

de genredidactiek zijn, hebben wij vergeefs gezocht.

Easy reading is damn hard writing. Naar analogie van dit *bon mot* zouden wij tot slot willen opmerken dat onze vragen en kritiek de verdienste zijn van het expliciete en transparante karakter van *Genres in schoolvakken*. Ook willen we onze waardering uitspreken voor de aanzet die geleverd is met dit boek. Een Nederlands boek over taalgericht vakonderwijs vanuit een genre-optiek is belangrijk. Wereldwijd zijn de afgelopen decennia - vanuit diverse disciplines en gericht op alle onderwijsniveaus - de analytische en pedagogische implicaties van genre verkend. Begrip van genres biedt ons zicht op de manieren waarop wij in het onderwijs en in allerlei andere contexten communicatieve praktijken vormgeven en reguleren. Vanuit dit gezichtspunt vormen genres de 'central nexus of human sense-making' (Charles Bazerman in Bawarshi & Reiff (2012, p. xi). *Genres in schoolvakken* zet genre als centraal concept voor een taalgerichte didactiek in het voortgezet onderwijs in Nederland op de kaart.

KEES DE GLOPPER & JACQUELINE VAN KRUININGEN

LITERATUUR

- Bawarshi, A. S., & Reiff, M. J. (2010). *Genre. An introduction to history, theory, research, and pedagogy*. West Lafayette: Parlor Press / The WAC Clearinghouse.
- Bhatia, V. (1993). *Analysing genre: language use in professional settings*. London: Longman.
- Burns, A., & Joyce, H. (1991). *Teachers' voices 4: Staying learner-centered in a competency-based curriculum*. Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Fang, Z., & Schleppegrell, M.J. (2008). *Reading in secondary content areas: A language-based pedagogy*. Michigan: University of Michigan Press.
- Hajer, M., & Meestringa, T. (2009). *Handboek taalgericht vakonderwijs*. Bussum: Coutinho.
- Heesters, K. (2000). *Een wereld vol verhalen: ontwikkeling van verhaalstructuur bij 9- tot 13-jarige eerste- en tweedetaalsprekers van het Nederlands*. Groningen: Rijksuniversiteit Groningen (dissertatie).
- Martin, J., & Rose, D. (2008). *Genre relations: mapping culture*. London: Continuum.
- Miller, C. R. (1994). Genre as social action. In A. Freedman, & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 23-42). Bristol: Taylor and Francis.
- Rose, D., & Martin, J. (2012). *Learning to write, reading to learn. Genre, knowledge and pedagogy in the Sydney school*. London: Equinox.
- Russell, D. R. (1997). Rethinking genre in school and society: an activity theory analysis. *Written Communication*, 14(4), 504-554.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge (UK), New York: Cambridge University Press.

VERSCHENEN PROEFSCHRIFT

Hiske Feenstra (2014). *Assessing writing ability in primary education. On the evaluation of text quality and text complexity*. (Proefschrift). Enschede: Universiteit Twente. ISBN 978 90 365 3725 4, 200 blz.

De centrale onderzoeksvraag in het proefschrift van Hiske Feenstra luidt: *Op welke wijze kan de schrijfvaardigheid van beginnende schrijvers betrouwbaar en valide getoetst worden?*

Haar onderzoek speelt zich af in de context van het grootschalige Periodiek Peilingsonderzoek van het Onderwijsniveau (PPON), dat al enkele decennia het prestatieniveau in kaart brengt van leerlingen aan het einde van en halverwege het basisonderwijs, onder andere op het gebied van hun schrijfvaardigheid in de Nederlandse taal.

Om haar onderzoeksvraag te beantwoorden onderzoekt Feenstra in drie deelstudies verschillende methoden van schrijfvaardigheidstoetsing:

1. de beoordelingsschaal met voorbeeldopstellen, gecombineerd met een analytische vragenlijst;
2. de revisietoets, in meerkeuzevorm en in open vraagvorm;
3. de geautomatiseerde beoordeling van opstellen (Automated Essay Evaluation).

De eerste twee methoden omschrijft ze als gericht op het beoordelen van tekstkwaliteit, de derde als gericht op het evalueren van tekstcomplexiteit.

