

## LITERATUUR

- Beeker A., Canton, J. & Fasoglio, D. (2008). *Chinees op school: Voorstel voor een leerplan Chinese Taal en Cultuur voor het vwo*. Enschede: SLO.
- Bennett, M. (1997) How not to be a fluent fool: Understanding the cultural dimension of language. In A.E. Fantini & J.C. Richards. *New Ways in Teaching Culture*. New Ways in TESOL Series II: Innovative Classroom Techniques, Alexandria, VA: TESOL.
- Byram, M. (1991). *Teaching Culture and Language: Towards an Integrated Model*. In D. Buttjes & M. Byram (red). *Mediating Languages and Cultures*. Bristol: Multilingual Matters.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.
- Byram, M., Gribkova, B. & Starkey, H. (2002). *Developing the Intercultural Dimension in Language Teaching, A Practical Introduction for Teachers*. Strasbourg: Council of Europe.
- Corbalan, G., Folmer, E. & Tigelaar, D. (2012). *Curriculumevaluatie Chinese taal en cultuur*. Enschede: SLO.
- Fasoglio, D. & Canton, J. (2009). *Vreemdetalenonderwijs: een (inter)cultureel avontuur?* Enschede: SLO.
- Hanban/Confucius Institute (2010). *Internationaal Curriculum voor Chinees Taalonderwijs*. Beijing: Foreign Language Teaching and Research Press.
- Hofstede, G. & Hofstede, G.J. (2005) *Allemaal andersdenkenden, omgaan met cultuurverschillen*. Amsterdam/Antwerpen: Contact.
- Kramsch, C. (1996). The Cultural Component of Language Teaching. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*. Geraadpleegd op 2-6-2012 op <[http://www.spz.tu-darmstadt.de/project\\_ejournal/jg-01-/beitrag/kramsch2.htm](http://www.spz.tu-darmstadt.de/project_ejournal/jg-01-/beitrag/kramsch2.htm)>.
- Ochs, E. & Schieffelin, B. B. (1984). Language acquisition and socialization. In R.A.

Shweper and R.A. Levine (red.) *Culture Theory: Essays on Mind, Self, and Emotion*. (pp. 277-320). Oxford: Oxford University Press.

Tigelaar, D., Corbalan, G., & Folmer, E. (2012). *Chinees als nieuw schoolvak?* *Levende Talen Tijdschrift* 13(2), 3-13.

GWENDOLYN TATES (1959) is sinoloog met jarenlange zakelijke China-ervaring. Tegenwoordig deelt zij haar passie voor China door het geven van workshops en trainingen, maar ook als docent Chinees bij de Haagse Hogeschool en bij het Rijnlands Lyceum in Wassenaar. In 2012 behaalde zij (cum laude) de Educatieve Master Chinese Taal en Cultuur, ICLON, Rijksuniversiteit Leiden. E-mail: <[gwendolyn@chinaways.nl](mailto:gwendolyn@chinaways.nl)>

JUDITH ZOETELIEF (1976) studeerde sinologie aan de universiteit van Leiden en Taipei, Taiwan en behaalde (cum laude) in 2012 een Educatieve Master aan het ICLON te Leiden. Zij is het sectiehoofd Chinese Taal en Cultuur van het Sint-Montfort College te Rotterdam, pilotschool Chinees. Daarnaast is zij werkzaam als onafhankelijk wetenschappelijk onderzoeker en gespecialiseerd in de Chinese aanwezigheid in Afrika. E-mail: <[jzoetelief@lmc-vo.nl](mailto:jzoetelief@lmc-vo.nl)>

## Kan een meerkeuzetoets bijdragen aan het meten van schriftelijke taalvaardigheid?

THEO PULLENS, HANNY DEN OUDEN,

WOLFGANG HERRLITZ & HUUB VAN DEN BERGH

Het beoordelen van schrijfvaardigheid is een complexe materie. Om de schrijfontwikkeling van leerlingen te kunnen volgen is er behoefte aan een valide en betrouwbaar instrument. Mogelijk kan de meerkeuzetoets hiervoor ingezet worden: op het eerste gezicht kan met deze toetsvorm snel en betrouwbaar gemeten worden. In dit onderzoek zijn de scores op een meerkeuzetoets schrijfvaardigheid vergeleken met de globale schaalbeoordeling van acht opstellen van dezelfde groep basisschoolleerlingen. Er kan geen positief verband aangetoond worden, met uitzondering van het onderdeel spelling.

De laatste tijd is er een toenemende aandacht waarneembaar in het basis- en voortgezet onderwijs voor de schriftelijke taalvaardigheid bij het moedertaalonderwijs. In een kritisch rapport over het schrijven van teksten in het basisonderwijs spreekt de onderwijsinspectie grote zorgen uit over de kwaliteit van het schrijfonderwijs (Henkens, 2010). Onlangs nog verscheen de publicatie 'Focus op schrijven' (Inspectie van het Onderwijs, 2012) waarin enkele aanwijzingen geformuleerd zijn om de didactiek van het schrijven van teksten in het basisonderwijs te verbeter-

ren. Men pleit er onder meer voor om instrumenten te ontwikkelen waarmee leraren de schrijfvorderingen van leerlingen kunnen volgen. Hiermee zou dan bepaald kunnen worden welke leerlingen de goede en welke de zwakke schrijvers zijn. De inspectie redeneert dat leraren op deze manier handvatten aangereikt krijgen om zwakkere schrijvers gerichte ondersteuning te bieden in specifieke onderdelen van het schrijfproces.

