

'Een verkenningsstocht naar een betrouwbare beoordeling'

Een instrumentatieonderzoek naar het
beoordelen van schrijfvaardigheden
door middel van een gesloten toets

19 april 2011
Universiteit Utrecht

Stijn Kling
3372715

Bachelor Eindwerkstuk Communicatiekunde
Scriptiebegeleider: Prof. dr. Huub van den Bergh
Versie 2



PDF Complete
Your complimentary use period has ended.
Thank you for using PDF Complete.

[Click Here to upgrade to Unlimited Pages and Expanded Features](#)

H1: INLEIDING

1.1 Schrijfvaardigheden	blz. 2
1.2 Taalgebruik	blz. 3

H2: METHODE

2.1 Materiaal	blz. 6
2.2 Pretest	blz. 7
2.3 Procedure	blz. 8
2.4 Essays	blz. 9

H3: RESULTATEN

3.1 Interne resultaten	blz. 10
3.1.1 Inhoudsvaliditeit	blz. 10
3.1.2 Constructvaliditeit	blz. 10
3.2 Externe resultaten	blz. 11

H4: CONCLUSIE & DISCUSSIE

4.1 Conclusie	blz. 13
4.2 Discussie	blz. 14

BIBLIOGRAFIE	blz. 16
---------------------	---------

Bijlage 1: Afnamemateriaal test taalgebruik (met aangegeven items)

Bijlage 2: Overzicht gebruikte items

Bijlage 3: Overzicht items en de eisen

1.1 Schrijfvaardigheden

Al vanaf het begin van de twintigste eeuw is er een discussie gaande over welke methode de beste is om schrijfvaardigheid te meten (Van Schooten, 1988). Braddock (1969) noemt het beoordelen van schrijfproducten zelfs een *especially frustrating problem*. De problemen bij het meten van schrijfvaardigheid worden veroorzaakt doordat de hoeveelheid foutenvariantie bij schrijfvaardigheidsmetingen vaak erg groot is (Van Schooten, 1988). Hierdoor ontstaan er geen valide metingen en kunnen er dus ook geen legitieme uitspraken over schrijfvaardigheden worden gedaan. Een meting wordt valide genoemd indien met die meting datgene gemeten wordt wat men met die meting beoogde te meten (Van Den Bergh & Meuffels, 2000). Schrijfvaardigheidsonderzoekers proberen al jaren specifieke middelen te bedenken en deze ook empirisch te beproeven om de beoordeling van schrijfproducten betrouwbaarder en daarmee meer valide te maken (Van Den Bergh & Meuffels, 2000). Alvorens we ingaan op mogelijke oplossingen voor dit probleem, worden de belangrijkste gebreken bij de beoordeling van schrijfvaardigheden kort toegelicht.

Het eerste probleem heeft te maken met de beoordelaarsbetrouwbaarheid. Het is niet gemakkelijk om een schrijfoopdracht te beoordelen (Wesdorp, 1981). Uit empirisch onderzoek van de afgelopen zeven decennia is bij herhaling gebleken dat beoordelaars weinig stabiel zijn (Van Den Bergh & Meuffels, 2000). Een opstel dat bij een eerste beoordeling een hoog cijfer kreeg, kan bij een volgende beoordeling van dezelfde beoordelaar een gemiddeld cijfer krijgen. Dit kan bijvoorbeeld liggen aan vermoedigheid van de beoordelaar, aan onvermogen om de kwaliteit van schrijfproducten exact te beoordelen of aan de volgorde waarin opstellen worden aangeboden (Van Schooten, 1988). De instabiliteit geldt ook voor de overeenstemming tussen verschillende beoordelaars onderling. De correlatie tussen beoordelaars ligt over het algemeen zeer laag (Van Den Bergh & Meuffels, 2000). De ene beoordelaar kan een schrijfproduct een veel hogere of lagere beoordeling geven dan een andere beoordelaar. Volgens De Groot (1961) kunnen hieraan verschillende effecten ten grondslag liggen. Voorbeelden hiervan zijn het signifisch effect (veroorzaakt doordat beoordelaars verschillende ideeën hebben over wat goed en wat minder goed is) en het contaminatie-effect (veroorzaakt door de neiging van beoordelaars die belang hebben bij een bepaalde uitslag, om de beoordelingen aan de door hen gewenste uitslag aan te passen). Wanneer verschillende beoordelaars het niet met elkaar eens zijn, weten we niet wat we meten. Gevolg hiervan is dat de meting direct invalide is.

