

Boekbespreking

Henny Uiterwijk

Hoe het meten van schrijfvaardigheid goed voorbereid kan worden

Een bespreking van:

Rob Schoonen, *De evaluatie van schrijfvaardigheidsmetingen, een empirische studie naar de betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs*. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam (proefschrift), 1991.

1. Inleiding

Om aan de discussie over wat kinderen op de basisschool moeten leren en feitelijk leren een empirische basis te verschaffen is in 1986 in opdracht van het toenmalige Ministerie van Onderwijs en Wetenschappen het project Periodieke Peiling van het Onderwijsniveau (PPON) van start gegaan. Het project PPON, waarvan de uitvoering is ondergebracht bij het Cito, heeft inmiddels van vrijwel alle vak- en vormingsgebieden onderzocht wat de kinderen wordt onderwezen en wat de leerresultaten daarvan zijn. Het is de bedoeling om deze peilingen periodiek te herhalen, waardoor ook vergelijkingen in de tijd mogelijk zijn. Het project PPON zal naar verwachting in 1996 kunnen rapporteren over de ontwikkelingen in de vakgebieden Rekenen en Taal.

Aan de start van het project PPON is de Voorstudie PPON voorafgegaan, waarin de haalbaarheid van het peilen van het onderwijsaanbod en de leerresultaten van het vakgebied Taal in de achtste groep van het basisonderwijs onderzocht werd. De voorstudie werd in de periode 1983 - 1985 door de Stichting Centrum voor Onderwijsonderzoek (SCO) van de Universiteit van Amsterdam uitgevoerd.

De hier te bespreken dissertatie van Rob Schoonen, die voortvloeit uit de Voorstudie PPON, is het verslag van een empirisch onderzoek naar de betrouwbaarheid, validiteit en bruikbaarheid van verschillende operationalisaties van schrijfvaardigheid met het oog op het beantwoorden van de vraag hoe schrijfvaardigheid in periodieke taalpeilingen goed en efficiënt geoperationaliseerd kan worden. In deze boekbespreking wordt eerst ingegaan op het doel, de opzet en de resultaten van de studie, tot slot volgt een beoordeling.

2. Doel van de studie

In de Voorstudie PPON was geconcludeerd dat onderzoek gewenst was enerzijds naar de scoring van schrijfprestaties en anderzijds naar de verschillende operationalisaties van schrijfvaardigheid. De in de Voorstudie gehanteerde schrijfvaardigheidsopdrachten die de leerlingen betrekkelijk weinig structuur boden, lieten zien dat de leerlingen de opdrachten zeer verschillend interpreteerden en uitvoerden. De grote variatie aan schrijfproducten stelde ook hoge eisen aan de deskundigheid van de beoordelaars en vroeg veel inspanning om de betrouwbaarheid van de beoordelingen op een aanvaardbaar niveau te krijgen.

Het ligt voor de hand om te veronderstellen dat opdrachten die meer sturing aan de leerlingen geven, minder variatie tussen de schrijvers laten zien en eenvoudiger te beoordelen zijn. Het is echter niet zeker dat gestructureerdere opdrachten inzake betrouwbaarheid en validiteit te beschouwen zijn als gelijkwaardige operationalisaties van schrijfvaardigheid. Schrijfvaardigheid wordt in deze studie gedefinieerd als de mate waarin iemand in staat is een tekst te produceren die de door hem/haar bedoelde boodschap adequaat verwoordt op een wijze die recht doet aan de communicatieve context en de gebezigde schrijftaal.

Als operationalisatie van schrijfvaardigheid worden in deze studie vijf opdrachtvormen gehanteerd, die oplopen in de mate waarin de leerling structuur krijgt aangeboden.

- Bepaalde opdrachten, waarbij de schrijver alleen een communicatieve context (schrijfdoel, beoogde lezer) krijgt aangereikt.
- Gestructureerde opdrachten, waarbij de schrijver een context krijgt aangereikt en enkele sturende vragen als hulp bij het genereren en structureren van de tekst.
- Correctie-opdrachten zonder markering, waarbij de schrijver een tekst krijgt aangereikt die aan de schrijfp opdracht voldoet maar die nog gereviseerd moet worden.
- Correctie-opdrachten met markering, waarbij de schrijver een tekst krijgt aangereikt die aan de schrijfp opdracht voldoet maar die op reeds gemarkeerde plaatsen revisie behoeft.
- Gesloten opdrachten, waarbij de schrijver bij tekstfragmenten een aantal alternatieven met betrekking tot de inhoud of het taalgebruik krijgt aangeboden waaruit het beste alternatief gekozen moet worden.

De beoordelingsmethoden verschillen per operationalisatie en verondersteld wordt dat de deskundigheid van de beoordelaar bij sturender opdrachtvormen en striktere beoordelingsmethoden van minder belang wordt.