Daarnaast wordt de mogelijkheid onderzocht om, net als in de eindtoets basisonderwijs, schrijfvaardigheid op te nemen in de nieuw te ontwikkelen diagnostische tussentijdse toets voor het voortgezet onderwijs. In het streven om ook voor schrijven een objectieve toets te construeren wordt bij de eindtoets basisonderwijs al jarenlang gebruik gemaakt van een meerkeuzetoets. Het is de vraag of met behulp van deze toetsvorm een betrouwbare en valide uitspraak gedaan kan worden over deze taalvaardigheid.

In dit artikel gaan we allereerst in op de beoordelingsproblematiek bij schrijfvaardigheid. Daarna doen we verslag van een onderzoek naar het inzetten van een meerkeuzetoets als beoordelingsinstrument voor schrijven. Hiermee willen we een bijdrage

leveren aan de discussie over nut, noodzaak en inrichting van het toetsen van schrijfvaardigheid in het basis- en voortgezet onderwijs.

### Beoordelingsproblematiek

De inspectie (Henkens, 2010) signaleert dat in de dagelijkse praktijk van het basisonderwijs de voortgang van de ontwikkeling in schrijfvaardigheid nauwelijks zichtbaar wordt gemaakt, dit in tegenstelling tot andere taalvaardigheden, zoals spellen en lezen. Zeer waarschijnlijk heeft deze situatie te maken met de complexiteit van het geven van oordelen over geschreven teksten. In studies naar de betrouwbaarheid en validiteit van metingen van schrijfvaardigheid geven Wesdorp (1974) en Schoonen (1991) een overzicht van de bruikbaarheid van verschillende toetsinstrumenten en de problemen bij verschillende soorten tekstbeoordelingen.

#### *Verschillen binnen schrijvers en taken*

De knelpunten bij het beoordelen zijn op de eerste plaats terug te voeren op de zogenaamde schrijversvariabiliteit. Een schrijver presteert niet altijd constant bij verschillende schrijfoopdrachten. Hiervoor worden enkele verklaringen genoemd (Van den Bergh & Meuffels, 2000). Een schrijver kan zich bijvoorbeeld beter thuis voelen bij het schrijven van een verhaal dan bij het schrijven van een brief. De schrijfprestaties van een leerling kunnen variëren afhankelijk van de affiniteit met het onderwerp en de achtergrondkennis die hij heeft (Stokking, 2000). Tevens kunnen de toegestane schrijftijd, de communicatieve stijl (bijvoorbeeld een persuasieve, informatieve of diverterende schrijfoopdracht) en de omstandigheden waaronder de schrijfoopdracht wordt gemaakt van invloed zijn op de schrijfprestatie (Braddock, Lloyd-Jones, & Schoer, 1963; Wesdorp, 1974). Dit betekent dat het niet mogelijk is om op basis van één

enkele schrijfoopdracht een betrouwbaar beeld te krijgen van de schrijfvaardigheid van een leerling.

Daarnaast blijken er grote verschillen te zijn, gerelateerd aan de verschillen in taken. De schrijfprestaties lopen immers uiteen afhankelijk van de verschillende soorten opdrachten en van de tekstsoorten die gevraagd worden, en ten slotte zijn prestaties sterk afhankelijk van het onderwerp (Wesdorp, 1981). Schoonen (1991) concludeert dat de voorspellende waarde van de prestaties van de ene schrijftaak laag is ten opzichte van de andere.

#### *Verschillen bij beoordelaars*

Een ander probleem is de betrouwbaarheid van beoordelaars van opstellen. Beoordelaars hebben grote moeite om over dezelfde schrijfproducten een zelfde oordeel te vellen, waardoor de betrouwbaarheid van de beoordeling niet stabiel te noemen is. Zowel binnen als tussen beoordelaars zijn verschillen zichtbaar. Schoonen (1991) stelt dat deze beoordelaarsvariabiliteit allereerst veroorzaakt kan worden doordat de beoordelingstaak onduidelijk is gedefinieerd. Daarnaast is het bekend dat een groot aantal spellingfouten een negatief effect heeft op de beoordeling van bijvoorbeeld de stijlkenmerken van een opstel. Ook oppervlakkige kenmerken, zoals tekstlengte en uiterlijke verzorging, hebben invloed op de beoordeling (De Glopper, 1988). De volgorde van de opstellen die nagekeken worden, kan zorgen voor ongewenste sequentie-effecten. Een sterk opstel dat aan het begin moet worden beoordeeld, kan een stempel drukken op de beoordeling van de volgende opstellen (Schoonen, 1991).