Een tweede probleem heeft te maken met de scorebetrouwbaarheid. In het empirisch onderzoek naar schrijfvaardigheden wordt in het gros van de gevallen uitgegaan van de aanname dat

Ulp van één schrijfo opdracht ook direct geldig zijn voor en alle soortgelijke opdrachten (Van Den Bergh & Meuffels, 2000). Volgens Van Schooten (1988) zijn echter niet alle schrijfo opdrachten gelijk waardoor ook de prestaties van schrijvers per schrijfo opdracht verschillen. Ook in diverse andere studies is de samenhang tussen verschillende schrijfo opdrachten nader bestudeerd, met als onveranderlijke conclusie dat de correlatie tussen prestaties op verschillende schrijfo opdrachten bedroevend laag is (Van Den Bergh & Meuffels, 2000). Prestaties op de ene taak hebben vrijwel geen voorspellende waarde voor de prestaties op een andere taak (Schoonen, 1991). Het gebrek aan samenhang tussen prestaties op verschillende schrijfo opdrachten vormt voor schrijfvaardigheidsonderzoekers dan ook een enorm probleem (Van Den Bergh & Meuffels, 2000).

Naast het onderwerp bepaalt bovendien de in de opdracht gevraagde communicatieve stijl de kwaliteit van een geschreven stuk (Van Schooten, 1988). Braddock e.a. (1963) noemen als verschillende zaken die de prestaties van schrijvers over taken doen verschillen niet alleen het onderwerp van een schrijfo opdracht, maar ook de communicatieve stijl van het te schrijven stuk, de toegestane schrijftijd en de situatie waaronder de toets wordt afgenomen. De schrijfvaardigheid die schrijvers krijgen toebedeeld is dus ook op die te beoordelen aspecten afhankelijk van de schrijftaak die zij uitvoeren.

Het gebruik van meerkeuzetoetsen voor het bepalen van schrijfvaardigheden is een oplossing voor de beoordelingsproblemen. Immers, een meerkeuzetoets is door iedereen gelijk te beoordelen en bovendien door één beoordelaar ook op een verschillend moment evenwichtig te beoordelen. De validiteit van meerkeuzetoetsen om schrijfvaardigheid te meten is over het algemeen matig tot redelijk (Van Schooten, 1988). De belangrijkste bezwaren zijn dat de meerkeuzetoetsen alleen receptieve en geen productieve vaardigheden kunnen meten en dat de beheersing van aspecten als stijl, organisatie en creativiteit niet goed met meerkeuzevragen gemeten kan worden (Van Schooten, 1988). Tegenover deze bezwaren kan gesteld worden dat de vaardigheden gemeten met meerkeuzetoetsen en schrijfo opdrachten weliswaar niet identiek, maar toch zo verwant zijn, dat meerkeuzetoetsen wel gebruikt kunnen worden om schrijfvaardigheid te meten (Van Schooten, 1988).

Om op een betrouwbare en valide manier schrijfvaardigheden te kunnen bepalen, proberen wij in dit paper een instrument te ontwikkelen waarmee de hierboven toegelichte beoordelingsproblematiek wat meer naar de achtergrond verschuift. Toch zullen de problemen met beoordelaars nooit volledig kunnen worden voorkomen. Schrijfo opdrachten en meerkeuzetoetsen om schrijfvaardigheid te meten lijken in grote mate gelijke vaardigheden te



*Your complimentary
use period has ended.
Thank you for using
PDF Complete.*

[Click Here to upgrade to
Unlimited Pages and Expanded Features](#)

de vaardigheden gemeten met open opdrachten en de
toetsen. Omdat de grootte van dit effect echter lijkt mee te
vallen (Van Schooten, 1988) en we de beoordelingsproblematiek van 'open toetsen' juist buiten
beschouwing willen houden, wordt het instrument waarmee we in ons onderzoek gaan werken
een gesloten toets.

illende onderdelen. Voorbeelden hiervan zijn spelling, werkwoordflexie, interpunctie, grammaticale bouw, structuur en correct taalgebruik (Meuffels, 1996). In dit artikel richten wij ons op het specifieke onderdeel taalgebruik.

Meuffels (1996) bespreekt verschillende toetsen voor het meten van deze onderdelen van schrijfvaardigheid. Op basis van een gedetailleerde foutenanalyse van betogen van studenten (waarvan de betrouwbaarheid en validiteit in verschillende empirische onderzoeken nader zijn beproefd) zijn deze toetsen tot stand gekomen. Hierbij geeft hij ook een overzicht van de meest voorkomende fouten op het gebied van taalgebruik. Drie types hiervan zijn niet-bestaande uitdrukkingen, niet-bestaande woorden, verkeerde voornaamwoorden en verkeerde verwijswwoorden.

