

# Het gebruik van foute woordbeelden in spellingtoetsen

## Rapport over een experimenteel onderzoek

L. HEYERICK

*Pedagogisch Centrum, Gent*

### *Samenvatting*

*Aan spellingtoetsen in multiple-choice vorm zijn twee problemen verbonden die hier nader onderzocht worden. Een experiment werd opgezet om na te gaan of het aanbieden van spelfouten in de toets een negatief effect heeft op de verdere spellingvaardigheid van de leerlingen. Geen enkel van de hier verkregen gegevens wees in die richting.*

*Een tweede probleem is de validiteit. Op basis van de in het hier beschreven experiment en in andere onderzoeken verkregen correlaties, kan niet uitgesloten worden dat een meerkeuzetoets een enigszins ander aspect meet van spellingvaardigheid dan dictees of meer spontane schriftelijke taalprodukten. Naast een kennis van de juiste schrijfwijze is voor het goed oplossen van meerkeuze spellingitems ook een vaardigheid in kritisch nalezen vereist. Ook die vaardigheid kan op school geleerd worden.*

*In de mate dat een correcte spelling belangrijk gevonden wordt, lijkt een meerkeuze spellingtoets derhalve een voor het onderwijs bruikbaar evaluatie-instrument.*

### *1. Problemen bij het toetsen van spellingvaardigheid*

Meer en meer wordt bij evaluatie van onderwijsleerprocessen gebruik gemaakt van objectieve studietoetsen, doorgaans in meervoudige keuzevorm. Deze wijze van evalueren is in onderwijskringen echter aan vrij veel kritiek onderhevig (zie bv. Bonset en Griffioen, 1976). Deze kritiek heeft niet alleen aanleiding gegeven tot verdere discussie (o.a. Prick en Wesdorp, 1976), maar ook tot enig onderzoek. Een aantal bezwaren tegen multiple-choice toetsen kon daardoor ondervangen worden: discriminatie van leerlingen uit lagere milieus kon niet aangetoond worden, evenmin kon geconcludeerd worden dat meerkeuzetoetsen eenzijdige en slechts weinig complexe doelstellingen zouden

nakijken (Zijlmans en Wesdorp, 1979; Zijlmans, e.a. 1978).

Meer specifieke bezwaren worden echter geuit tegen het toetsen van de spellingvaardigheid door middel van meerkeuzetoetsen. Eerst en vooral is er sprake van een validiteitsprobleem. De uiteindelijke doelstelling van spellingonderricht is dat de leerlingen zo weinig mogelijk spelfouten zouden maken in alles wat ze schrijven. De vraag is dan in welke mate die doelstelling adequaat geoperationaliseerd wordt door een multiple-choice spellingtoets. Schäfer (gecit. in Ingenkamp, 1962, p. 181-182) vond tussen de resultaten op een meerkeuzetoets en het niet genormeerde oordeel van leerkrachten over de spellingvaardigheid van hun leerlingen een correlatie van .65. Freyberg (1970) en Wesdorp (1972) hanteerden het aantal fouten in respectievelijk één en vijf opstellen als criterium en verkregen eveneens eerder lage validiteitscoëfficiënten voor hun meerkeuzetoetsen: .68 en .58.

De validiteit van zgn. kerndictees (één woord uit een voorgelezen zin moet door de leerlingen opgeschreven worden) is niet noodzakelijk hoger t.a.v. het opstelcriterium: bij Freyberg (1970) .72. Maar deze vorm van spellingtoets is in het onderwijs veel meer ingeburgerd en er worden heel wat minder bezwaren tegen geformuleerd. Wanneer een dergelijk dictee als criterium genomen wordt, is de validiteit van meerkeuzetoetsen doorgaans hoger, zij het ook niet altijd bevredigend: .74 bij Schäfer, .85 bij Freyberg (1970), .72 bij Wesdorp (1972) en .78 in een meer recent onderzoek van Kelso (1978). In de laatste twee vermelde onderzoeken waren de spellingmoeilijkheden in toets en dictee vrijwel identiek.

