

De nieuwe Cito-spellingtoets ter discussie¹

J.L.M. Schraven, A.M.T. Bosman & T. van Eekhout

SAMENVATTING

Het doel van deze studie was om na te gaan of de keuze die het Cito heeft gemaakt om ter toetsing van de spellingvaardigheid een onderdeel van het dictee te vervangen door een meerkeuzetoets, verantwoord is. In dit onderzoek is de toets Spelling groep 4 op twee verschillende manieren aan leerlingen voorgelegd. Eerst werd de meerkeuzetoets afgenomen overeenkomstig de instructies in de handleiding. Daarna schreven de leerlingen alle spellingwoorden die centraal stonden in de meerkeuzetoets op basis van een mondeling dictee. Statistische analyses lieten zien dat de psychometrische eigenschappen van de meerkeuzetoets onvoldoende zijn. Ook bleken de totaalscores op de meerkeuzetoets niet samen te hangen met de totaalscores op het dictee. Bovendien maakten de meeste leerlingen het dictee (een productietaak) beter dan de meerkeuzetoets (een herkenningstaak) en waren er grote individuele verschillen tussen de leerlingen op de twee taken. Ten slotte was er een substantiële discrepantie bij ruim een kwart van de woorden. Dat wil zeggen, in 16.0% van de gevallen was de spelling goed op het dictee en fout op de meerkeuzetoets en in 9.8% fout op het dictee en goed op de meerkeuzetoets. Gedetailleerde analyses op itemniveau doen het vermoeden rijzen dat leerlingen op andere gronden keuzes maakten in de meerkeuzetoets dan was bedoeld, namelijk het herkennen van de spellingfout. Kennis van woorden en leesvaardigheid spelen waarschijnlijk een belangrijke rol in deze test die toch vooral spellingkennis moet meten. De auteurs pleiten dan ook voor het herinvoeren van een woorddictee (met een eventuele aanvulling van een zinnen-dictee) om tegemoet te komen aan het eigenlijke doel van het spellingonderwijs, namelijk uit het hoofd kunnen spellen van woorden.

Inleiding

Sinds 2006 heeft het Cito nieuwe spellingtoetsen op de markt gebracht als vervanging van de oude SVS-toetspakketten. Deze nieuwe spellingtoetsen bestaan uit twee delen, namelijk een startmodule en twee typen vervolgmodes. Alle leerlingen beginnen met de startmodule. Deze wordt gevormd door een dictee van 25 woorden dat klassikaal wordt afgenomen. Op basis van de prestaties van de leerlingen op dit dictee worden de leerlingen onderverdeeld

in een groep minder goede spellers en een groep betere spellers. De leerlingen met een lage score krijgen vervolgens de module 'Spelling Vervolg 1', dat in het geval van de groepen 3 tot 5 opnieuw een dictee van 25 woorden betreft. De leerlingen met een hoge score op de startmodule maken de module 'Spelling Vervolg 2'. Dit is een meerkeuzetoets. Deze bestaat uit 25 opgaven waarbij uit telkens vier zinnen het foutgespelde woord omcirkeld dient te worden. In Figuur 1 staat de voorbeeldopgave uit de spellingtest voor groep 4. Overigens zullen de ver-

volgmodules voor zowel de zwakke als de goede spellers uit de groepen 6 tot 8 bestaan uit meerkeuzetoetsen. Het feit dat de betere spellers in de groepen 3-5 meerkeuzeopgaven maken in plaats van een dictee en dat vanaf groep 6 de toetsvorm voor alle leerlingen meerkeuze is, is een breuk met de oude spellingtoetsen.

In de handleiding van de toets Spelling Groep 4 worden redenen gegeven voor het veranderen van de *inhoud* van de spellingtoets. Dit zijn de wettelijk vastgestelde spellingveranderingen en het feit dat er nieuwe spellingmethoden op de markt zijn gekomen. De keuze van het Cito om de *vorm* van de spellingtoets te veranderen, wordt in dezelfde handleiding niet verantwoord. Uit het feit dat het Cito geen verantwoording geeft voor de vervanging van het dictee door een meerkeuzetoets leiden wij af dat de ontwikkelaars veronderstellen dat deze twee toetsvormen inwisselbaar zijn. Het doel van deze studie is om na te gaan of deze veronderstelling terecht is.