#### *Typen beoordelingen van schrijfvaardigheid*

Elke methode van beoordelen heeft te kampen met eigen knelpunten. Wesdorp (1981) onderscheidt vier typen beoordelingen: globale beoordeling, analytische beoordeling,

schaalbeoordeling, en objectieve toetsing. De globale tekstbeoordeling is relatief makkelijk en snel uitvoerbaar. De beoordelaar heeft de vrijheid om eigen criteria aan te leggen waarop hij de tekst beoordeelt. Maar dit zorgt ervoor dat dit type beoordeling niet erg objectief is. De globale oordelen geven een totaalindruk van de geschreven tekst, maar bieden niet de mogelijkheid om nauwkeurige analyses te maken van bijvoorbeeld de inhoud, de structuur of de stijlmiddelen in de tekst.

Bij analytische tekstbeoordelingen wordt de complexe schriftelijke taalvaardigheid uiteengehaald in een beperkt aantal deelvaardigheden die apart beoordeeld worden. Van den Bergh en Meuffels (2000) zetten vraagtekens bij de betrouwbaarheidswinst die analytische beoordelingsschema's opleveren.

Om tegemoet te komen aan grote beoordelaarsvariabiliteit bij globale beoordelingen en de lage betrouwbaarheid van analytische oordelen, zijn ook wel zogenaamde schaalbeoordelingen voorgesteld. Hierbij is het de bedoeling dat de beoordelaar zijn beoordeling afmeet aan een vooraf gegeven voorbeeldopstel, al dan niet met behulp van beoordelingscriteria. De te beoordelen opstellen worden dan op een schaal van slecht naar goed gezet. Schoonen (1991) rapporteert hoge beoordelaarsbetrouwbaarheden van rond .80 indien er gewerkt wordt met jury's van minstens drie deskundige beoordelaars.

De problemen op het gebied van de variabiliteit van schrijvers, van taken en van beoordelaars zijn het kleinst bij de objectieve toetsing in de vorm van meerkeuzetoetsen. Deze toetsvorm is te karakteriseren als een indirecte meting van de schrijfvaardigheid. Niet het geschreven opstel wordt beoordeeld, maar de leerling beantwoordt meerkeuzevragen over gegeven tekstfragmenten. Het is de bedoeling dat daarbij fouten en tekortkomingen in gegeven teksten gesignaleerd worden. In de Cito-eindtoets basisonderwijs zijn enkele meerkeuzevragen van dit type opgenomen. Tevens

heeft Cito een zogenaamde didactische toets schrijfvaardigheid ontwikkeld die bestaat uit meerkeuzevragen op basis van geschreven teksten door leerlingen (Staphorsius, 1995).

In theorie zijn er veel voordelen aan het gebruik van meerkeuzetoetsen. Zo spelen bij meerkeuzetoetsen de hierboven geschetste verschillen binnen schrijvers en taken geen rol meer en is de beoordelaarsvariabiliteit nihil. De antwoorden op de gecodeerde vragen kunnen simpel gecorrigeerd worden als 'goed' of 'fout'. Ook de praktische bruikbaarheid is erg groot, want de correctie van de antwoorden vraagt weinig tijd. Een ander voordeel is dat de deelvaardigheden beter af te bakenen zijn. De meerkeuzevragen kunnen bijvoorbeeld een beroep doen op de beoordeling van stijl, structuur, publiek- en doelgerichtheid van de tekst. Daarnaast kunnen spelling en grammaticale kwesties aan de orde komen.

Maar er kleven ook bezwaren aan de meerkeuzetoets schrijfvaardigheid. Zo is het bijvoorbeeld nog maar de vraag welke deelvaardigheden precies gemeten moeten worden en in welke verhouding (Van den Bergh & Meuffels, 2000). Een ander bezwaar is gelegen in de vraag in hoeverre de meerkeuzetoets valide is: het lijkt een contradictio in terminis om een toets te gebruiken die een aanspraak doet op de leesvaardigheid om daarmee vervolgens een uitspraak te doen over de vaardigheid in het schrijven. Er zou beredeneerd kunnen worden dat de leesvaardigheid die bij een meerkeuzetoets ingezet moet worden, ook van toepassing is in het schrijfproces. Schrijven is immers een complexe taalvaardigheid waarin ook receptieve deelvaardigheden op een recursieve wijze aan bod komen. De schrijver denkt afwisselend na over het onderwerp, zoekt woorden en construeert zinnen om uiting te geven aan zijn gedachten. Vervolgens herleest hij een voorgaande zin om daarmee aan te sluiten met een nieuwe uiting. Ook bij het redigeren van een concept-

**1. Inhoud & Organisatie**

Op basis van inhoudelijke logica en chronologie moeten leerlingen keuzes maken voor de volgorde van zinnen en tekstfragmenten en voor homonieme of synonieme taaluitingen. Bijvoorbeeld:

- Welke zin past het beste op de stippeltjes?  
Wij hebben thuis een kanarie. .... Soms fluit hij het hoogste lied.  
(a) Wij hebben ook een hond;  
(b) Mijn zusje heeft een blokfluit;  
(c) Hij zit op een stokje in de kamer.  
(d) Maar we hebben geen konijn.