In het eerste hoofdstuk zet Feenstra de problematiek uiteen van de toetsing van schrijfvaardigheid. Een tekst schrijven bestaat uit een grote reeks activiteiten, te beginnen met mentale processen: het genereren, organiseren en structureren van ideeën, en

het vertalen van deze ideeën naar woorden. Daarna zijn motorische vaardigheden nodig om handgeschreven of getypte letters te produceren. Tenslotte is een schrijver constant bezig om de tot dan toe geproduceerde tekst te evalueren en zo nodig te reviseren. Deze gelaagdheid van het construct schrijfvaardigheid bemoeilijkt het ontwerpen van een toets die recht doet aan alle componenten van schrijfvaardigheid en vormt een potentiële bedreiging voor de validiteit van de toets. Daarnaast kunnen de onderdelen waaruit een schrijfvaardigheidsmeting bestaat, tot gevolg hebben dat een verschil in schrijfscores tussen twee leerlingen niet veroorzaakt wordt door een verschil in vaardigheid, maar door een verschil in schrijftaak, beoordelaar(s) of beoordelingsprocedure. Zowel de betrouwbaarheid als de validiteit van de toetsing worden hierdoor negatief beïnvloed.

Hoofdstuk twee brengt verslag uit van de eerste deelstudie, waarin de combinatie van een beoordelingsschaal met voorbeeldopstellen en een analytische vragenlijst is onderzocht. Verwacht werd dat het aanbieden van voorbeeldopstellen, ofwel "ankers", ervoor zou zorgen dat de overeenstemming tussen beoordelaars hoger zou zijn dan wanneer alleen de analytische beoordeling werd toegepast. Ook werd verwacht dat de toevoeging van voorbeeldopstellen zou bijdragen aan een hogere generaliseerbaarheid

van de schrijfscores, doordat bronnen van ongewenste variantie zoals de schrijftaak en de beoordelaar er minder effect op zouden hebben.

Feenstra gebruikte drie schrijftaken: een narratieve, een descriptieve/directieve en een argumentatieve, die eerder deel hadden uitgemaakt van de toetsing van schrijfvaardigheid in het kader van PPO. Bij elke taak golden dezelfde overkoepelende beoordelingscriteria: *inhoud*, *structuur* en *taalverzorging*. Het criterium *inhoud* had betrekking op de mate waarin eisen van de taak op het gebied van informatie en doel- en publiekgerichtheid in de tekst waren terug te vinden. Bij *structuur* ging het om de organisatie van de inhoud: de opbouw van de tekst als geheel, de samenhang tussen tekstgedeelten en zinnen, en de lay-out. *Taalverzorging* had betrekking op zinsbouw, spelling, interpunctie en stijl.

Iedere schrijftaak bevatte voor elk van de drie overkoepelende criteria een beoordelingsschaal met drie voorbeeldopstellen, die op de schaal de scores 85, 100 en 115 vertegenwoordigden, waarbij 100 als gemiddelde gold. De keuze van de voorbeeldopstellen werd toegelicht via een korte typering ervan in termen van de beoordelingscriteria. 1476 teksten van 620 leerlingen in de leeftijd van 8 tot 12 jaar werden beoordeeld door tenminste twee beoordelaars, in twee condities. De eerste conditie was de controleconditie met alleen analytische vragen; de tweede de experimentele conditie met beoordelingsschaal en analytische vragen.

De toevoeging van voorbeeldopstellen blijkt een significant hogere betrouwbaarheid op te leveren bij het beoordelingscriterium *structuur*; bij *inhoud* en *taalverzorging* is er geen verschil. Feenstra geeft als mogelijke verklaring hiervoor dat de beide laatste criteria eenvoudiger te meten zijn dan de eerste: de hoeveelheid vereiste inhoudselementen en taalverzorgingsfouten kan worden geteld,

maar het tellen van structurelementen is zinloos, omdat tekststructuur niet noodzakelijk aangegeven hoeft te worden via expliciete structurelementen. Daardoor zou de beoordelingsschaal met voorbeeldopstellen wel iets toevoegen aan beoordeling op *structuur*, en niet op *inhoud* en *taalverzorging*.

Een generaliseerbaarheidsstudie laat zien dat de toevoeging van de voorbeeldopstellen aan analytische vragen in principe bijdraagt aan een hogere generaliseerbaarheid van de verkregen schrijfscores. Maar om een generaliseerbaarheid van meer dan .70 te bereiken, zouden voor het criterium *inhoud* vijf schrijftaken beoordeeld moeten worden door vijf beoordelaars. Voor *structuur* en *taalverzorging* zou het gaan om resp. drie taken door drie beoordelaars en twee taken door twee beoordelaars.

De beoordelaars is gevraagd naar hun oordeel over de beide beoordelingsmethoden. Ze spraken geen sterke voorkeur uit voor de experimentele beoordelingsmethode, maar vonden de voorbeeldopstellen wel een nuttig referentiepunt.