E. Ben je in de **zomer** op vakantie geweest?

F. De **caviea** van de burens heeft Sammie.

G. Mijn moeder heeft een **taart** gebakken

H. Weet jij waarom hij zo **raar** doet?

FIGUUR 1 Voorbeeldopgave Spelling M4 Vervolg 2

Meerkeuze als toetsvorm voor spellingvaardigheid

Sinds de invoering van de nieuwe spellingtoets uitlen leerkrachten en intern begeleiders hun bezorgdheid over de vervanging van het dictee door een meerkeuzetoets. Zij vragen zich af of a) het aanstrepen van een woord hetzelfde is als opschrijven ervan, b) wat de gevolgen kunnen zijn van de presentatie van

foutgespelde woorden, en c) welke diagnostische waarde een meerkeuzetoets heeft.

Opschrijven of aanstrepen. In de handleiding van de toets Spelling groep 4 staat de volgende definitie: 'Bij spellen gaat het erom woorden om te zetten in schriftbeelden' (p. 9, de Wijs, Krom & van Berkel, 2006). Wij leiden hieruit af dat spellen een vaardigheid is waar bij men uit het hoofd de orthografische vorm van een woord opschrijft dan wel inrypt. De makers van de spellingtoets lijken er dus vanuit te gaan dat spellen het opschrijven van woorden is en niet het aanstrepen. Het taakverschil tussen het aanstrepen van een woord en het opschrijven ervan, wordt ook wel aangeduid met het onderscheid tussen een herkenningstaak en een productietaak. Een meerkeuzetoets is een goed voorbeeld van een herkenningstaak en een dictee is een voorbeeld van een productietaak. Dat deze twee taken of toetsvormen meestal niet inwisselbaar zijn, is reeds uit onderzoek gebleken (e.g., Bosman & De Groot, 1992, 1996). In veel gevallen, maar niet altijd, is een herkenningstaak makkelijker dan een productietaak.

Presenteren van foutgespelde woorden. Uit de dagelijkse schoolpraktijk kennen we de negatieve consequenties van het aanbieden van foutgespelde woorden; dit wordt ondersteund door wetenschappelijk onderzoek. Het blijkt dat het regelmatig aanbieden van foutgespelde woorden de spellingvaardigheid van spellers negatief kan beïnvloeden. Zowel goede als zwakke spellers blijken niet gebaat te zijn bij blootstelling aan foute spellingen (Brown, 1988; Dixon & Kaminska, 1997; Jacoby & Hollingshead, 1990).

Diagnostische mogelijkheden. Voor leerkrachten is de belangrijkste waarde van een toets of deze informatief is met betrekking tot het soort fouten dat gemaakt wordt. Pas als een toets inzicht geeft in de spellingproblemen van de leerling kan deze adequaat geholpen worden. Vanuit diagnostisch oogpunt bezien is de oorsprong van de fouten die in een dictee worden gemaakt meestal wel duidelijk. Immers, een leerling die GJIT spelt in plaats van GEIT heeft in dit geval een probleem met de foneem-grafeemkoppeling [ei] -Jf of Ei. In

het geval dat er ZOOMER geschreven is in plaats van ZOMER, is het gerechtvaardigd om te veronderstellen dat deze leerling de regel voor de open lettergreep niet beheerst. Bij een meerkeuzetoets is het veel lastiger vast te stellen of en welk spellingprobleem een leerling heeft in het geval deze een fout maakt. Ook wanneer de leerling de juiste keuze maakt, blijft het de vraag in hoeverre hij of zij de spelling van het woord ook daadwerkelijk uit het hoofd kan produceren. Stel dat een leerling in het voorbeeld van Figuur 1 onterecht aankruist dat ZOMER het foutgespelde woord is. De leerling heeft dan weliswaar een mentale 'spelfout' gemaakt, maar de vraag is vervolgens: welke dan? Het kan zijn dat deze leerling dacht dat de juiste spelling ZOOMER is. Het kan echter ook zijn dat deze leerling denkt dat het SOMER moet zijn of zelfs ZOMUR, enzovoort. Omdat we geen idee hebben op grond waarvan deze leerling dacht dat het goed gespelde woord ZOMER fout was, krijgen we dus geen inzicht in het spellingprobleem van deze leerling.