Volgens Van Schooten (1988) bestaat het onderdeel taalgebruik uit de vragen waarbij de proefpersoon niet-grammaticale fouten in de vereiste woordreeks of klankvorm moet herkennen die niet onder de categorie spellingq vallen. Het gaat bij dit soort fouten om het niet correct hanteren van vaste uitdrukkingen. Onder deze categorie vallen fouten die betrekking hebben op verkeerd overgenomen uitdrukkingen (±Hij is van alle marktkramen thuisq), het toekennen van verkeerde betekenissen aan woorden en uitdrukkingen, contaminatie van verschillende zegswijzen (±Dat kost duurq) en het hanteren van verkeerde voorzetsels.

Op basis van deze literatuur hebben we in dit onderzoek gebruik gemaakt van drie verschillende foutieve vormen van taalgebruik: verhaspelingen van woorden, verhaspelingen van spreekwoorden en het onjuist gebruik van voornaam- en verwijswwoorden. In het vervolg van dit artikel zullen deze drie vormen worden aangehaald met de term *disciplines*.

2.1 Materiaal

Ons onderzoek is een instrumentatieonderzoek met als doel het ontwikkelen van een instrument om schrijfvaardigheden te kunnen bepalen. In het bijzonder hebben wij ons gericht op drie specifieke onderdelen van schrijfvaardigheid: spelling & interpunctie, tekststructuur en taalgebruik. Proefpersonen hebben zes verschillende opdrachten gemaakt. Enkel studenten hebben deelgenomen aan ons onderzoek.

Voor het onderdeel taalgebruik hebben de proefpersonen een door de proefleider samengesteld verhaal van drie alinea's gelezen. Binnen het onderdeel taalgebruik is er gebruik gemaakt van in totaal dertig items waarmee schrijfvaardigheid is gemeten. Deze items hebben we verdeeld onder de volgende drie disciplines: verhaspelingen van woorden (10), verhaspelingen van spreekwoorden (10) en foute voornaam- of verwijswwoorden (10). Deze dertig items hebben we willekeurig over de tekst verdeeld. Proefpersonen hebben de taak gekregen de tekst te lezen en te onderstrepen wat zij vonden dat fout was. Wanneer proefpersonen wat anders hebben onderstreept dan een door ons bedacht item, is dit als verfouting gecodeerd.

In bijlage 1 is de volledige afnametekst terug te vinden. In onderstaand kader staat een gedeelte van de gebruikte tekst:

Sara liep peinzend door de zaak. Dennis was ondertussen bezig met het **uitsorteren (i01)** van wat verfpotten. De twee kenden elkaar al **van kinds af aan (i02)** en ze waren al **meer als (i03)** zes jaar gelukkig samen. Twee jaar geleden trouwden ze en vertrokken ze een paar dagen later naar Griekenland voor hun huwelijksreis. Ondanks wat oponthoud op de **vluchthaven (i04)**, was het een fantastische vakantie geweest.

2.2 Pretest

Alvorens we de toets onder een grote groep respondenten hebben afgenomen, hebben we een pretest gedaan bij een kleinere groep. Dit waren in tegenstelling tot de uiteindelijke afnamegroep geen studenten van een eerstejaars cursus, maar twaalf master-studenten aan de Universiteit Utrecht.

Met de pretest wilden we testen of de toets consistent was. Om dit te kunnen bepalen stelden we drie eisen aan de resultaten:

gemiddelde. Wanneer de respondenten een item allemaal
delde 1 zijn. Dit zou kunnen betekenen dat dit item te
gemakkelijk was. Wanneer de respondenten een item allemaal fout zouden hebben dan zou het
gemiddelde 0 zijn. Dat item zou dan mogelijk te moeilijk zijn, omdat niemand het correct heeft. De
door alle deelnemers goed of fout beantwoorde items zeggen ons niets en daarom is het beter
wanneer het gemiddelde niet gelijk was aan 0 of 1.

De tweede eis gaat over de correlatie van de verschillende items tot de totaalscore op de toets.
Correlatie gaat over samenhang, en door deze te berekenen bekijk je in hoeverre een specifiek
item overeenkomt met de totaalscores. Deze correlatie moet positief zijn, want we verwachten
dat deelnemers die hoog scoren op een bepaald item, ook hoog scoren in hun totaalscore.
Zodoende willen we dat ze correlatie tussen een item en de totaalscores positief is.

De derde eis betreft de correlatie van de verschillende items tot de verfoutringen. De
verfoutringen zijn de door deelnemers onderstreepte woorden die niet fout zijn en dus helemaal
niet tot de items behoren. De eis is dat deze correlatie negatief is, we verwachten immers dat
deelnemers die goed zijn in een specifiek item, niet snel verfoutringen zullen maken.