In hoofdstuk drie beschrijft Feenstra de tweede deelstudie, waarin wordt onderzocht hoe een revisietoets op een valide en betrouwbare manier kan bijdragen aan een grootschalige toetsing van schrijfvaardigheid in het basisonderwijs. Tijdens het schrijven van een tekst, aldus Feenstra, zal een ervaren schrijver de tot dan toe geproduceerde tekst constant evalueren en waar nodig reviseren om zo de kwaliteit van de tekst te verhogen, met daarbij de beoogde lezer in gedachten. Het kunnen reviseren is daarmee een belangrijke vaardigheid om te verwerven en een relevante vaardigheid om te toetsen. In het type revisietoets dat zij onder de loep neemt, beantwoorden de leerlingen vragen over fouten in andermans teksten, waarbij de plaats van de fout is gemarkeerd; de fout hoeft dus niet door de

leerling zelf te worden opgespoord.

Feenstra heeft allereerst de validiteit van een bestaande, gestandaardiseerde meerkeuzetoets voor revisievaardigheid geëvalueerd, afkomstig uit de Entreetoetsen groep 5-7 van Cito. De inhoud van de toets is vergeleken met de revisie-activiteiten die beschreven zijn door Hayes e.a. (1987). Daaruit bleek dat slechts 18% van de items betrekking heeft op *Inhoud* (kwaliteit van de informatie, doel- en publiekgerichtheid) en 82% van de items op *Taalgebruik* (correctheid, begrijpelijkheid, passendheid). De prestaties van de leerlingen op de revisietoets correleerde Feenstra vervolgens aan hun prestaties op de in het vorige hoofdstuk beschreven schrijftaken; het ging hierbij dus om dezelfde 620 leerlingen tussen de 8 en 12 jaar. Prestaties op de revisietoets bleken significant hoger te correleren met de andere taalonderdelen van de Entreetoets (woordenschat en begrijpend lezen) dan met de prestaties op de drie schrijftaken. Zelfs met het onderdeel rekenen uit de Entreetoets correleerden de prestaties op de revisietoets nog hoger dan met de prestaties op de schrijftaken.

Dat de meerkeuzetoets nagenoeg alleen revisievaardigheid meet op het niveau van *Taalgebruik*, vindt Feenstra op zich geen doorslaggevend bezwaar, want de meeste beginnende schrijvers reviseren hun tekst vooral op dat niveau. Aan het gebrek aan *face validity* tilt ze zwaarder: de meerkeuzetoets vraagt niet van de leerling om daadwerkelijk een tekst te reviseren.

Daarom heeft ze op basis van een bestaande revisietoets in meerkeuzevorm een open variant ontworpen: een interlineaire toets waarin leerlingen in de tekst fouten moeten verbeteren door woorden of zinsdelen weg te halen, toe te voegen, te verplaatsen of te vervangen. Deze interlineaire toets is afgenomen bij 522 leerlingen uit groep 8; de resultaten zijn vergeleken met resultaten op de bestaande meerkeuzetoets. Een evalu-

atie van de validiteit van de open revisietoets wijst uit dat deze uitsluitend betrekking heeft op *Taalgebruik*, terwijl in de meerkeuzevariant in ieder geval nog 8% van de items betrekking heeft op doel- en publiekgerichtheid. Het was niet mogelijk dit type items om te zetten in de open vraagvorm. De psychometrische kwaliteiten van de beide toetsvormen bleken vrijwel gelijk; wel waren de items in open vraagvorm iets moeilijker dan de meerkeuze-items en was de betrouwbaarheid van de interlineaire toets tegen de verwachting in iets groter dan die van de meerkeuzetoets.

Feenstra concludeert dat de beide revisietoetsen een waardevolle aanvulling kunnen vormen op complete schrijfopdrachten in het kader van grootschalige toetsing van schrijfvaardigheid: de toetsen bestrijken dan het reviseren op het gebied van de formuleervaardigheid, de schrijfopdrachten ook het reviseren op het gebied van structuur, inhoud en organisatie van de tekst. Maar de revisietoetsen op zich zijn geen valide representatie van het gehele construct revisievaardigheid; ze toetsen alleen de mate waarin leerlingen de oppervlaktenmerken van de tekst kunnen verbeteren.

In hoofdstuk vier verkent Feenstra de mogelijkheden van een geautomatiseerde beoordeling van opstellen. Het doel van deze studie is het onderzoeken van de relatie tussen maten van tekstcomplexiteit en tekstkwaliteit, en de mogelijkheid om de maten van tekstcomplexiteit in te zetten binnen schrijfvaardigheidsmeting in het basisonderwijs. Ze maakt hiervoor gebruik van T-Scan, een programma voor geautomatiseerde complexiteitsanalyse van Nederlandse teksten. Dit programma meet zaken als woord- en zinscomplexiteit, informatiedichtheid, samenhang, concreetheid en mate van persoonlijk appèl aan de hand van kwantificeerbare linguïstische kenmerken van de tekst.