Wat leerkrachten zowel in lager als voortgezet onderwijs echter nog meer lijkt te storen in spellingtoetsen is het feit dat daarbij aan de leerlingen spelfouten gepresenteerd worden. Zij vrezen dat die gedrukte spelfouten de leerlingen in verwarring brengen, wat uiteindelijk tot meer spel-

fouten aanleiding zou geven. De onderliggende redenering is hier dat foute woordbeelden ingeslepen geraken, net zoals ze er bij de spellingdidactiek van uitgaan dat spellingleren vrijwel uitsluitend een visuele aangelegenheid is. Van der Geest e.a. (1978) hebben reeds een aantal argumenten bijeengebracht tegen een eenzijdig visuele benadering van het spellingonderricht. Ondermeer blijkt lezen – en dus zien – alleen van de juiste schrijfwijze niet automatisch tot spellingbeheersing te leiden: de woorden waarin de meeste fouten worden gemaakt in het lager onderwijs behoren vrijwel allemaal tot de meest frequent gehanteerde leesteksten voor die leeftijd. Maar de tot dusver verzamelde evidentie laat eigenlijk geen definitieve conclusies toe.

Beide problemen rond spellingtoetsen – validiteit en aanbieder van spelfouten – lijken ons de moeite waard om verder onderzocht te worden. Wij hebben een experiment opgezet met de hoofdbedoeling het differentieel effect van de vorm van de toets op de spellingvaardigheid na te kijken. Door ook bij pre- en posttests verschillende vormen te hanteren konden wij zijdelings informatie omtrent de validiteit verkrijgen.

## 2. Het onderzoek: populatie en procedure

Aan dit onderzoek werkten tien klassen mee uit de eerste twee leerjaren voortgezet onderwijs: 7 eerste observatieklassen en 3 tweede observatieklassen uit vier verschillende middensholen. Dit vertegenwoordigde een onderzoekspopulatie van ongeveer 200 leerlingen van 12 tot 14 jaar.

Binnen elke klas werden op grond van een toevalsprocedure drie groepen gevormd. De op deze wijze samengestelde groepen verschilden vrijwel niet in spellingvaardigheid, zoals gebleken is uit de variantie-analyse van de resultaten van een begintoets (Kerndictee van 40 woorden – Tabel 1).

De drie groepen kregen vervolgens een verschillende experimentele behandeling, gespreid over 5 sessies, met telkens ongeveer 3 weken tussen twee opeenvolgende sessies.

In elke experimentele sessie kreeg groep A eerst een kerndictee met 20 spellingmoeilijkheden. Groep C kreeg dezelfde moeilijkheden aangeboden, maar in de vorm van een tekst waarbij zij moesten oordelen of de onderstreepte woorden goed of fout geschreven waren. Hierbij waren drie kwart van de woorden fout gespeld. We kozen voor meer dan de helft fouten omdat het juist de bedoeling was het effect van visueel aangeboden spelfouten na te kijken. Anderzijds konden we ook niet alles fout

schrijven omdat er dan geen vergelijking met een toetsituatie meer zou bestaan. Verder opteerden we voor een true-false toetsvorm, omdat we identieke spellingmoeilijkheden als bij het kerndictee wilden aanbieden. De leerlingen moesten over elk van de woorden beslissen of het goed of fout geschreven stond. Bij de klassieke multiple-choice moeten zij niet meer naar de andere woorden kijken als ze in 1 van de alternatieven een fout ontdekt hebben<sup>2</sup>.