Doel van dit onderzoek

In dit onderzoek zal dus worden nagegaan of het gerechtvaardigd is om voor het bepalen van de spellingvaardigheid het dictee te vervangen door een meerkeuzetoets. Hiervoor zal de toets Spelling medio groep 4 gebruikt worden. Module Spelling Vervolg 2 zal op twee verschillende manieren worden afgenomen. Eerst wordt de toets aan de leerlingen aangeboden zoals staat voorgeschreven in de handleiding van de toets, namelijk als meerkeuzetoets. Vervolgens worden alle toetswoorden in een dicteevorm aangeboden.

De (on)gelijkwaardigheid van de meerkeuzetoets en het dictee zal op meerdere niveaus worden bepaald. We beginnen met de vaststelling van de betrouwbaarheid en validiteit van de dictees en de meerkeuzetoets. Vervolgens gaan we op groepsniveau nader in op de vergelijking tussen de meerkeuzetoets en het dictee. Hierbij zal worden vastgesteld hoeveel leerlingen consistent gedrag laten zien op beide toetsvormen. Ten slotte

wordt er een vergelijking gemaakt tussen de meerkeuzetoets en het dictee op itemniveau. Hiermee wordt inzicht verkregen in de mate waarin elk item zich consistent gedraagt op beide toetsvormen.

Methoden

Proefpersonen

In dit onderzoek participeerden 18 leerlingen (6 meisjes en 12 jongens met een gemiddelde leeftijd van 94 maanden) uit groep 4 van een reguliere basisschool. Geen van de leerlingen had gedoubleerd in Groep 3 of 4. Een leerling (in oktober geboren) heeft Groep 2 twee keer gedaan. De school betreft haar leerlingen voor het grootste deel uit lagere sociaal-economische milieus. Op basis van de nieuwe gewichtenregeling basisonderwijs (Mimisterie van OCW, 2006) bleek één leerling een gewicht van 1,2 te hebben (ouders die niet meer dan basisonderwijs hebben gehad) en één leerling had een gewicht van 0,3 (hiervan hadden de ouders niet meer dan een lbo/vbo-opleiding genoten).

Materiaal en procedure

Het materiaal van dit onderzoek werd, zoals gezegd, gevormd door de spellingwoorden van de toets Spelling M4 (de Wijs e.a., 2006). In januari 2008 heeft de derde auteur van dit artikel (de remedial teacher van een reguliere basisschool) de nieuwe spellingtoets op twee verschillende manieren afgenomen bij de leerlingen van groep 4. Bij de eerste afname werd de in de handleiding voorgeschreven procedure gevolgd. Dit hield in dat alle leerlingen de startmodule aangeboden kregen. Aan de hand van een dictee schreven de leerlingen de 'Startwoorden' op. De remedial teacher bepaalde vervolgens van elke leerling de score en nam daarna Vervolgmodule 2 af. Omdat alle leerlingen een score hadden die hen tot de betere spellers rekenden, konden alle leerlingen meedoen aan dezelfde vervolgttest, dat wil

zeggen de meerkeuzetoets met 25 opgaven (zie Figuur 1 uit de inleiding). De leerlingen moesten uit vier vetgedrukte woorden die in vier verschillende zinnen staan (A, B, C, en D-zinnen), het woord dat fout gespeld is, kiezen (zie Tabel 5). In de tweede fase van het onderzoek werd er op vier verschillende dagen een dictee afgenomen van de vetgedrukte woorden. Op de eerste dag werden de woorden uit de A-zinnen gedictreed, op de tweede die van de B-zinnen, de derde dag werden de C-zinnen aangeboden en op de laatste dag de D-zinnen. Van alle leerlingen werd van elk woord bepaald of dit correct dan wel incorrect gespeld was. De score van elke leerling vormde de input voor de onderstaande analyses.

