het is dan ook verstandig om altijd te controleren of alle leerlingen de uitleg begrepen hebben.

Ook is gedemonstreerd dat effecten afhankelijk kunnen zijn van het gebruikte instrumentarium. Veel van ons onderzoek wordt helaas verricht op basis van een zogeheten 'single message design' (Meuffels & Van den Bergh, 2005). In zulke designs wordt er één

operationalisatie van de afhankelijke of onafhankelijke variabele gebruikt. Bijvoorbeeld: één tekst met bijbehorende vragen (terwijl we weten dat er grote verschillen zijn tussen teksten). Of, in het geval men geïnteresseerd is in effecten van aanspreekvormen (bijvoorbeeld: 'u' versus 'je') in charitatieve brieven, één operationalisatie van een 'je-brief' en één operationalisatie van een 'u-brief'. Met zo'n onderzoeksontwerp kan helaas niets aannemelijk, dan wel onaannemelijk gemaakt worden (vergelijk: Meuffels & Van den Bergh, 2005, 2006; Mulder, 2008). We moeten ons beseffen dat we in veel gevallen willen generaliseren naar twee populaties: de populatie leerlingen én de populatie teksten. Op basis van *single message designs* is dat laatste onmogelijk.

Een derde conclusie, die tot op heden nog slechts impliciet naar voren is gekomen, betreft de gehanteerde statistiek. In veel gevallen grijpen we direct naar de toetsende statistiek in pakketten als SPSS, BMDP en SAS, waarin we de platgetreden paden volgen. Ik denk dat we in de toekomst meer aandacht zullen gaan besteden aan individuele diagnostiek. Dan gaat het niet (alleen meer) om de vraag of er een gemiddeld verschil is tussen condities, maar om de vraag voor welke, of welk type, leerlingen de conditie een verschil uitmaakt? Waarom wijkt de groei voor deze leerling(en) zo veel af van het gemiddelde? Kunnen we dit wellicht toeschrijven aan ... Vreemd genoeg wordt statistiek daarmee van toepassing op individuen in de steekproef. Eigenlijk deden we dat altijd al, maar waren we daar nooit echt expliciet over (immers, variantie is al een afwijking van het gemiddelde). Kortom, statistiek wordt flexibeler gebruikt, waarbij niet alleen de gemiddelden geschat worden, maar waar ook de (verschillen) in variantie(s) expliciet gemodelleerd worden.

Het gaat dan niet alleen om meer verschillen in gemiddelden, maar om individuele ont-

wikkelings- of groeicurven. De bepaling van dergelijke curven is misschien wel wat ingewikkelder, maar de opbrengst is veel genuanceerder. Bovendien bieden dergelijke analyses zo veel meer mogelijkheden tot interpretatie (zie bijvoorbeeld de figuren 1 en 2), dat het ondenkbaar is dat dergelijke modellen in de nabije toekomst niet veelvuldig gebruikt gaan worden. Eén en ander staat los van de voorkeur voor zulke analyse vanuit een statistisch oogpunt. Immers, door rekening te houden met verschillen in variantie ten gevolge van de experimentele manipulatie wordt ook een zuivere schatting van de populatiegemiddelden verkregen. Het incorporeren van verschillen in variantie in de analyse heeft dan ook een duidelijk voordeel.

Dergelijke analyses verdienen niet alleen vanuit een statistisch oogpunt de voorkeur, maar bieden ook veel meer mogelijkheden tot inhoudelijk interpretatie van onderzoeksresultaten.

Noten

1. De gegevens zijn verzameld door studenten in het kader van de Educatieve Master van de UU.
2. Opgemerkt zij dat een dergelijk interactie-effect op een groot aantal manieren geschat kan worden. In het verleden is wel aanbevolen de groep in drieën te splitsen (bijvoorbeeld: hoog, gemiddeld en laag scorenden op de voormeting) en voor elke groep een apart regressiegewicht te schatten.
3. Alleen gerekend naar Mean Squares zijn de varianties geheel anders omdat het telkens om een ander aantal waarnemingen gaat (6 voor teksten; 100 voor personen en 600 voor de interactie tussen persoon en tekst).

LITERATUUR

Anastasi, A. (1958). *Differential psychology*. New York: Macmillan

Bergh, H. van den & Hoeksma, J.B. (1993). Modeling development in education: A three level model. J.H.L. Oud & R.A.W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioural sciences*. Nijmegen: ITS.

Bimmel, P., & Schooten, E. van. (2004). The relationship between strategic reading activities and reading comprehension. *L1-Educational studies in Language and Literature*, 4, 85-102.

Braaksma, M., Rijlaarsdam, G. & Bergh, H. van den. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, 94, 405-415.

Clark, H.H. (1973). The language-as-fixed-effect fallacy. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.

Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, N.J.: Lawrence Erlbaum ass.

Kieft, M., Rijlaarsdam, G. Galbraith, D., & Bergh, H. Van den. (2007). The effects of adapting a writing course to students' writing strategies. *British Journal of Educational Psychology*, 77, 565-578.

Inspectie van het Onderwijs (1999). *Werk aan de basis: Evaluatie van de basisvorming na vijf jaar*. Utrecht: Inspectie van het onderwijs.

Land, J.F.H. (2008). *Zwakke lezers, sterke teksten: experimenteel onderzoek naar effecten van tekststructuur op begrip en waardering*. Utrecht: Stichting Lezen.

Mayer, S. de & Rymenans, R. (2004). *Onderzoek naar kenmerken van effectieve scholen*. Antwerpen: Academia Press

Meuffels, B., & Bergh, H. van den. (2005). De ene tekst is de andere niet. De language-as-a-fixed-effect fallacy revisited: methodologische implicaties. *Tijdschrift voor Taalbeheersing*, 27, 106-125.

Meuffels, B., & Bergh, H. van den (2006).

De ene tekst is de andere niet: The language-as-a fixed-effect fallacy revisited: Statistische implicaties. *Tijdschrift voor Taalbeheersing*, 28, 323-245.

Moonen, M. (2008). *Testing the multi-feature hypothesis: Tasks, mental actions and second language acquisition*. Utrecht: IVLOS.

Mulder, G. (2008). *Understanding causal coherence relations*. Utrecht: LOT.

Overmaat, M. (1996). *Schrijven en lezen met tekstschemas: Effectief onderwijs in schriftelijke taalvaardigheid in de bovenbouw van het voortgezet onderwijs*. Amsterdam: SCO-Kohnstamm instituut.

HUUB VAN DEN BERGH geeft onderwijs bij het Departement Nederlands en doet onderzoek bij het Uil/OTS van de universiteit Utrecht, alsmede bij het ILO van de Universiteit van Amsterdam. Hij was betrokken bij diverse grootschalige onderzoeken (zoals periodiek peilingsonderzoek in Nederland en Vlaanderen, en de evaluatie van de Basisvorming). Maar is ook betrokken bij vele kleinschalig studies naar lees- en schrijfprocessen.

E-mail: <Huub.vandenBergh@let.uu.nl>.