Uit de pretest bleek allereerst dat zeventien van de dertig items aan alle drie de bovenstaande
eisen voldeden. Zeven items werden door alle proefpersonen terecht onderstreept. Omdat de
pretestgroep slechts twaalf proefpersonen betrof, is de kans groot dat alle deelnemers een
bepaalde fout zouden onderstrepen. Ook waren de respondenten in tegenstelling tot de
uiteindelijke afnamegroep geen eerstejaars studenten, maar master-studenten. Hierdoor
verwachtten we dat zij de test beter zouden maken dan de uiteindelijke respondenten. Om die
redenen besloten we niet alle items te veranderen, maar enkel de items waarvan van we bij
nader inzien dachten dat ze voor iedereen te gemakkelijk waren. Die twee items (i11 en i24)
veranderden we in een iets andere vorm waardoor het moeilijker werd deze te ontdekken. De test
bleef na de pretest dus grotendeels hetzelfde op twee kleine aanpassingen na.

2.3 Procedure

De deelnemers kregen individueel een boekje met opdrachten voorgelegd. De deelnemers
maakten deze opdrachten direct achter elkaar, zonder pauzes tussen de taken. De eerste twee
opdrachten die de proefpersonen maakten hadden te maken met het structureren van tekst, de
derde opdracht ging over spelling. De vierde opdracht ging vervolgens weer over tekststructuur.
Opdracht vijf toetste taalgebruik en de laatste opdracht ging tenslotte weer over spelling. Voor de
taken over tekststructuur gold een tijdslimiet.

deel aan het onderzoek. Al deze proefpersonen waren aalfilosofieqaan de Universiteit Utrecht.

Het onderzoek is afgenomen door drie verschillende proefleiders. Deze afnamen vonden plaats in drie verschillende groepen studenten. De proefpersonen maakten de opdrachten individueel en in het bijzijn van een proefleider. Het opdrachtenboekje bevatte duidelijke instructies. De proefleiders mochten de deelnemers niet helpen met de opdrachten. Proefpersonen konden tijdens de afname niet worden afgeleid door externe factoren als muziek of gezelschap van anderen. Het invullen van een vragenlijst duurde gemiddeld twintig minuten.

2.4 Essays

Voor ons onderzoek hadden we ook de beschikking over individuele essays van al onze respondenten. Deze essays beoordeelden de drie betrokken onderzoekers individueel op spelling & interpunctie, tekststructuur en taalgebruik. We hanteerden de kwaliteit van één essay als standaard (100) en de andere essays werden vervolgens aan de hand van dit standaardmodel beoordeeld. Wanneer een essay bijvoorbeeld door een onderzoeker als veel slechter dan het standaardmodel werd gezien, kreeg dit essay een lagere score dan 100. Andersom gold voor een beter essay dan de standaard dat het een score boven de 100 kreeg toebedeeld.

De essays zijn ook nog beoordeeld door de docent van de betreffende cursus waarin het essay werd ingeleverd. Dit gebeurde op een gebruikelijke schaal van 0 tot 10. Een goed essay kreeg een hoge score, een minder goed essay een lagere score. Zijn beoordelingen stonden verder volledig los van die van de andere drie beoordelaars.

3.1 Interne resultaten

Allereerst behandelen we de toets zelf. We kijken hierbij naar zowel de inhoudsvaliditeit als de constructvaliditeit. De inhoudsvaliditeit is van belang, omdat we willen weten in welke mate de toets het te meten domein bestrijkt. De constructvaliditeit is daarnaast van belang, omdat we ook benieuwd zijn naar de mate waarin het instrument het theoretische begrip meet dat het instrument beoogt te meten.

3.1.1 Inhoudsvaliditeit

Een schrijfvaardigheidstoets die alleen taalgebruik meet, is natuurlijk maar matig inhoudsvalide, aangezien schrijfvaardigheid meer behelst dan alleen taalgebruik. Toch proberen we met een goede constructvaliditeit een toets te ontwerpen waarmee in ieder geval het schrijfvaardigheids-onderdeel *taalgebruik* op een betrouwbare en valide wijze kan worden gemeten.

3.1.2 Constructvaliditeit

Om te bepalen of het gehanteerde instrument daadwerkelijk meet wat we willen meten, is het van belang te kijken naar de betrouwbaarheid van zowel de totale toets als de betrouwbaarheid van iedere afzonderlijke discipline. Een meetinstrument moet immers betrouwbaar meten om valide te kunnen zijn. In Tabel 1 zijn deze resultaten terug te vinden.