**2. Grammatica**

Hierbij worden vragen gesteld over de grammaticaliteit van woorden en zinnen. Bijvoorbeeld:

- goed / fout: Van de huiden van gedode dieren maken ze bontjassen van.
- Welke van de vier zinnen is juist?  
(a) Ik vind die film niet leuk omdat het zo saai is.  
(b) Ik vind die film niet leuk omdat het daarom saai is.  
(c) Ik vind die film niet leuk omdat het verhaal zo saai is.  
(d) Ik vind die film niet leuk omdat ik het zo saai vind.

**3. Idioom & Stijl**

Hierbij moet de leerling semantische problemen herkennen. De ene keer hebben die te maken met de betekenis van afzonderlijke woorden, de andere keer worden vragen gesteld over uitdrukkingen en zegswijzen. Ook het hanteren van de juiste voorzetsels en pleonasmen komen aan bod. Bij stijl gaat het er vooral om of de woordkeus adequaat is. De meerkeuzevragen gaan over storende herhalingen, onterechte verwisselingen van werkwoordstijden, niet consequent volgehouden perspectief. Bijvoorbeeld:

- Welke uitdrukking past het beste bij de zin:  
'Nauwelijks was de automobilist de overweg gepasseerd, of de trein raasde voorbij.'  
(a) Dat ging zijn neus voorbij;  
(b) Hij is door het oog van de naald gekropen;  
(c) Hij is geen knip voor zijn neus waard;  
(d) Hij zette grote ogen op.

**4. Spelling & Interpunctie**

Bij deze categorie komen zowel de spelling van de veranderlijke als de onveranderlijke woorden aan de orde. Bij interpunctie gaat het om afbreektekens, leestekens en hoofdletters. Bijvoorbeeld:

- goed / fout: De muzikant speelde fluit.
- goed / fout: Hij woont in Noordbrabant.

Figuur 1. Vier vraagcategorieën met voorbeelden van meerkeuzevragen.

tekst leest de schrijver de tekst of delen van een tekst nog eens over en neemt beslissingen om wijzigingen aan te brengen. Maar er moet rekening gehouden worden met de mogelijkheid dat het voor de taalgebruiker gemakkelijker is om fouten in de schriftelijke uiting van een ander te vinden dan deze in de eigen teksten te voorkomen. Productieve schrijfopdrachten en receptieve meerkeuzetoetsen meten waarschijnlijk verschillende kanten van het constructief vaardigheden. De cognitieve vaardigheden die gemeten worden bij de meerkeuzetoets worden soms aangeduid als *low-level skills* (Camp, 2009) in vergelijking met de complexe cognitieve vaardigheden die bij het actieve schrijven gevraagd worden.

#### Onderzoek naar meerkeuzetoetsen schrijfvaardigheid

Enkele onderzoeken rapporteren een substantiële correlatie tussen vrije opdrachtvormen en gesloten meerkeuzetoetsen (Veal & Hudson, 1981; Moss, Cole & Khampalikit, 1982), maar uit andere studies komt geen eenduidig beeld naar voren (Van Schooten, 1988; Camp, 2009). Het valt overigens op dat hier de laatste jaren nauwelijks onderzoek naar verricht is. In Nederland dateren de laatste onderzoeken op dit gebied van ruim 20 jaar geleden (Van Schooten, 1988; Schoonen, 1991).

In het streven naar een valide en betrouwbaar meetinstrument van schrijfvaardigheid heeft Van Schooten (1988) aansluiting gezocht bij de classificatie van Wesdorp (1981). Daarbij is een verdeling gemaakt in vier hoofdcategorieën, die in veel schrijfvaardigheidsmetingen onderscheiden worden. Het betreft: Inhoud & Organisatie, Grammatica, Idioom & Stijl en Spelling & Interpunctie. Van Schooten concludeert dat deze meerkeuzetoets een goede interne validiteit lijkt te hebben. Over de betrouwbaarheid rapporteert hij (1988, p. 106): 'De betrouwbaarheid van de verschillende subtests is redelijk en de toetsende factoranalyses ondersteunen de

divergente validiteit van de in de foutenclassificatie onderscheiden categorieën.' Voor wat betreft de externe validiteit adviseert hij om in het kader van vervolgonderzoek de scores van de meerkeuzetoets te vergelijken met de scores van andere schrijfvaardigheidstoetsen.

Dit alles overziende is het de vraag in hoeverre op grond van een meerkeuzetoets een uitspraak gedaan kan worden over de schrijfvaardigheid van leerlingen in het basisonderwijs. Een meerkeuzetoets zou een mogelijke oplossing kunnen zijn voor de bezwaren die in de dagelijkse praktijk van het onderwijs gevoeld worden om opstellen te beoordelen. De toets is immers gemakkelijk in te zetten en biedt de leerkracht de mogelijkheid om de antwoorden snel na te kijken. Tevens zouden met deze toets uitspraken gedaan kunnen worden over de groei van leerlingen in deelvaardigheden die bij schrijfvaardigheid aan de orde zijn, namelijk Inhoud & Organisatie, Stijl & Idioom, Grammatica, Spelling & Interpunctie. In figuur 1 worden enkele voorbeelden gegeven van meerkeuzevragen.