Groep B fungeerde als controlegroep en werkte gedurende de spellingoefeningen van de andere groepen zelfstandig aan een opdracht die niet rechtstreeks met spelling te maken had: doorgaans was dat het begrijpend lezen van een tekst. Na de toetsen werd de juiste schrijfwijze van de 20 woorden klassikaal besproken, ook met de leerlingen uit de controlegroep. Om de procedure in alle klassen zo gelijk mogelijk te laten verlopen, werden de sessies overal door dezelfde leerkracht, gedetacheerd naar het Stedelijk Pedagogisch Centrum te Gent, geleid. Uiteindelijk werden op de bovenbeschreven wijze 100 spellingmoeilijkheden met de leerlingen behandeld. De eindtoets werd voor 90% samengesteld uit de behandelde moeilijkheden: identieke voor de vaste woordbeelden, toepassing van dezelfde regels voor de werkwoorden. Deze toets bevat twee delen: een kerndictee met 25 vaste woordbeelden en 25 werkwoordvormen en een goed/fout gedeelte met dezelfde 25/25 verdeling<sup>3</sup>. Deze subtoetsen werden onderscheiden omdat, als het zien van spelfouten een negatief effect heeft, we dit vooral op de vaardigheid in het spellen van vaste woordbeelden zouden verwachten. Op het einde van het schooljaar 1978-'79 werd nog een meerkeuzespellingtoets (met 4 alternatieven) afgenomen in alle middensholen van het Gents stedelijk onderwijs, waartoe ook de tien klassen behoorden, die bij ons experiment betrokken waren.

Analyse van de resultaten op die centraal verwerkte toetsen, voor die leerlingen die aan het experiment deelnamen, kon ons nog verdere informatie leveren. Deze toetsen waren wel verschillend voor eerste en tweede observatiejaar.

Tabel. 1 *Groepsgemiddeldes op de pretest*

	groep A	groep B	groep C	totaal	
M	24,93	25,34	25,06	25,11	F. = 0,09
s	5,68	5,98	4,67	5,45	P > .90
n	67	68	67	202	

Tabel 2 *Betrouwbaarheidscoëfficiënten eindtoets*

	K R .20
Totaal (100 items)	.840
Subtoets 1 (25 items, werkwoord)	.670
Subtoets 2 (25 items, vast)	.591
Tot. kerndictee (50 items)	.767
Subtoets 3 (25 items, werkwoord)	.532
Subtoets 4 (25 items, vast)	.638
Tot. Goed/Fout (50 items)	.711

### 3. De resultaten

#### 3.1 De betrouwbaarheid van de toetsen

De betrouwbaarheidscoëfficiënten, berekend volgens de interne consistentie-benadering (KR.20 – Nunally, 1967), hebben we uitgezet in Tabel 2. Gezien het vrij groot aantal items lijken deze coëfficiënten eerder laag, tenminste als we veronderstellen dat op dertienjarige leeftijd spelling reeds een homogene vaardigheid is. De betrouwbaarheid werd o.i. echter ook gedrukt door toetsgebonden factoren. Een aantal items bleken te gemakkelijk – een tiental bereikte een p-waarde van .95 terwijl de range van de p-waarden tussen .33 en .99 lag. Ook werden een aantal items opgenomen waarbij de toepassing van de voorkeurspelling eerder arbitrair is (*collega, vakantie, ...*): 'goede' en 'zwakke' leerlingen maken hier evenveel fouten. Verder is er voor de subtoetsen 3 en 4 (goed/fout) een hoge raadkans die .50 bedraagt. Wij hebben daar toch voor geopteerd om dezelfde spellingmoeilijkheden als tijdens het experiment te kunnen aanbieden. Overigens is de raadkans bij heel wat problemen in het kerndictee nauwelijks lager: allerlei, pauw, noch, parade ..., ook heel wat werkwoordsvormen worden slechts op één manier fout geschreven. Leerlingen die twijfelen kunnen ook met 50% kans op succes de juiste spelling raden.

Bij meervoudige keuzetoetsen met 4 alternatieven is de raadkans kleiner, en verkrijgen we een betrouwbaarheid met 40 items die vergelijkbaar is met die van de subtoetsen kerndictee en goed/fout met 50 items: respectievelijk bedroeg K R 20. 77 voor de toets in het eerste observatiejaar en .72 voor die van het tweede observatiejaar<sup>4</sup>.

#### 3.2 Het effect van 'foute woordbeelden'

In Tabel 3 kunnen we zien dat de verschillen tussen de groepsgemiddeldes op verre na niet statistisch

significant zijn. Alle bijkomende analyses wijzen in dezelfde richting. Op geen enkele subtoets is er enig statistisch significant verschil tussen de groepsgemiddeldes merkbaar.