De teksten die leerlingen geschreven hadden in het kader van de argumentatieve schrijftaak uit hoofdstuk twee zijn geanalyseerd met behulp van T-Scan, en de uitkomsten van die analyse zijn vergeleken met de oordelen van twee beoordelaars met behulp van een beoordelingsschaal met voorbeeldopstellen. Hierbij ging het om 418 teksten.

De resultaten lieten allereerst zien dat de validiteit van verschillende maten van tekstcomplexiteit wordt beïnvloed door het feit dat schrijfproducten in het basisonderwijs van gebrekkige kwaliteit zijn. Spelfouten maakten het bijvoorbeeld het programma vaak onmogelijk om de juiste woordsoort van een woord te identificeren, en het ontbreken van interpunctie had invloed op het vaststellen van zinslengte en zinscomplexiteit. Het bleek dus nodig fouten in spelling, interpunctie en grammatica handmatig op te sporen en te verbeteren voordat het programma kon worden toegepast.

Om te bepalen welke kenmerken van tekstcomplexiteit indicatief zijn voor de schrijfvaardigheid van basisschoolleerlingen, heeft Feenstra eerst een kwantitatieve analyse uitgevoerd van de overeenstemming tussen 37 mogelijke maten uit T-Scan en twee maten van schrijfvaardigheid, namelijk leerjaar en opstelscore. Op grond daarvan selecteerde ze 13 maten van tekstcomplexiteit, waarop ze een kwalitatieve analyse uitvoerde om de interpreteerbaarheid en validiteit ervan als maten van schrijfvaardigheid te evalueren. Uit deze analyse bleek dat verschillende maten aangepast dienen te worden om ze specifiek geschikt te maken voor de evaluatie van de schrijfvaardigheid van beginnende schrijvers. Ook bleek dat verschillende factoren, zoals tekstlengte, de gemeten complexiteit van een tekst onrechtmatig beïnvloeden.

Feenstra sluit haar proefschrift af met aanbevelingen voor het toetsen van schrijfvaardigheid binnen grootschalig peilingsonderzoek

in het basisonderwijs. Wanneer alle drie de beschreven methoden daarin zouden worden toegepast, zou dat naar haar mening gunstige effecten hebben op zowel de betrouwbaarheid als de validiteit van de meting. De toevoeging van ankeropstellen aan een analytische beoordelingsprocedure blijkt immers een gunstig effect te hebben op de overeenstemming tussen beoordelaars, de validiteit van de schrijfscores en de mate waarin de schrijfscores generaliseerbaar zijn over verschillende taken en beoordelaars. Een revisietoets in open vraagvorm blijkt een valide methode om te meten in hoeverre schrijvers in staat zijn de kwaliteit van een tekst, qua oppervlaktekenmerken, te evalueren en te verbeteren. En de geautomatiseerde evaluatie van tekstcomplexiteit is een veelbelovende objectieve methode om tekstenmerken te evalueren die indicatief zijn voor de schrijfvaardigheid van beginnende schrijvers.

De resultaten van Feenstra's eerste deelstudie bevestigen wat uit onderzoek steevast naar voren komt: dat de combinatie van een schaal met voorbeeldopstellen en analytische vragen een valide en betrouwbare methode is om schrijfvaardigheid te beoordelen (zie recent in dit tijdschrift bijvoorbeeld Van Weijen, Van den Bergh en Rijlaarsdam, 2009; Pollmann, Prenger en De Glopper, 2012; Hoogeveen en Van Gelderen, 2014, en in dit nummer de bijdragen van Elving en Van den Bergh en Beeker et al.). Dat is een bevinding die niet alleen van belang is in de context van periodiek peilingsonderzoek in het basisonderwijs, maar ook in de context van landelijke toetsing en examinering, zoals de eindtoets basisonderwijs, de eindexamens Nederlands en een eventueel in te voeren diagnostische tussentijdse toets in het voortgezet onderwijs. Het praktijkprobleem van de geringe betrouwbaarheid van de toetsing van schrijfvaardigheid in deze context is in principe oplosbaar. Wat in de weg staat, is de majeure investering die gedaan

zou moeten worden in het ontwikkelen van beoordelingsschalen en het organiseren van een adequate beoordelingsprocedure; majeur omdat het gaat om een jaarlijkse toetsing van zeer grote hoeveelheden leerlingen. Dat deze investering niet gedaan wordt, is vooral een kwestie van politieke (on)wil.