	Gemiddelde	Standaard-deviatie	Cronbach's Alpha (α)
Verhaspelingen van woorden	0.46	0.19	0.58
Verhaspelingen van spreekwoorden	0.68	0.18	0.62
Fouten in voornaam- of verwijzwoorden	0.39	0.15	0.69
Totaal	0.56	0.17	0.79

Tabel 1: Gemiddelden, standaarddeviaties en betrouwbaarheid per discipline en over de totaalscore op de toets

Wanneer we de vuistregel voor betrouwbaarheid hanteren (een resultaat is betrouwbaar wanneer de Cronbach's Alpha minimaal 0.70 is) is de totale toets betrouwbaar ($\alpha = 0.79$). Voor de afzonderlijke disciplines gelden betrouwbaarheidsscores die (net) onder deze grens scoren. De disciplines afzonderlijk zijn dus niet zo betrouwbaar als dat we eigenlijk zouden willen.

De disciplines niet dezelfde constructen meten, hebben we met elkaar correleren. Dit was met name interessant voor de correlatie tussen verhaspelingen van woorden en de verhaspelingen van spreekwoorden, omdat je zou kunnen verwachten dat deze disciplines veel met elkaar samenhangen. Tussen de verschillende disciplines gold echter geen enkele volledige correlatie. Zoals verwacht was de samenhang tussen verhaspelingen van woorden en verhaspelingen van spreekwoorden wel het grootst (0.61). Tussen verhaspelingen van woorden en fouten in voornaam- of verwijswaarden gold een veel lagere correlatie (0.28) en de correlatie tussen verhaspelingen van spreekwoorden en fouten in voornaam- of verwijswaarden lag daar tussenin (0.38). Essentieel is echter dat géén van de disciplines volledig met elkaar samenhangen en de verschillende disciplines dus verschillende constructen hebben gemeten.

Het beoordelen van de essays is gedaan door drie individuele beoordelaars. De essays kregen een gemiddelde beoordeling van 86,74 met een standaarddeviatie van 17,49. Om vervolgens te bepalen of deze beoordelingen valide zijn, is ook de samenhang tussen de beoordelaars berekend. Tussen deze drie beoordelaars gold een vrij hoge gemiddelde correlatie (0.64). Deze correlatie had een hoge betrouwbaarheidsscore ($r = 0.84$). De beoordelingen van beoordelaar I en III hingen het minst samen (0.62), terwijl de beoordelingen van beoordelaar II en III het meest correleerden (0.67). Beoordelaar I en II correleerden daar tussenin (0.63).

3.2 Externe resultaten

Nu we weten dat onze toets constructvalide is en de samenhang tussen de verschillende beoordelaars op de beoordeling van de essays betrouwbaar en valide is, kunnen we kijken naar de relatie tussen beiden. Immers, de meest interessant verwachte uitkomst van ons onderzoek is de samenhang tussen de scores op de essays en de scores op de gesloten toets. Wanneer we deze link zullen leggen, begeven we ons op het gebied van de criteriumvaliditeit. Dit is de mate waarin de scores op de toets correleren met een relevant criterium (in dit geval de essaybeoordelingen). Als de taalgebruiktoets criteriumvalide is en het criterium deugt, zal de correlatie tussen de twee groepen scores substantieel moeten zijn. In Tabel 2 staat een overzicht van de correlaties tussen de toets en de essayscores. Hierin is overigens ook de samenhang tussen de essaybeoordelingen van de docent van de cursus Φ aalfilosofieq en de essaybeoordelingen van de drie beoordelaars opgenomen.

	Correlatie
Verhaspelingen van woorden – Essay	0.20
Verhaspelingen van spreekwoorden – Essay	0.27
Fouten in voornaam- of verwijzwoorden – Essay	0.22
Essaybeoordeling docent - Essay	0.24
Totaal – Essay	0.29

Tabel 2: Correlaties tussen de essayscores en de scores op de toets (zowel op de totaalscore als de scores op de afzonderlijke disciplines).

De eerste discipline (verhaspelingen van woorden) correleert positief met de essayscore. Dit wil zeggen dat proefpersonen die een goed essay schrijven ook goed scoren op de discipline verhaspelingen van woorden en andersom. Deze positieve samenhang geldt ook voor de andere twee onderzochte disciplines. De hoogste correlatie is er tussen de totaalscore op de toets en de essayscore. Proefpersonen die de taalgebruiktoets goed maken, krijgen gemiddeld ook een hogere essaybeoordeling dan proefpersonen die de taalgebruiktoets minder goed maken. Hierbij moet wel vermeld worden dat zowel de afzonderlijke disciplines als de totaalscore op de toets niet erg hoog met de essaybeoordelingen correleren.