Om tegemoet te komen aan de hypothese van enkele leerkrachten als zou een eventueel nadelig effect van foute woordbeelden vooral bij zwakkere leerlingen voorkomen, hebben we alle analyses overgedaan met enkel die leerlingen die op de pretest een score beneden het gemiddelde gehaald hadden. Ook binnen deze groep van minder spellingvaardige leerlingen vonden we geen significant verschil tussen de drie groepen.

Een laatste controle werd één maand na de posttest uitgevoerd. Hierbij hebben we gebruik gemaakt van centraal verwerkte spellingtoetsen in meerkeuzevorm. Analyses moesten apart voor 1ste en 2de jaar gebeuren omdat er verschillende toetsen per leerjaar werden afgenomen. Ook uit de prestaties op deze toetsen blijkt dat in de spellingvaardigheid van de 3 groepen geen noemenswaardige verschillen optreden. Verder hebben we nog nagekeken of er significante verschillen op bepaalde items voorkwamen.

Slechts 8 woorden van de 100 werden door één groep significant slechter gespeld dan door één van de andere groepen. Dit lijkt ons evenmin de toevalskans te overstijgen. Bovendien is er geen lijn te ontdekken in de aard van die fouten die in een bepaalde groep meer gemaakt werden.

Alles samen genomen vinden we geen enkele aanwijzing tegen het aanbieden van spelfouten in meerkeuzetoetsen. De twijfel of verwarring die hierdoor bij de leerlingen geïntroduceerd werd, tast hun reeds verworven spellingvaardigheid blijkbaar niet aan, net zo min als een dictee waarbij de leerlingen zelf moeten schrijven. Dit laatste werd door leerkrachten nooit betwijfeld hoewel leerlingen daarbij als ze fouten schrijven zowel met een visueel als een motorisch fout woordbeeld geconfronteerd worden. Vergeten we echter niet dat het bij toetsen gaat om een éénmalige aanbieding van fouten. We kunnen moeilijk verwachten dat na die ene keer een foutieve schrijfwijze ingeslepen is. Spellingdidactiek is immers in de eerste plaats gericht op het

Tabel 3 *Gemiddeldes op de posttest*

	groep A	groep B	groep C	totaal	
M	75,37	75,84	74,98	75,41	F. = 0,06
s	9,83	10,02	8,89	9,57	p > .90
n	67	67	59	193	

Tabel 4 *Correlaties tussen de verschillende toetsen*

		1ste observatiejaar (n = 132)			
		1	2	3	4
1.	pretest (Kerndictee)		.786(1.00)	.596 (.786)	.571(.751)
2.	posttest (Kerndictee)	.653(.861)		.665(.901)	.688(.895)
3.	posttest (Goed/Fout)	.404(.553)	.424(.574)		.767(1.00)
4.	Meerv. keuze toets	.312(.425)	.383(.515)	.495(.692)	

2de observatiejaar (n = 61)

verwerven van een automatisme, wat vooral door herhaling, zij het in variërende context, te bereiken is. Wel kunnen we in dit verband stellen dat meerkeuzetoetsen voor sommige leerlingen moeilijker zijn, omdat ze niet op het automatisme, dat ze bij het zelf schrijven verworven hebben, kunnen terugvallen. Vandaar zouden we het advies willen formuleren dat leerlingen bij het oplossen van spellingtoetsen de woorden waaraan ze twijfelen mogen opschrijven. We hebben trouwens geconstateerd dat leerlingen bij het oplossen van de 'Goed/fout'-toets spontaan naar dit verworven motorisch automatisme teruggrepen.

### 3.3 *Het validiteitsprobleem*

Tussen de verschillende toetsen die bij dit spellingonderzoek werden afgenomen hebben we correlaties berekend. In Tabel 4 drukken we die afzonderlijk af voor eerste (boven diagonaal) en tweede observatiejaar (onder diagonaal). In tabel 5 voor die toetsen die van de totale groep werden afgenomen.