Op basis van het materiaal dat door Van Schooten is ontwikkeld, stellen wij de volgende onderzoeksvragen:

1. Kan met behulp van een meerkeuzetoets een uitspraak gedaan worden over het constructief schrijfvaardigheid?
2. Is het mogelijk om met behulp van een meerkeuzetoets de ontwikkeling in schrijfvaardigheid in kaart te brengen?

#### Methode

In het schooljaar 2008-2009 is een experiment uitgevoerd waarbij de schrijfontwikkeling van 168 leerlingen van groep acht is gevolgd. Doel van dit longitudinale experiment was om de effecten van het computerprogramma *TiO-schrijven* gedurende één schooljaar te meten. Dit computerprogramma helpt leer-

lingen gedurende het gehele schrijfproces. Er wordt ondersteuning geboden bij het verkennen van het onderwerp, leerlingen kunnen kiezen om bepaalde schrijfstrategieën in te zetten, er worden voorbeeldteksten gegeven en de schrijver wordt gedwongen om voldoende tijd te besteden aan de verschillende deelprocessen van het schrijven. De leerlingen werden verdeeld in drie condities. In de eerste conditie schreven de proefpersonen wekelijks aan de hand van schrijfopdrachten uit de taalmethode, gebruikmakend van de daarin aangeboden schrijfdidactiek. De tweede conditie was de zogenaamde TiO-conditie, waarin leerlingen wekelijks met het computerprogramma schreven. De derde conditie was een controleconditie waarin maar een of twee keer per maand werd geschreven, zonder gebruikmaking van het computerprogramma. Op drie meetmomenten (aan het begin, in het midden en aan het eind van het schooljaar) zijn per leerling in totaal twaalf opstellen verzameld. Vervolgens zijn ruim 2200 opstellen van 168 leerlingen uit de drie condities beoordeeld door een grote groep ervaren leerkrachten en pabostudenten. Aan deze beoordelaars is gevraagd om de te beoordelen tekst te vergelijken met een gemiddelde voorbeeldtekst (die met 100 punten gewaardeerd was). Op basis van enkele criteria gaven beoordelaars vervolgens een waardering boven of onder 100 punten. Met behulp van deze globale schaalbeoordelingen zijn de verschillen tussen condities in kaart gebracht (Pullens, Den Ouden, Herrlitz & Van den Bergh, 2012).

Parallel aan dit onderzoek hebben dezelfde leerlingen een meerkeuzetoets schrijfvaardigheid gemaakt op twee toetsmomenten: aan het begin en aan het einde van groep acht. Hierbij is gebruik gemaakt van een van de twee parallelle subtests zoals ontwikkeld door Van Schooten (1988). Zoals al vermeld, gaat het hier om een relatief oude toets. Er is immers gebleken dat er in de laatste jaren, buiten de

eindtoets basisonderwijs van het Cito, weinig of niets op dit gebied is ontwikkeld dat ook systematisch is onderzocht. Ten behoeve van dit onderzoek zijn de uitkomsten van het hiervoor genoemde longitudinaal onderzoek met globale schaaloordelen te beschouwen als zogenaamde *known-group-resultaten*. De beoordelingen van de opstellen konden op deze manier op twee meetmomenten vergeleken worden met de meerkeuzetoetsen (de opstelbeoordelingen van de tussenmeting zijn in dit onderzoek niet meegewogen, omdat ernaar gestreefd is de metingen van de opstellen en de meerkeuzetoetsen op min of meer hetzelfde tijdstip uit te voeren).

Bij de constructie van de toets is rekening gehouden met de eisen die aan een goede meerkeuzetoets moeten worden gesteld (Mellenbergh, 1971). Allereerst zijn er de eisen van relevantie en evenwichtigheid (logische validiteit). Daarnaast moet de toets een korte en eenduidige instructie bevatten. De antwoordmogelijkheden moeten leiden naar één objectief goed antwoord. Tevens moeten de vragen voldoende discrimineren. De vragen moeten unidimensioneel zijn, dat wil zeggen dat ze bedoeld moeten zijn om één bepaalde vaardigheid te meten. Ten slotte is er de eis dat leerlingen bij de ene vraag geen antwoord moeten kunnen vinden op een andere vraag. Bij de voormeting aan het begin van het schooljaar en bij de nameting aan het eind van het schooljaar is een identieke subtest (de zgn. A-variant) gehanteerd. Gezien het feit dat er minstens tien maanden tussen de voor- en nameting lagen, is het niet waarschijnlijk dat de herinneringseffecten groot waren. De gehele test bestond uit 113 items, verdeeld over de deelvaardigheden Inhoud & Organisatie (12 items = 10,6%) Idioom & Stijl (24 items = 21,2%), Grammatica (29 items = 25,7%) en Spelling & Interpunctie (48 items=42,5%). De toetsen zijn klassikaal afgenomen. Om vermoeidheidseffecten te vermijden is de toets zowel op de voor- als de

nameting opgedeeld in twee afnamemomenten binnen één week. Gezien de tijdsdruk in groep acht, de specifieke eisen van het gehele onderwijsprogramma en de dwingende toetsmomenten, kon uitval van proefpersonen niet geheel vermeden worden. Een groep van tien beoordelaars (pabostudenten) heeft alle antwoorden gescoord. Per deelvaardigheid is het aantal goede en foute antwoorden verzameld.

