



Kritiek op toetsen Spelling steunt op losse gronden

Een reactie op het artikel 'De nieuwe Cito-spellingtoets'

A. de Wijs

SAMENVATTING

In het artikel 'De nieuwe Cito-spellingtoets' wordt een onderzoek van Schraven, Bosman en Van Eekhout beschreven dat tot doel had te bepalen of het gerechtvaardigd is om voor het vaststellen van de spellingvaardigheid van leerlingen dicteeopgaven te vervangen door meerkeuzeopgaven. Deze onderzoeksvraag is relevant omdat Cito in de nieuwe toetsen Spelling niet alleen dictees aanbiedt, maar ook meerkeuzeopgaven. Na bestudering van het artikel zetten wij echter vraagtekens bij de opzet van het onderzoek. Op basis van een toetsafname bij slechts 18 leerlingen van één school kun je naar onze stellige mening geen uitspraken doen over de psychometrische eigenschappen (betrouwbaarheid en interne consistentie) van de toetsen Spelling, of van welke toets dan ook. Cito heeft in 2008 en 2009 een vergelijkend onderzoek dictee-meerkeuze uitgevoerd bij 1318 leerlingen in groep 4 en 1928 leerlingen in groep 8. Uit dat onderzoek blijkt dat dicteeopgaven en meerkeuzeopgaven niet precies hetzelfde meten (dat beweren we ook nergens), maar wel redelijk hoog correleren (.80 in groep 4 en .86 in groep 8). Veel hoger dan de correlaties die Schraven, Bosman en Van Eekhout vonden.

In de opinie van Cito kent spelling meer aspecten dan het zelf foutloos schrijven van woorden. Ook het kunnen herkennen en verbeteren van spelfouten behoort tot de vaardigheid spellen. Als leerlingen deze 'passieve' kant van spelling niet beheersen, heeft het bijvoorbeeld weinig zin hen te vragen een opstel nog eens na te kijken alvorens het in te leveren. Zowel de actieve kant van spelling (zelf schrijven) als de passieve kant (opsporen van spelfouten) zou in het onderwijs aan bod moeten komen. En zoals het dictee een passende vorm van toetsing is voor het zelf foutloos schrijven, kan het herkennen van spelfouten prima getoetst worden met meerkeuzeopgaven.

De toetsen Spelling

De nieuwe toetsen Spelling van het Leerling- en onderwijsvolgsysteem (LOVS) van Cito vervangen sinds 2006 geleidelijk de oude toetsen Schaal Vorderingen in Spellingvaardigheid (SVS). Naast dicteeopgaven bevatten de nieuwe toetsen ook meerkeuzeopgaven. Die zijn niet bestemd voor alle groepen (in groep 3 krijgen de leerlingen alleen dictee) noch voor alle leerlingen (in groep 4 en 5 krijgen de zwakkere spellers alleen dictee), maar Schraven, Bosman en Van Eekhout constateren terecht dat het opnemen van meerkeuzeopgaven een breuk betekent met de oude SVS-toetsen. De breuk is echter oppervlakkiger dan zij denken: de vaardigheidsschaal die ten grondslag ligt aan de nieuwe toetsen Spelling is namelijk dezelfde als de schaal die ten grondslag lag aan de SVS-toetsen. Op deze vaardigheidsschaal zijn (na analyse met OPLM¹) zowel de dicteeopgaven als de meerkeuzeopgaven afgebeeld. Dat betekent niet dat dicteeopgaven identiek zouden zijn aan meerkeuzeopgaven. Het betekent dat de opgaven in een meerkeuzetaak een beroep doen op dezelfde kennis en vaardigheden als waarop bij het maken van een dictee een beroep wordt gedaan. Met andere woorden: de onderliggende vaardigheid is dezelfde; niet de opgaven waarmee je die vaardigheid wil meten.

¹ One Parameter Logistic Model: meetmodel uit de testleer waarin de antwoorden die een leerling op de opgaven geeft, worden beschouwd als indicator voor de mate waarin de te meten vaardigheid aanwezig is.