We zien dat in elk geval van de onderscheiden groepen leerlingen het hoogste verband bestaat tussen die toetsen die qua vorm gelijk van aard zijn: de correlaties tussen de twee kerndictees enerzijds en tussen de goed/fout-toets en de meervoudige

keuzetoets anderzijds zijn telkens hoger dan die tussen de twee posttesthelften (Kerndictee en goed/fout) die nochtans met de minste tussentijd zijn afgenomen (hoogstens een week). Tussen de twee kerndictees lag een tussentijd van 5 maanden en tussen de posttest en de meerkeuzetoets meer dan één maand. Ook na correctie voor attenuatie (de coëfficiënten tussen haakjes in Tabel 4 en 5) blijft dit patroon behouden. De twee kerndictees enerzijds en de goed/fout en meerkeuzetoets anderzijds zijn te beschouwen als echte paralleltoetsen. De bovengrens van die intercorrelaties is dan ook 1.00 bij maximale betrouwbaarheid. Ook na correctie voor onbetrouwbaarheid blijven de correlaties tussen toetsvormen die verschillen van aard telkens lager.

Op basis van deze correlaties zijn we dan ook geneigd te concluderen dat het schrijven van gedichteerde woorden en het herkennen van spelfouten in gedrukte teksten twee aspecten van spellingvaardigheid zijn die elkaar niet helemaal overlappen. Beide deelvaardigheden lijken ons noodzakelijk om het uiteindelijke doel van het spellingondericht te bereiken en derhalve is het ook zinvol om beide toetsvormen te laten hanteren in het onderwijs. Bovendien lijkt het ook zinvol om beide deelvaardigheden aan te leren. Naast spellingbeheersing moet ook kritische zin aangeleerd worden: leerlin-

Tabel 5 *Correlaties tussen de verschillende toetsen*  
(totale groep n = 193)

	1	2	3
1. pretest (Kerndictee)	1.000	.782(1.00)	.582(.797)
2. posttest (Kerndictee)		1.000	.634(.859)
3. posttest (Goed/Fout)			1.000

gen moeten leren fouten ontdekken bij het lezen van teksten van anderen of bij het herlezen van wat ze zelf geschreven hebben. Doordat meer en meer teksten met de machine geschreven worden, moeten zij ook leren tikfouten te herkennen. Deze vorm van kritisch nalezen op spelling is iets anders dan het weten hoe een woord gespeld moet worden, omdat tikfouten doorgaans niet te wijten zijn aan een gebrek aan spellingvaardigheid van schrijver of typist.

Ook de Nijmeegse werkgroep Taaldidactiek (1974) legt de nadruk op dit kritisch nalezen. 'In de eindfase van het spellingonderricht – en dit is het belangrijkste argument om een deel van de correctie uit handen te geven – moet de schrijver de attitude bezitten om zijn eigen werk na te kijken. Niet alleen op spelling maar ook op andere aspecten. Deze zelfstandigheid moet van het begin af gestimuleerd en geoefend worden. Wanneer het geen controledictees betreft, kan de leerkracht daarom nakijkwerk verantwoord uit handen geven, mits hij natuurlijk wel attent bij de groep aanwezig is' (p. 127). Dat leerlingen elkaars teksten zouden corrigeren wordt – althans in Vlaanderen – door heel wat onderwijsinspecties stellig afgeraden, omdat zij vrezen voor een nadelig effect van het zien van spelfouten. Het hier beschreven onderzoek heeft echter op geen enkele wijze een dergelijk negatief effect kunnen aantonen.