De resultaten van de tweede deelstudie sluiten aan bij een discussie die ook recent in dit tijdschrift gevoerd is: kan een meerkeuzetoets, of kunnen gesloten toetsen in het algemeen bijdragen aan de toetsing van schrijfvaardigheid? (Pullens, Den Ouden, Herrlitz en Van den Bergh, 2013; De Glopper en Willemsen, 2014). Pullens e.a. vergeleken de scores op een meerkeuzetoets schrijfvaardigheid met een globale schaalbeoordeling van opstellen van dezelfde basisschoolleerlingen, en vonden geen positief verband, behalve bij spelling. Ze stellen mede op grond daarvan dat de meerkeuzetoets ongeschikt is voor de toetsing van schrijfvaardigheid. De Glopper en Willemsen halen studies aan die wel degelijk verband laten zien tussen meerkeuzetoetsen en open schrijfopdrachten, en voegen daar een eigen onderzoek aan toe dat eveneens een positief verband laat zien. Op grond daarvan typeren ze het standpunt van Pullens e.a. als eenzijdig. Feenstra geeft deze discussie een nieuwe, zinvolle wending door de bewuste meerkeuzetoetsen niet te beschouwen als toetsing van het gehele construct schrijfvaardigheid, maar alleen als toetsing van één aspect van schrijfvaardigheid, namelijk revisievaardigheid, die dan nog beperkt is tot formulevaardigheid. Maar ook dan blijft het vreemd dat er in haar studie zo weinig verband gevonden wordt tussen de prestaties van de leerlingen op de meerkeuzetoets en hun prestaties op complete schrijfopdrachten, een punt waar Feenstra helaas niet nader op ingaat. Dit gegeven lijkt een ondersteuning van het standpunt van Pullens e.a.

De derde deelstudie betreft iets volkomen nieuws: voor zover mij bekend is er in ons

land niet eerder gepubliceerd over de geautomatiseerde beoordeling van teksten in de context van het voortgezet onderwijs. Feenstra zal hierover dit jaar publiceren in *Levende Talen Tijdschrift*; reden dat ik de beschrijving van haar derde deelstudie globaal heb gehouden en nu ook afzie van commentaar op deze deelstudie.

Feenstra's aanbevelingen met betrekking tot periodiek peilingsonderzoek vind ik wat overhaast, ook al formuleert ze niet heel stellig. In de eerste plaats lijken drie metingen van schrijfvaardigheid in een peiling mij nogal veel van het goede, vooral wanneer deze weinig verband met elkaar zouden blijken te vertonen. In de tweede plaats zou opgehelderd moeten worden waarom er geen verband wordt gevonden tussen prestaties op een revisietoets en prestaties op een schrijfopdracht, voordat een revisietoets wordt ingevoerd. Ten derde lijkt invoering van een vorm van geautomatiseerde beoordeling nog ver weg, alleen al vanwege het probleem dat alle teksten daartoe moeten worden voorbereid.

Het proefschrift van Feenstra is al met al een interessante bijdrage aan de kennis van en de discussie over de beoordeling van schrijfvaardigheid.

HELGE BONSET

LITERATUUR

- Glopper, K. de, & Willemsen, A. (2014). Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid? *Levende Talen Tijdschrift*, 15(1), 31-39.
- Hayes, J. R., Flower, L. S., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (ed.), *Advances in Applied Psycholinguistics*, Vol. 2: *Reading, Writing and Language Learning*. Cambridge: Cambridge University Press, pp. 176-240.

- Hoogeveen, M., & Gelderen, A. van (2014). Effecten van leren schrijven met peer response en instructie in genrekennis. *Levende Talen Tijdschrift*, 15(2), 15-27.
- Pollmann, E., Prenger, J., & Glopper, K. de (2012). Het beoordelen van leerlingteksten met behulp van een schaalmodel. *Levende Talen Tijdschrift*, 13(3), 15-25.
- Pullens, T., Den Ouden, H., Herrlitz, W., & Bergh, H. van den (2013). Kan een meerkeuzetoets bijdragen aan het meten van schriftelijke taalvaardigheid? *Levende Talen Tijdschrift*, 14(2), 31-42.
- Weijen, D. van, Bergh, H. van den, & Rijlaarsdam, G. (2011). Het gebruik van T₁ tijdens schrijven in T₂. *Levende Talen Tijdschrift*, 12(2), 3-14.