Meerkeuze als toetsvorm voor spellingvaardigheid

Opschrijven of aanstrepen

Als dicteeopgaven precies hetzelfde zouden meten als meerkeuzeopgaven, maakt het niets uit welke vorm je kiest. Je kunt dan alle dicteeopgaven vervangen door meerkeuzeopgaven, of andersom, zonder het toetsresultaat te beïnvloeden. Dat klopt natuurlijk niet, en Cito beweert dat ook niet. Het zelf schrijven van een woord is niet identiek aan het herkennen van een foute schrijfwijze van dat woord. Beide deelvvaardigheden hangen wel samen: als je een woord correct kunt schrijven, kost het je waarschijnlijk weinig moeite om een spelfout in dat woord te ontdekken. En soms weet je niet precies hoe je een woord schrijft, maar wel dat het zó - op deze manier geschreven - niet correct is. Een concreet voorbeeld daarvan is dat je bij twijfel over de spelling van een woord, het woord op verschillende manieren opschrijft op een kladblaadje, om dan via eliminatie de juiste schrijfwijze te kiezen. Zelf schrijven ('actief spellen') en herkennen ('passief spellen') gaan dan vrijwel naadloos in elkaar over. Actief en passief spellen zijn echter niet identiek. Leerlingen halen niet precies dezelfde scores op dictees en op meerkeuzeopgaven, ook niet als de getoetste woorden hetzelfde zijn.

Presenteren van fout gespelde woorden

Schraven, Bosman en Van Eekhout voeren als argument tegen meerkeuzeopgaven spelling aan dat het regelmatig aanbieden van fout gespelde woorden de spellingvaardigheid van leerlingen negatief kan beïnvloeden. Bij dat argument zijn wel wat kanttekeningen te maken. De toetsen Spelling worden twee keer per jaar afgenomen. In totaal krijgt een gemiddeld vaardige leerling via de toetsen dus maximaal twee keer per jaar gedurende een half uur een aantal zinnen onder ogen waarin een fout gespeld woord staat. Dát er een fout in de zinnen staat, wordt de leerling expliciet verteld. De vraag is immers: 'In welke zin is het dikgedrukte woord fout gespeld?' Het gaat hier om een geïsoleerde toetservaring, waarbij het de leerling duidelijk is dat de aangeboden woorden niet geleerd hoeven te worden. De kans dat woorden tijdens de toetsafname foutief worden ingeprent door leerlingen die de juiste schrijfwijze van het woord blijkbaar eerder niet hadden kunnen onthouden lijkt ons dan ook klein. (Ging het leren maar zo makkelijk!)

Diagnostische mogelijkheden

Volgens Schraven, Bosman en Van Eekhout bieden meerkeuzeopgaven weinig mogelijkheden tot diagnostiek. Zij hebben daarin gelijk. Het door hen aangehaalde voorbeeld van de leerling die denkt dat 'zomer' fout geschreven is, snijdt hout. Waarom vindt deze leerling 'zomer' fout? Had het volgens de leerling 'somer' moeten zijn, of 'zomur', of 'zoomer'? Met een meerkeuzeopgave kom je daar niet achter.

Toch kennen ook de meerkeuzeopgaven diagnostische mogelijkheden, via een omgekeerde bewijsvoering. Als de leerling een meerkeuzeopgave fout beantwoordt, heeft hij of zij gemeend een fout te zien die er in werkelijkheid niet was (in het bovenstaande voorbeeld: een fout in 'zomer'). Welke fout de leerling dacht te zien, is niet bekend. Maar in de opgave stond ook een echte fout, die de leerling niet gezien heeft. De echte fout valt in een bepaalde spellingcategorie. Blijkbaar beheerst de leerling die categorie nog niet volledig, anders had hij of zij de fout wel ontdekt. Als een leerling veel meerkeuzeopgaven van een bepaalde categorie fout heeft, is dat dus wel degelijk een aanwijzing dat de leerling die categorie onvoldoende beheerst. Op deze manier is het mogelijk om ook bij meerkeuzeopgaven een categorieënanalyse te maken.