Aan de verschillende items werden in het methodehoofdstuk echter ook drie eisen gesteld. Voor vijf van de in totaal dertig items gold dat zij aan één of twee eisen niet voldeden. De correlaties met het criterium veranderden echter nauwelijks wanneer we enkel de items die aan de inhoudelijke eisen voldeden in onze berekeningen meenamen. Ook bij dit resultaat moeten we stellen dat de gevonden samenhang weliswaar positief is, maar dat de (onderdelen van de) toets en de essaybeoordelingen niet erg hoog correleren (0.28).

4.1 Conclusie

Met ons onderzoek hebben we geprobeerd een meetinstrument te ontwerpen waarmee we schrijfvaardigheid zonder de beoordelingsproblematiek konden bepalen. Dit instrument wilden we vervolgens toepassen in de praktijk. Allereerst was het om die reden van belang om te onderzoeken of het meetinstrument wel betrouwbaar en valide was. De betrouwbaarheid van de gehele taalgebruiktoets was voldoende, de betrouwbaarheid van de afzonderlijke disciplines waarmee taalgebruik werd gemeten (verhaspelingen van woorden, verhaspelingen van spreekwoorden en fouten in voornaam- of verwijzwoorden) scoorde iets onder de grens van de geschikte betrouwbaarheid. We kunnen daarmee stellen dat de toets enigszins betrouwbaar was.

Vervolgens is gekeken naar de verschillende constructen (disciplines) en hun onderlinge correlaties. De disciplines bleken niet hetzelfde construct te meten, waardoor we hebben kunnen concluderen dat de disciplines weliswaar samenhangen, maar allemaal met succes een ander aspect van taalgebruik hebben gemeten.

De essays zijn beoordeeld door drie afzonderlijke beoordelaars. Deze beoordelingen bleken erg betrouwbaar en correleerden hoog met elkaar. De gemiddelde beoordeling van de beoordelaars hing ook samen met die van de cursusdocent die de essays beoordeelde, deze samenhang was echter niet erg hoog.

Nadat we concludeerden dat onze toets constructvalide was en de samenhang tussen de verschillende beoordelaars op de beoordeling van de essays betrouwbaar en valide kon worden genoemd, konden we kijken naar een eventuele correlatie tussen beide scores. Hiervoor hebben we naar twee verschillende taalgebruiktoetsen gekeken: allereerst de toets bestaande uit alle items, maar ook naar de toets bestaande uit enkel de items die aan de drie vooraf gestelde eisen voldeden. Tussen de totaalscore op de taalgebruiktoets en de essaybeoordelingen van de beoordelaars was sprake van een positieve correlatie. Hierdoor lijkt het erop dat proefpersonen die hoog scoren op de taalgebruiktoets ook hoger scoren met hun essay. Andersom geldt dit natuurlijk ook: een hoog cijfer voor een essay hangt samen met een hoge score op de taalgebruiktoets. Ook tussen alle scores op de verschillende disciplines afzonderlijk en de essayscores werd een positieve samenhang gemeten. Deze correlaties, inclusief die van de totaalscore met het essay, waren echter niet erg hoog.

om veel mogelijk eventuele tekortkomingen uit te sluiten, zijn er wel een aantal onvolkomenheden te bedenken. De belangrijkste heeft te maken met het feit dat we in dit onderzoek schrijfvaardigheid hebben gemeten aan de hand van één specifiek aspect van schrijfvaardigheid. Het is dus weliswaar mogelijk uitspraken te doen over taalgebruik, maar om dit te overkoepelen naar schrijfvaardigheden in het geheel is niet terecht.

Een andere tekortkoming heeft te maken met de samenstelling van het materiaal. De proefleiders hebben zonder ervaring op het gebied van toetsontwerp zelf het gebruikte tekstmateriaal geschreven. Ondanks dat deze procedure altijd in overleg met een begeleidende docent is verricht, hebben zij hierbij mogelijk te weinig rekening gehouden met zaken als tekstlengte, tekstonderwerp en woordkeuze. Daarbij komt dat dertig items (waarvan tien per discipline) er misschien wat te weinig zijn om generaliserende uitspraken te doen over het meten van (aspecten van) taalgebruik door middel van een gesloten toets.

Een laatste eventuele tekortkoming heeft te maken met het feit dat de proefpersonen bij het maken van de taalgebruiktoets wisten dat ze deelnamen aan een onderzoek. Mogelijk hebben ze hierdoor met nadruk gelet op eventuele taalfouten met als gevolg dat de uitslag geforceerd is.