Resultaten

De beschrijving van de resultaten valt uiteen in drie onderdelen. Eerst worden de psychometrische eigenschappen (i.e. de betrouwbaarheid en de validiteit) van de onderscheiden toetsen bepaald. Daarna zal worden nagegaan of het herkennen van de correcte spelling (meerkeuze) vergelijkbaar is met het produceren ervan (dictee) voor de groep als

geheel. Vervolgens wordt bekeken wat de uitslagen/consequenties zijn voor de individuele leerlingen van deze onderzoeksgroep en ten slotte zal er een vergelijking op itemniveau plaatsvinden.

Bij de presentatie van de resultaten worden de volgende aanduidingen gehanteerd. *Startwoorden* zijn de 25 dicteewoorden uit de startmodule. *Meerkeuzetoets* verwijst naar de meerkeuzeopgaven van vervolgmodule 2 (zie Figuur 1 voor een voorbeeld). De *Dicteewoorden* zijn de foutgespelde woorden uit de meerkeuzetoets die in dicteevorm zijn aangeboden. De *A-woorden*, *B-woorden*, *C-woorden* en *D-woorden* zijn alle dikgedrukte woorden uit respectievelijk de A-zinnen, B-zinnen, C-zinnen en D-zinnen die in dicteevorm zijn aangeboden. De *A-D woorden* zijn alle dikgedrukte woorden uit alle vier de zinnen van de meerkeuzetoets die in dicteevorm zijn aangeboden.

Betrouwbaarheid en validiteit

Op basis van het criterium dat een toets minimaal een alfawaarde van .80 moet hebben wil deze betrouwbaar genoemd kunnen worden, blijkt dat feitelijk geen van de toetsen aan

TABEL 1 Interne consistentiewaarden, Pearson correlaties en gemiddeld aantal goed op alle onderscheiden toetsvormen (standaardafwijkingen tussen haakjes)

Toets	Cronbachs alfa	Correlaties met Meerkeuzetoets	Gemiddeld score ¹
Meerkeuzetoets	.75	-	19.7 (3.5)
Dicteewoorden	.66	.45	21.2 (2.8)
Startwoorden	.74	.64*	21.4 (3.1)
A-woorden	.65	.64*	20.7 (2.9)
B-woorden	.40	.36	20.1 (2.2)
C-woorden	.27	.38	21.9 (1.6)
D-woorden	.39	.65*	20.6 (2.2)
A-D woorden ²	.79	.68*	20.8 (1.8)

* significante correlaties, $p < .05$;

¹ De maximumscore op alle toetsen is 25;

² De score op de A-D woorden is het gemiddelde van de vier dictrees

die betrouwbaarheidsvoldoet (zie tweede kolom van Tabel 1). Ten aanzien van de criteriumvaliditeit zien we dat hoewel vier van de acht correlaties significant zijn, de hoogten van de correlaties allemaal onvoldoende zijn (lager dan .80) om de meerkeuzetoets valide te noemen (zie derde kolom van Tabel 1). De belangrijkste correlatie, die tussen de meerkeuzetoets en de dicteewoorden, was zelfs niet statistisch significant. De conclusie is dus dat de meerkeuzetoets niet hetzelfde meet als het dictee en dat ze dus *niet* inwisselbaar zijn.

Meerkeuze versus dictee op groepsniveau

In de laatste kolom van Tabel 1 staan de gemiddelde scores op alle toetsvormen. Om te bepalen of de leerlingen de meerkeuzetoets beter dan wel slechter maakten dan de dictrees zijn de gemiddelde scores met elkaar vergeleken. Een eerste interessante bevinding is dat de gemiddelde score op de meerkeuzetoets lager is dan 20, terwijl die van alle verschillende dictrees hoger is dan 20. De statistische toetsen uit Tabel 2 laten zien dat het verschil tussen de meerkeuzetoets en de dicteevormen significant is. Dit betekent dat leerlingen de dicteevormen significant beter maakten

dan de meerkeuzetoets. Ook de dictrees over de C-woorden, de A-D woorden en de Startwoorden waren significant beter gemaakt dan de meerkeuzetoets. Het dictee over de Startwoorden en de Dicteewoorden werd even goed gemaakt.