Het onderzoek van Schraven, Bosman en Van Eekhout

Schraven, Bosman en Van Eekhout willen nagaan of het gerechtvaardigd is om voor het bepalen van de spellingvaardigheid het dictee te vervangen door een meerkeuzetoets. Ook zonder onderzoek is die vraag eenvoudig te beantwoorden als je - zoals Schraven, Bosman en Van Eekhout doen - stelt dat spelling is: het foutloos schrijven van woorden. Dat kun je inderdaad niet goed meten met

meerkeuzeopgaven. Wij zouden spelling echter willen omschrijven als: weten wat de juiste schrijfwijze van een woord is. Spellingvaardigheid heeft dan, zoals hierboven al gezegd werd, twee aspecten: zelf foutloos schrijven (de 'actieve' kant van spelling) en fout gespelde woorden herkennen ('passief' spellen). In de toetsen Spelling van Cito wordt het eerste aspect getoetst met dicteeopgaven, het tweede met meerkeuzeopgaven.

Opzet van het onderzoek

De onderzoekspopulatie van Schraven, Bosman en Van Eekhout bestaat uit 18 leerlingen uit één groep 4 van een basisschool in Zevenaar (Gelderland). Eerst krijgen de leerlingen de toetsmodule M4 Start uit het toetspakket LOVS Spelling groep 4 van Cito. Deze module bestaat uit 25 dicteezinnen.

Een voorbeeld van zo'n dicteezin is:

De ridder ging op zijn **paard** zitten. Schrijf op: **paard**.

Daarna krijgen ze toetsmodule M4 Vervolg 2 voorgelegd. Deze module bevat 25 meerkeuzeopgaven.

De meerkeuzeopgaven Spelling bestaan uit vier (niet samenhangende) zinnen A, B, C en D. In elke zin is één woord dikgedrukt. Een van de vier dikgedrukte woorden is fout gespeld. De vraag is telkens: In welke zin is het dikgedrukte woord fout gespeld?

Een voorbeeld van zo'n meerkeuzeopgave is:

A Aafke sprong in de **sloot**.

B Aan de **balk** hing een haak.

C Hij viel van het **flot** in het water.

D Snij het **uitje** in stukjes.

Enige tijd na de afname van deze twee modules krijgen de leerlingen op vier achtereenvolgende dagen alle dikgedrukte woorden uit module M4 Vervolg 2 nogmaals aangeboden, maar nu in de vorm van een dictee. De eerste dag worden de 25 dikgedrukte woorden uit de A-zinnen voorgelezen, de tweede dag de 25 dikgedrukte woorden uit de B-zinnen, enzovoort. In vier dagen tijd schrijven de leerlingen dus alle 100 dikgedrukte woorden uit de meerkeuzemodule op. Diezelfde woorden hebben ze al een keer in een toetsboekje zien staan. Toen was een vierde deel van die woorden fout gespeld; de leerlingen weten echter niet welke woorden (ervan uitgaande dat de meerkeuzemodule na afloop niet met hen besproken is). De kans is klein dat de leerlingen de betreffende woorden hebben onthouden. Ze hebben ze immers maar één keer eerder onder ogen gehad, in een toetssituatie die niet deed vermoeden dat ze later die woorden zelf moesten schrijven. Waarschijnlijk hebben ze de dikgedrukte woorden niet aandachtiger bekeken dan willekeurig welke leerling die een meerkeuzemodule maakt.

We verwachten dus niet dat het zien van de dikgedrukte woorden de resultaten van het woorddictee heeft beïnvloed, maar helemaal zeker is dat natuurlijk niet.

Een belangrijker probleem van de gekozen onderzoeksopzet is de grootte en samenstelling van de onderzoeksgroep. Een populatie van 18 leerlingen uit één groep is zowel te klein als te eenvormig om uitspraken te kunnen doen over een bij die groep afgenomen toets. Om aan te tonen dat een toets betrouwbaar is, of juist niet, zijn vele honderden proefpersonen nodig, afkomstig uit een steekproef waarin de totale populatie waarvoor die toets bestemd is, wordt vertegenwoordigd. Dus leerlingen uit een grote stad en leerlingen uit een dorp, witte en zwarte leerlingen, jongens en meisjes, zittenblijvers en bollebozen, leerlingen van openbare, christelijke en islamitische scholen, leerlingen die les krijgen volgens methode A en leerlingen die les krijgen volgens methode B, leerlingen met hoogopgeleide en laagopgeleide ouders, leerlingen die thuis een dialect spreken, leerlingen die tweetalig opgroeien, ... Kortom, leerlingen die een representatieve afspiegeling vormen van de totale Nederlandse populatie van groep 4-leerlingen.