In de toekomst zal er onderzoek moeten blijven worden gedaan naar de mogelijkheden om schrijfvaardigheden op een betrouwbare en valide wijze te toetsen. Met behulp van meerkeuzetoetsen, of de zoals in dit onderzoek gehanteerde gesloten toetsen, worden de problemen die er bestaan met beoordelaars voor een belangrijk deel teruggedrongen. Het is zaak hierin in de toekomst nog meer te kunnen slagen, zodat schrijvers ook daadwerkelijk op hun schrijfvaardigheden worden beoordeeld.

Bergh, H. van den & Meuffels, B. (2000). *Schrijfvaardigheden en schrijfprocessen*. In: Braet, A. Taalbeheersing als communicatiewetenschap. Een overzicht van theorievorming, onderzoek en toepassingen. Bussum: Coutinho

Braddock, R. (1969). *English composition*. In: R.L. Ebel (Ed.), *Encyclopedia of Educational Research* 443-451. London: Macmillan.

Groot, A.D. de (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. Gravenhage: Mouton en Co.

Meuffels, B. (1996). *Diagnostische schrijfvaardigheidstoetsen. De constructie van objectief of semi-objectief te scoren diagnostische schrijfvaardigheidstoetsen voor schrijvers van universitair niveau*. Amsterdam: IFOTT.

Schoonen, R. (1991). *De evaluatie van taalvaardigheidsmetingen. Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs*. Amsterdam: Universiteit van Amsterdam.

Schooten, E. van (1988). *De constructie van een meerkeuzetoets voor het meten van schrijfvaardigheid*. (SCO-rapport 178) Amsterdam: Stichting Centrum voor Onderwijsonderzoek.

Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid*. SVO-reeks nr. 42, Staatsuitgeverij 's-Gravenhage.

gebruik (met aangegeven items)

Dennis was ondertussen bezig met het **uitsorteren (i01)** van wat verfpotten. De twee kenden elkaar al **van kinds af aan (i02)** en ze waren al **meer als (i03)** zes jaar gelukkig samen. Twee jaar geleden trouwden ze en vertrokken ze een paar dagen later naar Griekenland voor hun huwelijksreis. Ondanks wat oponthoud op de **vluchthaven (i04)**, was het een fantastische vakantie geweest. Dennis probeerde Sara de vliegtickets te **ontfrutselen (i05)**, maar de douane kon er de humor niet van inzien en Dennis moest drie uur lang verklaringen afleggen. Achteraf konden ze er allebei wel om lachen. Vlak voor hun reis had het pasgetrouwde stel een huwelijksfeest voor vrienden en familie gegeven. Het was een geweldige avond. Na afloop hadden de mensen het zelfs over het leukste feestje **wat (i06)** ze ooit hadden meegemaakt. Dennis had de bruiloftspartij zelf georganiseerd. Hij werkte tot voor kort als bedrijfsleider in een groot café in Haarlem en wist dus wel hoe hij de aanwezigen een **onvergeten (i07)** avond moest bezorgen. Een aantal weken geleden bekeken ze samen de foto's nog eens. "Eigenlijk zouden we binnenkort weer eens een feest moeten geven", zei Dennis. "Maar dan wel op een andere locatie. Kan dat niet in de aula bij jou op de basisschool?" Sara keek op. Daar had ze eigenlijk nog nooit aan gedacht. Angela en Daisy, haar collega's **waarmee (i08)** ze op woensdagen weleens ging joggen, hadden dat ook al eens gedaan. Als **hun (i09)** dat kunnen, kunnen wij dat ook, dacht Sara. **Ik zal morgen eens een balletje neerleggen (i10)**," zei ze tegen Dennis. **Het moet wel weer een geslaagd feest worden, anders zakken we door de mand (i11)**. Misschien een feest ter opening van het restaurant?" Dennis knikte instemmend.

Ze wilden samen een eetzaak beginnen en een aantal maanden geleden was het **schaap van de dam (i12)**: ze besloten een pand te kopen. Op **de lange duur (i13)** wilden ze zelfs dat het restaurant zou uitgroeien tot een hotel. Toch verliep niet alles op rolletjes en de verbouwing van het pand had nog heel wat **voeten in de waarde (i14)**. De bouw was nog maar in een vroeg **stadion (i15)** en het ging allesbehalve voortvarend. Zo moest er veel meer gerepareerd worden dan ze in eerste instantie dachten. Ook alle muren moesten **overnieuw (i16)** geleverd worden en Dennis wist eigenlijk nog niet hoe hij dat ging aanpakken. Sara **rook al nattigheid (i17)** toen ze het pand kochten, maar Dennis was vastbesloten de deal te sluiten. Voor Sara was de dag waarop **hun (i18)** het pand kochten echter langzamerhand al een **gifzwarte (i19)** dag aan het worden. Ze vond dat sommige ingrepen overbodig waren, maar haar man wilde alles in het gebouw veranderen. Hierdoor kreeg Sara een **dubbelstrijdig (i20)** gevoel over de gang van zaken en ze wist niet zeker of het wel zo'n verstandige beslissing was geweest om hun droom te gaan verwezenlijken.