## Resultaten

We concentreren ons eerst op de interne validiteit van de meerkeuzetoets, anders kunnen de eventueel geconstateerde verschillen

nauwelijks geïnterpreteerd worden. Als we naar de eis van het discriminerend vermogen van de toets kijken, dan zou de groep items met een gemiddelde score van 0,4 – 0,7 het meest vertegenwoordigd moeten zijn. Dat blijkt echter niet het geval te zijn; slechts een kwart (26,6%) van alle items voldoet aan de eis van het discriminerend vermogen. Als we kijken naar de afzonderlijke toetsonderdelen blijkt ook een relatief grote groep items relatief makkelijk te zijn. Idioom & Stijl bevat een relatief groot aantal items met een gemiddelde score die lager is dan 0,4. Vergeleken met de andere toetsonderdelen blijkt dat de meerkeuzevragen van dit type moeilijker zijn dan die van de andere onderdelen. Tweederde

| Toetsonderdeel                          | $\bar{\chi}$ | SD   | $\alpha$ | Aantal items $r_{it} = < 0$ |
|---|--------------|------|----------|-----------------------------|
| <b>Inhoud &amp; Organisatie N=12</b>    |              |      |          |                             |
| voor                                    | 0,37         | 0,07 | 0,56     | 1                           |
| na                                      | 0,37         | 0,07 | 0,58     |                             |
| totaal                                  | 0,37         | 0,07 | 0,57     |                             |
| <b>Idioom &amp; Stijl N=24</b>          |              |      |          |                             |
| voor                                    | 0,66         | 0,14 | 0,59     | 3                           |
| na                                      | 0,68         | 0,15 | 0,67     |                             |
| totaal                                  | 0,66         | 0,14 | 0,63     |                             |
| <b>Grammatica N=29</b>                  |              |      |          |                             |
| voor                                    | 0,61         | 0,10 | 0,61     | 5                           |
| na                                      | 0,62         | 0,10 | 0,62     |                             |
| totaal                                  | 0,61         | 0,10 | 0,62     |                             |
| <b>Spelling &amp; Interpunctie N=48</b> |              |      |          |                             |
| voor                                    | 0,78         | 0,10 | 0,72     | 0                           |
| na                                      | 0,81         | 0,10 | 0,73     |                             |
| totaal                                  | 0,80         | 0,10 | 0,73     |                             |

Tabel 1. Gemiddelde percentuele scores en betrouwbaarheid van de toetsitems per toetsonderdeel (N=aantal items)

van de scores (66,4%) heeft een gemiddelde score die groter is dan 0,7. Dat betekent dus dat deze toets te veel items bevat die niet of nauwelijks discrimineren.

Tabel 1 geeft een overzicht van de gemiddelde scores en de betrouwbaarheid van de scores per toetsonderdeel. De items met een negatieve item-rest correlatie zijn buiten beschouwing gelaten.

Er kan geconcludeerd worden dat de betrouwbaarheid van de toetsitems van de vier verschillende toetsonderdelen niet erg hoog is. Inhoud & Organisatie heeft de laagste gemiddelde betrouwbaarheid, terwijl de toetsitems van Spelling & Interpunctie de hoogst gemeten betrouwbaarheid hebben. Daar waar het mogelijk was om een vergelijking te maken met de betrouwbaarheidsanalyse van Van Schooten zijn er geen opvallende verschillen geconstateerd.

De correlaties tussen de scores van de vier toetsonderdelen zijn vermeld in tabel 2. De correlaties in de diagonaal links onder betreffen zogenaamde geattenuerde correlaties. Dat zijn de geobserveerde correlaties die gecorrigeerd zijn voor de onbetrouwbaarheid

van de meting. Indien er sprake is van lage correlaties, kan geconcludeerd worden dat de items van deze toetsonderdelen verschillende aspecten van de schrijfvaardigheid meten.

Uit tabel 2 blijkt dat de correlaties tussen de vier verschillende toetsonderdelen vrijwel allemaal laag zijn. Echter, één correlatie springt eruit: na de correctie voor de onbetrouwbaarheid van de meting blijkt er een hoge correlatie (0,78) te bestaan tussen de meerkeuzetoetsen van Grammatica en Idioom & Stijl. Uit de lage correlaties kan geconcludeerd worden dat deze toetsonderdelen kennelijk alle een verschillende kant van het construct schrijfvaardigheid meten. Het valt overigens op dat de correlaties die door Van Schooten zijn geconstateerd in enkele gevallen aanmerkelijk hoger zijn.