Schraven, Bosman en Van Eekhout voeren diverse statistische bewerkingen uit op de door hen verzamelde onderzoeksresultaten. Zij berekenen waarden voor de betrouwbaarheid van de toetsen, stellen correlaties vast tussen dicteeopgaven en meerkeuzeopgaven en bepalen itemkarakteristieken tot op twee cijfers achter de komma, dit alles op basis van één toetsafname bij één groep van 18 leerlingen. Een dergelijk onderzoek kan naar onze mening de toets der kritiek niet doorstaan.

Resultaten

Betrouwbaarheid

Schraven, Bosman en Van Eekhout hebben de betrouwbaarheid van module Spelling M4 Vervolg 2 bepaald en komen daarbij op een alfawaarde van .75. Dat is te laag om te kunnen spreken van een betrouwbare toets, zeggen zij. Maar module M4 Vervolg 2 is geen toets, het is de helft van een toets. M4 Vervolg 2 bevat slechts 25 opgaven en dat is erg weinig om een betrouwbaar beeld te kunnen geven van iemands spellingvaardigheid. De volledige toets Spelling M4 bevat 50 opgaven: eerst maken de leerlingen M4 Start (25 opgaven dictee) en daarna ofwel M4 Vervolg 1 (25 opgaven dictee) ofwel M4 Vervolg 2 (25 opgaven meerkeuze). Van de aldus opgebouwde toets Spelling M4 heeft Cito de betrouwbaarheidscoëfficiënten vastgesteld op .90 (M4 Start + Vervolg 1) en .91 (M4 Start + Vervolg 2) in een normeringsonderzoek bij 782 leerlingen van 55 verschillende scholen. Overigens bedraagt de door Schraven, Bosman en Van Eekhout berekende betrouwbaarheid van (alleen) de dicteemodule M4 Start slechts .66, lager dus dan de .75 van de meerkeuzemodule. Aan dit verschil besteden zij in hun artikel nauwelijks aandacht. Hun voorkeur voor dictee is dus blijkbaar niet gebaseerd op het belang dat zij hechten aan betrouwbaarheid.

Naast de betrouwbaarheid van de modules M4 Vervolg 2 en M4 Start berekenen Schraven, Bosman en Van Eekhout ook de betrouwbaarheden van de door henzelf gemaakte dictees met A-zinnen, B-zinnen etc. De gevonden waarden zijn lager dan die van de bestaande modules. Het is ons niet duidelijk wat zij hiermee willen aantonen.

Correlatie tussen meerkeuze en dictee

Schraven, Bosman en Van Eekhout vinden in hun onderzoek een lage correlatie (.45) tussen de meerkeuzemodule en de verschillende dicteemodules (de bestaande module M4 Start en de zelf gemaakte dictees van de dikgedrukte woorden uit de meerkeuzemodule). Waarschijnlijk hangt dat samen met hun kleine onderzoekspopulatie. Cito heeft in 2008 onder 1318 leerlingen uit groep 4 een onderzoek uitgevoerd, waarin de fout gespelde woorden uit module M4 Vervolg 2 als dicteewoord aan de leerlingen werden voorgelegd (net als bij Schraven, Bosman en Van Eekhout). Maar in het onderzoek van Cito kregen de leerlingen nooit twee keer hetzelfde woord voorgelegd. De kans dat leerlingen woorden herkennen of onthouden is dus gegarandeerd nihil. Dit werd gerealiseerd volgens het design (onderzoeksontwerp) dat hieronder is weergegeven.

groep	D1, opg 1-12	D2, opg 13-25	M1, opg 26-37	M2, opg 38-50	aantal leerlingen
1					323
2				292	
3					350
4				353	

Eerst werd van de meerkeuzemodule een dicteeversie gemaakt, die alleen de zinnen bevatte waarin het fout gespelde woord stond. Vervolgens werden deze dicteeversie (25 zinnen) en de meerkeuzemodule zelf (25 opgaven) in tweeën geknipt, waardoor vier deelttoetsjes ontstonden: twee met dicteezinnen (D1 en D2) en twee met meerkeuzeopgaven (M1 en M2).