Plotseling stond er buiten een man door het raam te kijken. Sara zag hem en zei tegen Dennis: "Daar heb je die man **waarmee (i21)** ik laatst over onze bouwplannen gesproken heb. Hij leek er wel verstand van te hebben. Zal ik vragen of hij binnen wil komen?" Dennis keek op en zei: "Denk je echt dat hij in zijn eentje meer weet **als (i22)** wij?" Maar Sara luisterde niet en liep naar de voordeur om de man binnen te laten. Zijn naam was Freddie en hij had nog wel wat goede tips voor de twee jonge ondernemers. "Die verbouwing gaat vrijwel zeker langer duren **als (i23)** jullie denken," zei Freddie. "**Als perkje bij paaltje komt (i24)** moet er veel meer gebeuren dan jullie verwachten." Sara zuchtte. Het opstarten van een restaurant was misschien wel het moeilijkste **dat (i25)** ze ooit had gedaan. Freddie had heel wat aan te merken op de verbouwingsplannen. Hij keek de mappen in waarin alle werkzaamheden stonden en keek verbaasd op. "Jullie hadden de boel wel wat beter mogen voorbereiden, ik lees het nu pas vier minuten en ik **erger me er nu al aan (i26)**", riep hij naar Dennis. "Is er dan iets **dat (i27)** niet klopt?" vroeg Dennis geïrriteerd. Sara keek haar man boos aan. Ze accepteerde niet dat Dennis zo reageerde en vond dat hij zich tegenover Freddie moest **verexcuseren (i28)**. "Luister nou maar naar Freddie, hij heeft al jarenlang een goedlopende broodjeszaak hier verderop. Hij weet echt wel **hoe de vork aan de steel (i29)** zit." Maar Frank weigerde te luisteren. "Een broodjeszaak? Dat is toch iets heel anders? Stop toch eens met die onzin, je **vergelijkt appels met eieren (i30)**. Bovendien kunnen we het prima zelf. Help je me even met het dragen van deze spullen?"

orden (1)

Discipline II – Verhaspelingen in spreekwoorden (2)

Discipline III – Fouten in voornaam- of verwijswwoorden (3)

Item	Correctie	Discipline
Item i01: uitsorteren	sorteren	1
Item i02: van kinds af aan	van jongs af aan	2
Item i03: meer als	meer dan	3
Item i04: vluchthaven	in deze context luchthaven	1
Item i05: ontfrutselen	ontfutselen	1
Item i06: wat	dat	3
Item i07: onvergeten	onvergetelijke	1
Item i08: waarmee	met wie	3
Item i09: hun	zij	3
Item i10: een balletje neerleggen	een balletje opgooien	2
Item i11: zakken we door de mand	vallen we door de mand	2
Item i12: het schaap van de dam	het hek van de dam	2
Item i13: op de lange duur	op de lange termijn	2
Item i14: voeten in de waarde	voeten in de aarde	2
Item i15: stadion	stadium	1
Item i16: overnieuw	opnieuw	1
Item i17: nattigheid ruiken	nattigheid voelen	2
Item i18: hun	zij	3
Item i19: gifzwarte	gitzwarte	1
Item i20: dubbelstrijdig	onbestaand woord	1
Item i21: waarmee	met wie	3
Item i22: als	dan	3
Item i23: als	dan	3
Item i24: perkje bij paaltje	puntje bij paaltje	2
Item i25: dat	dat	3
Item i26: erger me aan	irriteer me aan	1
Item i27: dat	dat	3
Item i28: verexcuseren	excuseren	1
Item i29: de vork aan de steel	de vork in de steel	2
Item i30: appels met eieren	appels met peren	2

[Click Here to upgrade to Unlimited Pages and Expanded Features](#)

is

Eis 1: Gemiddelde k 0 of 1; **Eis 2:** Correlatie item . totaalscore = positief; **Eis 3:** Correlatie items . verfouteringen = negatief

Voldoet aan geen enkele eis																			
Voldoet aan één of twee eisen	i02	i14	i15	i26	i28														
Voldoet aan alle eisen	i01	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i16	i17	i18	i19	i20	i21	i22
Voldoet aan alle eisen (vervolg)	i23	i24	i25	i27	i29	i30													