Parallel aan de meerkeuzetoetsen hebben leerlingen bij de voor- en de nameting opstellen geschreven die met behulp van een globale schaalbeoordeling gewaardeerd zijn door beoordelaars. Per leerling zijn acht opstellen beoordeeld (vier opstellen tijdens de voormeting en vier bij de nameting). Dit heeft per leerling een gemiddelde score opgele-

verd die aangeduid wordt als 'schaalbeoordeling'. Door middel van een regressieanalyse is nagegaan of alle toetsonderdelen op de voormeting van de meerkeuzetoets samen een voorspellende waarde hebben op de hoogte van de schaalbeoordeling op de voormeting. De mate waarin de toetsonderdelen samen de schaalbeoordeling op de voormeting verklaren is niet al te hoog ( $r^2=0,08$ ). Voor de afzonderlijke toetsonderdelen is de voorspellende waarde op de voormeting voor geen van de toetsonderdelen significant: Inhoud & Organisatie ( $t=1,20$ ;  $p=0,23$ ), Idioom & Stijl ( $t=0,48$ ;  $p=0,63$ ), Grammatica ( $t=1,11$ ;  $p=0,27$ ), Spelling & Interpunctie ( $t=1,78$ ;  $p=0,08$ ). De nameting levert hetzelfde beeld op ( $r^2=0,14$ ).

Als er naar de afzonderlijke toetsonderdelen wordt gekeken, kunnen de verschillen tussen de voor- en de nameting in kaart gebracht worden. Bij Inhoud & Organisatie zijn geen van de genoemde verschillen significant. Bij Idioom & Stijl is er wel een significant verschil tussen de voormeting en de nameting ( $F=8,47$ ;  $df 1, 157$ ;  $p=0,004$ ). Ook bij Grammatica is er een significant verschil tussen de voormeting en de nameting ( $F=13,32$ ;  $df 1 151$ ;  $p<0,001$ ). Bij Spelling & Interpunctie is er eveneens een significant verschil tussen de voormeting en de nameting ( $F=15,32$ ;  $df 1, 158$ ;  $p<0,001$ ).

Met behulp van een optimale weging van de onderdelen van de meerkeuzetoets kunnen op de voor- en nameting respectievelijk slechts 8 procent en 14 procent van de verschillen met de schaalscores verklaard worden. Geen van de toetsonderdelen heeft een significant voorspellende waarde op de schaalscore, met uitzondering van het toetsonderdeel Spelling ( $t=1,78$ ;  $p<0,001$ ). Kennelijk bestaat er wel een verband tussen de spellingprestaties van leerlingen en hun prestaties bij het schrijven van opstellen. Van de overige gemeten onderdelen van de meerkeuzetoets kan niet aangetoond worden dat zij gecorreleerd zijn

met echte schrijfprestaties. Daarom kan geen valide uitspraak gedaan worden over het construct schrijfvaardigheid.

## Conclusie en discussie

Om de bruikbaarheid van meerkeuzetoetsen bij het meten van de schrijfvaardigheid te toetsen zijn de scores van de meerkeuzetoetsen vergeleken met beoordelingen van schrijfteksten uit een parallel uitgevoerd longitudinaal onderzoek.

De gebruikte meerkeuzetoets is verdeeld in vier categorieën met vragen over Inhoud & Organisatie, Idioom & Stijl, Grammatica en Spelling & Interpunctie. Hoewel de onderscheiden deelvaardigheden vaak gehanteerd worden bij de beoordeling van opstellen, kan niet worden aangetoond dat de toetsonderdelen een significant voorspellende waarde op de globale opstelbeoordeling hebben. Voor spelling kan wel een positief verband aangetoond worden. Er zou beredeneerd kunnen worden dat spellingkennis deel uitmaakt van het construct schrijfvaardigheid.

De vier verschillende toetsonderdelen hebben een lage onderlinge correlatie. Hieruit kan geconcludeerd worden dat deze toetsonderdelen afzonderlijk een verschillende kant van het construct schrijfvaardigheid meten. Gezien de lage correlatie met de globale schaalscores van opstellen van dezelfde leerlingen, kan niet geconcludeerd worden dat deze toetsonderdelen samen een uitspraak kunnen doen over het construct schrijfvaardigheid. Dit is overigens wel begrijpelijk als we in ogenschouw nemen dat schrijfvaardigheid een productieve taalvaardigheid is en dat een meerkeuzetoets een grote aanspraak doet op de receptieve leesvaardigheid.

De meerkeuzetoets laat wel significante verschillen zien tussen de voor- en nameting. Kennelijk is het mogelijk om de groei in bepaalde taalvaardigheden (zoals spelling en

| Toetsonderdeel          | Inhoud & organisatie | Idioom & Stijl | Grammatica  | Spelling & Interpunctie |
|-------------------------|----------------------|----------------|-------------|-------------------------|
| Inhoud & Organisatie    |                      | 0,26           | 0,31        | 0,20                    |
| Idioom & Stijl          | 0,43 (1,00)          |                | 0,49        | 0,37                    |
| Grammatica              | 0,52 (0,47)          | 0,78 (0,85)    |             | 0,36                    |
| Spelling & Interpunctie | 0,31 (0,60)          | 0,55 (0,70)    | 0,54 (0,50) |                         |