De totale groep leerlingen werd in vier groepen verdeeld. De eerste groep maakte eerst D1 (12 dicteeopgaven) en daarna D2 (13 dicteeopgaven). De tweede groep maakte eerst D1 en daarna M2 (13 meerkeuzeopgaven), de derde groep maakte eerst D2 en daarna M1 (12 meerkeuzeopgaven) en de vierde groep maakte eerst M1 en daarna M2.

Op basis van de resultaten van de leerlingen werd de latente correlatie tussen de dicteeopgaven en de meerkeuzeopgaven bepaald; deze was exact .80.

Cito heeft dit onderzoek in 2009 op dezelfde wijze herhaald in groep 8 bij in totaal 1928 leerlingen, met module M8 Vervolg 2 en een daarvan afgeleide dicteeversie. De latente correlatie was .865, hoger dus dan in groep 4. Een mogelijke verklaring voor dit verschil is dat in groep 4 (en in groep 5) de meerkeuzeopgaven eigenlijk niet bedoeld zijn voor alle leerlingen, maar alleen voor de betere spellers. De leerlingen in het onderzoek van Schraven, Bosman en Van Eekhout hoorden volgens de auteurs allemaal tot de betere spellers. (Hun onderzoeksgroep is erg homogeen.) In het onderzoek van Cito maakten ook de minder goede spellers in groep 4 de meerkeuzeopgaven.

Moeilijkheid van de opgaven

In het onderzoek van Cito in groep 4 zijn de scores van de leerlingen omgerekend naar een (feitelijk niet bestaand) aantal van 50 dicteeopgaven en 50 meerkeuzeopgaven. Gemiddeld hadden de leerlingen 39,9 dicteeopgaven goed beantwoord, tegen 40,2 meerkeuzeopgaven. Het verschil in moeilijkheid tussen dictee en meerkeuze was in dit onderzoek dus verwaarloosbaar.

Schraven, Bosman en Van Eekhout vinden wel verschillen in moeilijkheidsgraad tussen dictee en meerkeuze. Gemiddeld maken hun leerlingen 19,7 van de 25 meerkeuzeopgaven goed en 21,4 van de 25 van die meerkeuzeopgaven afgeleide dicteewoorden. Zij noemen dit een interessante bevinding, te meer omdat zij via statistische toetsing middels t-waarden kunnen laten zien dat de geconstateerde verschillen significant zijn. Echter de door hen gevonden resultaten zijn afkomstig van slechts 18 leerlingen, een te klein aantal om überhaupt t-waarden over te kunnen berekenen.

Schraven, Bosman en Van Eekhout noemen als probleem dat de scores van hun leerlingen op een inconsistente wijze variëren tussen dictee- en meerkeuzeopgaven. Voor elf leerlingen was de meerkeuzemodule moeilijker dan het dictee, voor vier leerlingen was het net andersom, en voor drie leerlingen maakte het niet uit. Wat had hier consistent moeten zijn? Hadden alle leerlingen lager moeten scoren op de meerkeuzemodule dan op het dictee? Dat zou vreemd geweest zijn, want uit het vergelijkend onderzoek van Cito blijkt dat dictee en meerkeuze nauwelijks in moeilijkheid verschillen. Hadden de twee groepen leerlingen dan precies even groot moeten zijn (negen leerlingen maken dictee beter, negen leerlingen meerkeuze)? Dat zou een toevalstreffer geweest zijn met een dergelijk klein aantal leerlingen.