Meerkeuze versus dictee op individueel niveau

In Tabel 3 staan de individuele scores op de meerkeuzetoets en de dicteevormen. Voor vier leerlingen (22.2%) was de meerkeuzetoets makkelijker dan de dicteevormen, voor drie leerlingen (16.7%) was er geen verschil, terwijl voor 11 leerlingen (61.1%; de leerlingen 2, 3, 5, 6, 9, 11, 12, 13, 14, 17, en 18) de meerkeuzetoets moeilijker was dan het uit het hoofd opschrijven van de dicteevormen. Met name voor de leerlingen 2, 12 en 14 bleek het bijzonder moeilijk te zijn indien hun spellingvaardigheid middels een meerkeuzetoets bepaald werd.

Meerkeuze versus dictee op itemniveau

In Tabel 4 staan een aantal belangrijke statistische gegevens per item. In de eerste drie kolommen staat per item het percentage leer-

TABEL 2 Statistische toetsing van het verschil tussen twee toetsvormen

Vergelijking	t-waarden ¹	p-waarden ²
Meerkeuze vs. Dicteewoorden	$t(17) = -1.89$	$p < .04^*$
Meerkeuze vs. A-woorden	$t(17) = -1.43$	$p = .09$
Meerkeuze vs. B-woorden	$t(17) = -0.42$	$p = .34$
Meerkeuze vs. C-woorden	$t(17) = -2.89$	$p < .005^*$
Meerkeuze vs. D-woorden	$t(17) = -1.33$	$p = .10$
Meerkeuze vs. A-D woorden	$t(17) = -1.70$	$p < .05^*$
Startwoorden vs. Meerkeuze	$t(17) = -2.52$	$p < .01^*$
Startwoorden vs. Dicteewoorden	$t(17) = -0.34$	$p = .37$

* significante verschillen;

¹ Eenzijdige toetsing; gepaarde waarnemingen

TABEL 3 Individuele scores en verschildscores

Leerling	Meerkeuzetoets	Dictewoorden	Verschild
1	23	19	4
2	15	23	-8
3	23	25	-2
4	19	18	1
5	22	24	-2
6	23	24	-1
7	19	19	0
8	24	24	0
9	20	23	-3
10	18	16	2
11	19	21	-2
12	17	22	-5
13	18	19	-1
14	10	19	-9
15	20	17	3
16	23	23	0
17	19	22	-3
18	23	24	-1

* significante verschillen;

† Eenzijdige toetsing; gepaarde waarnemingen

lingen dat dit item goed had aangekruist op de meerkeuzetoets, het percentage leerlingen dat dit item goed gespeld had op het dictee en het verschil tussen beiden. De *t*-testen laten vervolgens van ieder item zien of er een significant verschil was in de prestaties op de meerkeuzetoets en het dictee. Dit was het geval voor 7 van de 25 items. Het betrof *zuwak*, *scherm*, *sloep*, *aanrecht*, *lusje*, *goed* en *poetje*. Uit de analyse over alle items blijkt dat het gemiddelde verschil tussen de prestaties op de meerkeuzetoets en het dictee in het voordeel is van het dictee. Het op het oog slechts geringe gemiddelde significante verschil van 6% (zie laatste regel van Tabel 4) krijgt een geheel andere betekenis wanneer het tweede deel van Tabel 4 in beschouwing wordt genomen.