Tabel 2. Correlaties tussen de verschillende toetsonderdelen op de voor- en nameting; geobserveerde correlatie boven diagonaal; geattenuerde correlaties onder diagonaal. Alle correlaties zijn significant (de door Van Schooten [1988] geconstateerde correlaties tussen haakjes)

grammaticale kennis) bij individuele leerlingen vast te stellen met dit instrument. Maar omdat de correlatie met taken, waarbij echt geschreven moet worden laag is, blijft het onduidelijk wat deze meerkeuzetoets precies meet. Gezien de kwaliteit van de items en de interne validiteit van de toets als geheel zijn we geneigd dit te generaliseren naar een lage constructiviteit van meerkeuzetoetsen in het algemeen. Er moeten daarom vraagtekens gezet worden bij de ambitie van Staphorsius (1995) om een vergelijkbare meerkeuzetoets in te zetten teneinde een leergesprek over schrijfvaardigheid op gang te brengen. Tevens moeten twijfels geuit worden bij de vaststelling in de handleiding van de Cito-toets dat leerlingen die goede teksten schrijven, over het algemeen ook goed presteren bij de Didactische Toets Schrijfvaardigheid (Staphorsius, 1995, p. 11). In zijn algemeenheid kunnen er vraagtekens gezet worden bij het gegeven dat meerkeuzevragen schrijfvaardigheid onderdeel uitmaken van de eindtoets basisonderwijs, en zou geadviseerd moeten worden om dit type toetsing niet langer op te nemen in deze eindtoets en de nog te ontwikkelen diagnostische tussentijdse toets in het voortgezet onderwijs.

## LITERATUUR

- Bergh, H. van den & Meuffels, B. (2000). Schrijfvaardigheden en schrijfprocessen. In: A. Braet (red.) *Taalbeheersing als communicatiewetenschap. Een overzicht van theorievorming, onderzoek en toepassingen.* (pp. 122–153). Bussum: Coutinho.
- Braddock, R. Lloyd-Jones, R. & Schoer, L. (1963). *Research in written composition.* Urbana, IL: National Council of Teachers of English.
- Camp, R. (2009). Changing the Model for the Direct Assessment of Writing. In: B. Huot & P. O'Neill (Eds.). *Assessing Writing. A Critical Sourcebook.* Urbana: Bedford / St. Martin's.
- Glopper, K. de (1988). Opstelkenmerken en opstelbeoordelingen. *Tijdschrift voor Taalbeheersing*, 6, 176–189.
- Henkens, L. (2010). *Het onderwijs in het schrijven van teksten. De kwaliteit van schrijfonderwijs in het basisonderwijs.* Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2012). *Focus op schrijven. Het onderwijs in het schrijven van teksten (stellen).* Utrecht: Inspectie van het Onderwijs.
- Mellenberg, G. (1971). *Studies in studietoetsen.* Amsterdam: Psychologisch Laboratorium, Universiteit Amsterdam. (Academisch proefschrift).
- Moss, P., Cole, N. & Khampalakit, C. (1982). A comparison of procedures to assess written language skills at grades 4, 7 and 10. *Journal of Educational Measurement*, 19, 37–48.
- Pullens, T., Ouden, H. den, Herrlitz, W., & Bergh, H. van den (2012). Effecten van het computerprogramma TiO-schrijven op de schrijfvaardigheid in leerlingen in groep acht. Een longitudinaal onderzoek. *Pedagogische Studiën*, 89(2), 104–119.
- Schoonen, J. (1991). *De evaluatie van schrijfvaardigheidsmetingen: Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs.* Haarlem, SCO. (Academisch proefschrift).
- Schooten, E. van (1988). *De constructie van een meerkeuzetoets voor het meten van schrijfvaardigheid.* Amsterdam: SCO, Stichting Kohnstamm Fonds voor onderwijsresearch (SCO-rapport 178).
- Staphorsius, G. (1995). *Didactische Toetsen Schrijfvaardigheid. Handleiding.* Arnhem: Cito.
- Stokking, K. (2000). *Ontwikkeling en beoordeling van schrijfvaardigheden. Aanwijzingen en suggesties voor docenten Nederlands rond het werken met het 'schrijfdossier' in de Tweede Fase.* Utrecht: ICO/ISOR.

- Veal, L. & Hudson, S. (1981). Direct and indirect measurements for large-scale evaluation of writing. *Research in the Teaching of English*, 17(3), 290–296.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid.* SVO-reeks 42, 's-Gravenhage: Staatsuitgeverij.

THEO PULLENS is onlangs gepromoveerd aan de Universiteit Utrecht en hogeschooldocent Nederlands bij Hogeschool Avans, Pabo Breda. E-mail: <tjm.pullens@avans.nl>

HANNY DEN OUDEN is universitair docent en onderzoeker op de afdeling Taalbeheersing, departement Nederlandse Taal en Cultuur van de Universiteit Utrecht.

WOLFGANG HERRLITZ is emeritus hoogleraar linguïstiek, in het bijzonder taalonderwijskunde en taalkunde van het Duits bij de Faculteit Geesteswetenschappen van de Universiteit Utrecht.

HUUB VAN DEN BERGH is hoogleraar op de afdeling Taalbeheersing, departement Nederlandse Taal en Cultuur en het onderzoeksinstituut OTS van de Universiteit Utrecht.