Metten van spellingvaardigheid

Schraven, Bosman en Van Eekhout besteden veel aandacht aan het feit dat leerlingen in hun onderzoek in ongeveer een kwart van de gevallen het dicteewoord wel goed schrijven, en de meerkeuzeopgave van dezelfde woord fout beantwoorden, of omgekeerd. (In driekwart van de gevallen komt de score op dictee en meerkeuze overigens wél overeen.) Zij concluderen daaruit dat meerkeuzeopgaven niet hetzelfde meten als dicteeopgaven en zij vragen zich af hoe je het spellingniveau van een leerling kunt bepalen met twee verschillende soorten opgaven. Volgens ons kan dat wel degelijk. Dicteeopgaven zijn geschikt voor het meten van de actieve kant van spellingvaardigheid (zelf schrijven); meerkeuzeopgaven voor het meten van de passieve kant (spelfouten opsporen). De beide soorten opgaven leiden niet tot precies dezelfde resultaten; ze correleren echter redelijk hoog en liggen op dezelfde vaardigheidsschaal.

Overigens zijn bijna alle schoolse vaardigheden te meten met verschillende soorten opgaven. Als je wil nagaan of een leerling de Europese hoofdsteden kent, kun je die laten aanwijzen, opnoemen, aankruisen, intikken, een niet bestaande hoofdstad (bijvoorbeeld 'Rotterdam') laten opsporen in een lijst, etc. En als je wil nagaan of een leerling weet hoe 'hoofdstad' geschreven moet worden, kun je dat woord laten opschrijven, aanvinken, letter voor letter hardop zeggen, een foute schrijfwijze (bijvoorbeeld 'hoofstad') laten opsporen in een meerkeuzeopgave, etc.

De invloed van de context in meerkeuzeopgaven

Een dicteeopgave over het woord 'zomer' gaat over het woord 'zomer', en nergens anders over. Een meerkeuzeopgave over het woord 'zomer' gaat ook over de andere drie dikgedrukte woorden. Het maakt natuurlijk wel wat uit of 'zomer' vergezeld gaat van de dikgedrukte woorden 'sequentieel', 'fragmentarisch' en 'conciërge' (opgave x), of van de dikgedrukte woorden 'appeltaart', 'hondje' en 'speeltuin' (opgave y). Stel dat 'zomer' zowel in opgave x als in opgave y het fout gespelde woord is ('somer'). De kans is groot dat er bij opgave x meer leerlingen zijn die denken dat een van de andere woorden fout geschreven is dan bij opgave y. Het is dus niet alleen zo dat 'zomer' in een dicteeopgave soms een andere score oplevert dan in een meerkeuzeopgave; waarschijnlijk levert 'somer' in de ene meerkeuzeopgave een andere score op dan in de andere meerkeuzeopgave.

Dat is minder erg dan het op het eerste gezicht lijkt. De spellingvaardigheid van leerlingen wordt natuurlijk niet vastgesteld op basis van één meerkeuzeopgave. En om even bij het bovenstaande voorbeeld te blijven: in een toets zit maar één opgave met 'somer' (ofwel opgave x, ofwel opgave y) en die opgave is voor alle leerlingen gelijk. Het is dus niet zo dat de ene leerling een score krijgt voor het herkennen van 'somer' in opgave x en een andere leerling voor het herkennen van 'somer' in opgave y.

Maar de context van de andere woorden in een meerkeuzeopgave heeft wel degelijk invloed op de moeilijkheid van de opgave, en Schraven, Bosman en Van Eekhout vestigen daar terecht de aandacht

op. Zij vinden een dergelijke invloed zonder meer ongewenst. In hun visie worden leerlingen bij het opsporen van een spelfout op het verkeerde been gezet als rondom die spelfout woorden staan die ze niet goed kennen. Dat beeld willen wij graag nuanceren. In teksten staan wel vaker woorden die leerlingen niet kennen en met het oog op hun taalontwikkeling lijkt ons dat heel goed. Nu is een toets natuurlijk niet bedoeld om taal te ontwikkelen, maar om de al ontwikkelde taal te meten. Toetsen en opgaven die alleen woorden bevatten die alle leerlingen kennen, zijn voor die meting echter niet geschikt. Hoe kan de goede speller zich dan van de minder goede speller onderscheiden? Hoe kan de leerkracht te weten komen welke leerlingen extra hulp nodig hebben? Om betrouwbaar te kunnen meten, moet een toets zo goed mogelijk aansluiten bij de vaardigheid van de gehele doelgroep. Een goede toets bevat dus zowel uitdagende opgaven voor de goede spellers als makkelijke opgaven voor de minder goede spellers.