#### Noten

1. Dat deze vrees niet alleen in Vlaanderen en Nederland bij leerkrachten moedertaal leeft, maar bv. ook in Duitsland, moge blijken uit volgend citaat van Ingenkamp (1962): 'Die Aufstellung dieser Tests wird grundsätzlich Untersuchungen über die Frage erfordern, ob falschgedruckte Wörter die spätere Richtig-schreibung negativ beeinflussen. Unter Lehrern gibt das meist als Dogma, ein empirischer Beweis steht aus' (p. 103-104).
2. De aanbieding in deze experimentele toetsen was bv. als volgt:
  - Omtrent de grootte van de buit heeft de bank nog niets *meegedeelt*. goed/foutIn klassieke multiple-choice met 4 alternatieven zou de aanbieding er zo kunnen uitzien:
  - Omtrent de *grootte* (a) van de *buit* (b) heeft de bank nog niets *meegedeelt* (c). *Geen fout* (d).Als een leerling hierbij a aanduidt omdat hij denkt dat daar een fout staat, weten we dus niet meer zeker of hij de schrijfwijze van *meegedeelt* nog wel in overweging genomen heeft, omdat hij slechts één alternatief mag aangeven.
3. De indeling en benoeming van de subtoetsen is enigszins arbitrair. Ten eerste zijn er werkwoordsvormen, waarbij leerlingen vooral fouten maakten die met de vervoeging niets te maken hebben: bv. bij

'geïmporteerd' kwamen drie fouten voor: het deelteken weglaten, n in plaats van m, en t in plaats van d. Enkel de laatste fout is een werkwoordfout. Ten tweede dekt de term 'vaste woordbeelden' niet helemaal de lading omdat ook de spellingmoeilijkheden verbonden aan verbuiging van adjectieven en meervoudsvorming van substantieven opgenomen werden.

4. Deze gegevens zijn gebaseerd op afname van de toetsen bij 340 leerlingen uit eerste observatiejaar en 472 uit tweede observatiejaar.

#### Literatuur

- Bonset, H., en J. Griffioen, CITO-materiaal voor 5e klas kritisch getoetst, Met toets en band door het ganse land, *Het Beste uit Moer* 69-73, Veghel, 1976, 60-69.
- Freyberg, P. S., The concurrent validity of two types of spellingtest, *British Journal of Educational Psychology*, 1970 (40) 68-71.
- Geest, A., van der, e.a., *Spellingwijzer*, Praxis 14 Malmberg, Den Bosch, 1978.
- Ingenkamp, K., *Die deutschen Schulleistungstests*, Verlag Julius Beltz, Weinheim, 1962.
- Kelso, G., *Spelling correlation on handwritten vs. multiple choice tests* El Centro College, Dallas, 1978, ERIC Reports no. ED 157088.
- Nijmeegse Werkgroep Taaldidactiek, *Taaldidactiek aan de basis*. Wolters-Noordhoff, Groningen, 1974.
- Nunnally, J. C., *Psychometric Theory*. McGraw Hill, New York, 1967.
- Prick, L., en H. Wesdorp, Bruikbaarheid van objectieve studietoetsen in het onderwijs. *Het Beste uit Moer* 69-73, Veghel, 1976, 69-76.
- Wesdorp, H., *De validiteit van de subtoetsspelling uit de schooltoetsen basisonderwijs van het CITO*, Memo no. 34. CITO, Arnhem, 1972.
- Zijlmans, S., e.a., *De onderwijskundige aspecten van het gebruik van objectieve studietoetsen*, Interim-rapport, S.V.O.-project 0313.. RITP, Amsterdam, 1978.
- Zijlmans, S., en H. Wesdorp, Objectieve studietoetsen en hun effect op het taalonderwijs, *Levende talen*, 1979 (341) 271-287.

#### Curriculum vitae

Heyerick, L. (geb. 1949) behaalde in 1971 het licentiaat psychologie aan de Rijksuniversiteit te Gent; was van 1971 tot 1976 wetenschappelijk medewerker aan de Rijksuniversiteit te Gent en de Universitaire Faculteit Sint-Ignatius te Antwerpen; van 1976 af adviseur van het Stedelijk Pedagogisch Centrum te Gent. Hij promoveerde in 1978 tot doctor in de ontwikkelingspsychologie met als dissertatie: 'De schoolse achterstand van arbeiderskinderen, een onderzoek naar het aandeel van de leerkracht – leerling relatie in het tot stand komen van die problematiek'.

Adres: Pedagogisch Centrum, Guinardstraat 30, 9000 Gent (België)