Als de meerkeuzetoets hetzelfde meet als het dictee dan zouden er geen leerlingen moeten voorkomen in de twee kolommen die onder discrepantie staan. Dit blijkt voor slechts één woord te gelden, namelijk STANK. Op de woorden HOND en KLANK is er bij een leerling een discrepantie. Bij 11 van de 25 woorden (44%) is er bij minimaal 25% van de leerlingen geen overeenstemming tussen dictee en meerkeuzetoets en bij 3 van de 25 woorden is er zelfs bij meer dan 50% van de leerlingen geen overeenstemming (BOOT-TOCHT, ZWAK en VLOT). Dit kan twee dingen betekenen. In het ene geval concludeert de leerkracht ten onrechte op basis van de meerkeuzetoets dat een leerling de spelling van een woord niet beheerst, terwijl op het dictee blijkt dat ze dat wel kunnen (e.g., ZWAK). In het andere geval zou onterecht geconcludeerd worden op basis van de meerkeuzetoets dat een leerling de spelling van het woord beheerst, terwijl het dictee uitwijst dat dit niet het geval is (e.g., AANRECHT). Uit de gemiddelden blijkt dat 16% van de leerlingen niet het fout gespeld woord op de meerkeuzetoets kiest, terwijl zij dit woord op het dictee wel correct spellen, en 9,8% van de leerlingen kiest wel het fout gespeld woord op de meerkeuzetoets, terwijl ze dit fout spellen op het dictee.

Dus, in een kwart van de gevallen (9,8% + 16,0% = 25,8%) was er een discrepantie tussen de prestaties op de meerkeuzetoets en het dictee. Statistische toetsing middels *Phi*, Pearson correlatie, laat zien dat er slechts voor één item (i.c. *scherm*) een significante samenhang is. Dit bevestigt het inconsistente patroon dat waargenomen kan worden in deze gegevens. De kans dat een item goed is gemaakt op de meerkeuzetoets heeft nauwelijks een relatie met de kans dat dit item op het dictee ook goed gespeld wordt en andersom. Dit betekent voor de diagnostiek dat een leerkracht op basis van de meerkeuzetoets per woord bij gemiddeld ruim een kwart van de leerlingen een onjuist beeld krijgt van de spellingvaardigheid van een leerling.

In Tabel 5 staat per meerkeuzetoets alle vier de mogelijke opties het percentage

leerlingen dat dit item als het fout gespeld woord koos. Ook staat per item aangegeven het percentage leerlingen dat dit woord op het dictee correct spelde. Dit overzicht brengt een aantal merkwaaardigheden aan het licht.

De eerste meerkeuzetoets is het duidelijkste voorbeeld dat een meerkeuzetoets leerlingen in verwarring kan brengen. Hier moesten de leerlingen kiezen tussen de vier alternatieven VLAG, WRAK, GRAP en SWAK. Minder dan een kwart van de leerlingen gaf correct aan dat SWAK het fout gespeld woord was; de meeste leerlingen besloren dat WRAK fout gespeld was. In het dictee bleek dat de overgrote meerderheid van de leerlingen wel degelijk de spelling van ZWAK kende, terwijl slechts een heel gering percentage leerlingen het woord WRAK correct spelde. Als alleen de meerkeuzetoets was afgenomen, dan was ten onrechte geconcludeerd dat de leerlingen de spelling van ZWAK niet beheersen. Onze analyse leidt echter tot een alternatieve conclusie: de leerlingen kenden zeer waarschijnlijk het woord WRAK niet.

De negende meerkeuzetoets is een voorbeeld van de rol die kennis van de woordbetekenis of het ontbreken daarvan kan spelen bij het maken van de meerkeuzetoets voor spelling. Hier moesten de leerlingen kiezen tussen de vier alternatieven GOUDGEEL, LICHT, LUSJU en GEULEN. Iets meer dan de helft koos correct voor LUSJU, terwijl ruim een kwart GEULEN koos als de foute spelling. Op de dictees bleek echter dat zowel LUSJE als GEULEN door vrijwel iedereen correct gespeld werd. Het vermoeden bestaat dat niet alle leerlingen het woord GEULEN kenden en dat dit in de meerkeuzetoets tot verwarring heeft geleid. Bij een dictee waar het gaat om toepassing van de kennis van de spelling, en in het geval van een verkleinwoord kennis hebben van het achtervoegsel JE, is het meestal niet nodig om de betekenis te kennen om het woord correct te spellen.

Ook bij de negentiende meerkeuzetoets lijkt verwarring te zijn ontstaan. De alternatieven waren: BLADZIJDE, DEKSOL, BRIEFJE en BUURMAN. Het feit dat 78% van de leerlingen wel degelijk in staat was om DEKSEL correct te spellen op het dictee, terwijl 61% dit