Cito hanteert strenge criteria voor de meerkeuzeopgaven in de toetsen Spelling. Elke opgave moet door minimaal 40% (maar niet meer dan 90%) van de leerlingen goed beantwoord zijn in een grootschalig normeringsonderzoek. En leerlingen die goed kunnen spellen moeten een grote kans hebben de betreffende opgave goed te maken (de opgave mag de goede spellers niet aan het twijfelen brengen).

Schraven, Bosman en Van Eekhout bespreken opgave 1 uit module M4 Vervolg 2, een meerkeuzeopgave met de dikgedrukte woorden 'vlag', 'wrak', 'grap' en 'swak'. Zij geven het voorbeeld van een leerling die het woord 'wrak' niet kent, denkt dat het fout geschreven is en dat daarom 'swak' goed gespeld moet zijn. Wordt deze leerling misleid? Hij kent 'wrak' niet, en hij kent 'zwak' niet goed genoeg. 'Wrak' is een heel moeilijk woord. Alle leerlingen die M4 Vervolg 2 maken, lopen tegen deze moeilijkheid aan. Leerlingen worden ten opzichte van elkaar dus niet benadeeld. De leerlingen die deze opgave goed beantwoorden (in het normeringsonderzoek: 43% van de leerlingen) weten misschien niet zeker hoe je 'wrak' schrijft, maar ze weten wel dat 'swak' fout is.

Conclusies

Schraven, Bosman en Van Eekhout concluderen uit hun onderzoek dat een meerkeuzetoets geen adequate vervanging is van een dictee, omdat de meerkeuzetoets iets anders meet. Die conclusie vinden wij om diverse redenen te kort door de bocht. In de toetsen Spelling is geen sprake van 'vervanging': alle leerlingen maken altijd minstens één dictee. Daarnaast maken de leerlingen in veel gevallen een meerkeuzemodule. Het is dus niet het een of het ander, maar het een én het ander. Meerkeuzeopgaven en dicteeopgaven kunnen beide gebruikt worden om spellingvaardigheid te meten. Met het dictee meet je of leerlingen zelf foutloos woorden kunnen schrijven (actieve spelling), met de meerkeuzeopgaven of leerlingen spelfouten kunnen herkennen (passieve spelling). De betrouwbaarheid van de aldus samengestelde toetsen Spelling is hoog (.90), veel hoger dan de door Schraven, Bosman en Van Eekhout berekende betrouwbaarheden. Maar die laatste waren dan ook gebaseerd op de afname van een halve toets, bij slechts 18 leerlingen.

Schraven, Bosman en Van Eekhout voorzien dat een leerkracht op basis van de door hen onderzochte meerkeuzeopgaven foute beslissingen zal nemen over of een leerling extra hulp nodig heeft, en zo ja, voor welke spellingcategorieën. Die vrees is niet ongegrond, maar om een andere reden dan de opgavenvorm. Zoals eerder al aangegeven werd, is het ook bij meerkeuzeopgaven mogelijk om - in beperkte mate, via omgekeerde bewijsvoering - uitspraken te doen over het wel of niet beheersen van een spellingcategorie. Maar op basis van één module met 25 opgaven, waarin wel dertien (!) verschillende spellingcategorieën aan de orde komen, is het nauwelijks mogelijk een uitspraak te doen over of een leerling een categorie wel of niet beheerst, ook niet als die module uit dicteezinnen zou bestaan. Gemiddeld krijgt de leerling immers slechts twee opgaven van elke categorie aangeboden. Stel dat hij of zij beide opgaven fout beantwoordt (of: beide woorden fout schrijft), is daarmee dan vastgesteld dat de leerling de betreffende categorie niet beheerst? De toets Spelling M4 bestaat uit twee modules en dus uit 50 in plaats van 25 opgaven, maar ook op basis van de hele toets blijft voorzichtigheid geboden bij het doen van een dergelijke uitspraak. Als een leerkracht vermoedt dat een leerling een categorie niet beheerst, op basis van de toetsscore en de categorieënanalyse, adviseert Cito om eerst het controledictee uit het hulpmiddel Spelling af te nemen alvorens gericht hulp te bieden. Een controledictee bevat tien woorden uit één spellingcategorie. Als de leerling de meeste van die woorden fout schrijft, is de kans groot dat er inderdaad sprake is van onvoldoende beheersing van de categorie. Dan kan zinvol hulp worden

geboden. De controledictees uit het hulpboek Spelling kunnen in geval van twijfel ook worden gebruikt als de toetsuitslag en de categorieënanalyse niet direct verontrustend waren. Cito raadt in elk geval af om op basis van alleen het toetsresultaat over te gaan tot categoriegerichte remediëring.

In de laatste alinea's van hun artikel spreken Schraven, Bosman en Van Eekhout zich ondubbelzinnig uit vóór het dictee en tegen meerkeuzeopgaven. Helaas is hun onderzoek uitgevoerd bij een veel te kleine groep leerlingen, waardoor hun argumentatie op losse gronden is gebaseerd. Toch menen de auteurs voldoende 'bewijsmateriaal' te hebben verzameld voor een oproep aan de scholen om spellingvaardigheid niet (meer) te toetsen met meerkeuzeopgaven, en een oproep aan Cito om ook voor de Eindtoets Basisonderwijs om te schakelen naar de dicteevorm.

Het zal duidelijk zijn dat wij hun oproep niet onderschrijven. Wij weten ons daarbij gesteund door de COTAN (Commissie Testaangelegenheden Nederland), die onlangs de toetsen Spelling van Cito beoordeeld heeft. De toetsen kregen op vijf van de zes beoordeelde aspecten het predicaat 'goed' (uitgangspunten, kwaliteit van het toetsmateriaal, kwaliteit van de handleiding, normen en betrouwbaarheid) en op één aspect 'voldoende' (begripsvaliditeit). Over de betrouwbaarheid werd expliciet opgemerkt: De toets is zeer betrouwbaar en voldoet ruimschoots aan de door de COTAN gestelde criteria.

Wij verwachten dat de spellingopgaven in de Eindtoets Basisonderwijs heus nog wel eens een andere vorm krijgen, maar niet naar aanleiding van het onderzoek van Schraven, Bosman en Van Eekhout. En wat de toetsen Spelling betreft: Cito is zojuist begonnen met de voorbereidingen voor wat wij intern 'de derde generatie LOVS-toetsen' noemen. Het gaat om nog te ontwikkelen toetsmateriaal dat in de toekomst de huidige nieuwe toetsen Spelling moet gaan vervangen. Het is nog niet bekend hoe de derde generatie er precies uit gaat zien, noch welke opgaventypen ze zal bevatten. Maar ook hiervoor geldt: op basis van het onderzoek van Schraven, Bosman en Van Eekhout zien we geen reden om bij voorbaat meerkeuzeopgaven uit deze toetsen te weren. Dergelijke opgaven zijn namelijk niet alleen betrouwbaar, maar ook efficiënt, objectief en praktisch!

LITERATUUR

COTAN Beoordelingssysteem voor de kwaliteit van tests (2009). *Amsterdam, NIP/Cotan*.

Schraven, J.L.M., Bosman, A.M.T., & Van Eekhout, T. (2010). De nieuwe Cito-spellingtoets ter discussie. *Tijdschrift voor Orthopedagogiek*, 49 (2010), 75-86.

Verhelst, N.D. (1992). Het één parameter model (OPLM). Een theoretische inleiding en een handleiding bij het computerprogramma. *Arnhem: Cito*.

Verhelst, N.D., & Glas, C.A.W. (1995). The one parameter logistic model. In: G.H. Fischer & I.W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications* (pp. 215-239). *New York: Springer*.

Wijs, A. de, Kamphuis, F., Kleintjes, F. & Tomesen, M. (2010). Wetenschappelijke verantwoording bij de toetsen Spelling. *Arnhem: Cito*.

Wijs, A. de, Krom, R. & Berkel, S. van (2006). LOVS Spelling groep 4. *Arnhem: Cito*.