In het proefschrift over de betrouwbaarheid, validiteit en bruikbaarheid van de vijf operationalisaties van schrijfvaardigheid wordt getracht de volgende vragen te beantwoorden:

- Leiden de sturender opdrachtvormen en beoordelingsmethoden tot betrouwbaardere beoordelingen?

- Zijn lekenbeoordelaars zonder betrouwbaarheids- en/of validiteitsverlies inzetbaar bij de beoordeling van de schrijfprestaties die verzameld worden met de diverse operationalisaties van schrijfvaardigheid?
- Zijn de operationalisaties van schrijfvaardigheid gelijk in hun criteriumvaliditeit als men het docentenoordeel als criterium hanteert respectievelijk als men de bias van het instrument onderzoekt?
- Zijn de operationalisaties van schrijfvaardigheid gelijk in hun constructvaliditeit als men de relaties tussen de operationalisaties onderling beschrijft, danwel die tussen de operationalisaties en intellectuele vaardigheden?

3. Opzet van de studie

3.1. Instrumentontwikkeling

Voor de vijf operationalisaties van schrijfvaardigheid werden opdrachten geconstrueerd in zoverre ze niet beschikbaar waren uit de Voorstudie PPON. Er zijn vier communicatieve situaties onderscheiden:

- mensen overtuigen;
- beschrijven van concrete zaken;
- uitleggen van een spel en;
- geven van een routebeschrijving.

Op basis van elke communicatieve situatie zijn de eerder genoemde vijf opdrachtvormen ontwikkeld. Dit is gedaan om te voorkomen dat zogenaamde taakeffecten een rol gaan spelen. Een taakeffect treedt op wanneer de correlatie tussen de deelaspecten van een schrijfvaardigheidso opdracht binnen dezelfde communicatieve situatie hoger is, dan de correlatie van hetzelfde deelaspect over verschillende communicatieve situaties. In dat geval wordt een leerlingsscore meer beïnvloed door de 'toevallige' communicatieve situatie en het thema dat daarin in aan de orde wordt gesteld (in het vervolg: het onderwerp), dan door de schrijfvaardigheid zelf. Om de constructvaliditeit van deze 5 x 4 instrumenten te onderzoeken is ook gebruik gemaakt van een reeds bestaande operationalisatie van het 'Structure of Intellect'-model van Guilford in zoverre die betrekking heeft op semantische vaardigheden.

De schrijfproducten van de leerlingen zijn beoordeeld op twee aspecten: Inhoud & organisatie en Taalgebruik. Bij Inhoud & organisatie ligt het accent op de kwaliteit van de boodschap die overgebracht wordt, bij Taalgebruik op de vorm van de taal met accent op de kwaliteit van het gebruikte idioom, op de gehanteerde grammaticale regels en op de mate waarin het taalgebruik effectief, passend en aantrekkelijk is. Bij de beoordelingen blijft de spelling en interpunctie buiten beschouwing.

De beoordelaars maken bij de bepaalde en gestructureerde opdrachten gebruik van reeksen 'geschaalde voorbeeldopstellen en van nauwkeurige scoringsvoor-schriften. De correctie-opdrachten met en zonder markering worden met scorings-

voorschriften beoordeeld en de gesloten opdrachten zijn geprecodeerd, waardoor objectieve verwerking mogelijk is. De scoringsvoorschriften zijn, in zoverre ze niet beschikbaar waren uit de Voorstudie PPON, in deze studie ontwikkeld.

De kwaliteit van de in deze studie ontwikkelde opdrachtvormen en beoordelingsinstrumenten is in een vooronderzoek beproefd en waar nodig aangepast.

3.2. Afname en scoring

Ten behoeve van het validatie-onderzoek hebben 22 basisscholen met in totaal 442 leerlingen meegedaan aan de afname van de schrijfvaardigheidstoetsen. De leerlingen zijn hiervoor in vijf groepen verdeeld. Vier groepen leerlingen maakten elk van één onderwerp alle vijf operationalisaties van schrijfvaardigheid. Zij maakten telkens eerst de minst sturende opdrachtvorm, gevolgd door steeds gestructureerdere vormen. De vijfde groep leerlingen maakte als controlegroep van alle vier onderwerpen alleen de minst gestructureerde opdrachtvorm (bepaalde opdrachten). Alle vijf groepen maakten de gesloten opdrachten en de 'Structure of Intellect'-toets (SI-toets).

Afgezien van de gesloten opdrachten en de SI-toets werden alle schrijfopdrachten op Inhoud & organisatie en Taalgebruik geëvalueerd door deskundige beoordelaars. De bepaalde en gestructureerde opdrachten werden per onderwerp door een jury van vijf personen op de aspecten Inhoud & organisatie of op Taalgebruik beoordeeld. Deze twee opdrachtvormen werden per onderwerp zowel met geschaalde voorbeeldopstellen als met scoringsvoorschriften beoordeeld. De correctie-opdrachten met en zonder markering werden per onderwerp door een jury van drie personen op de aspecten Inhoud & organisatie of op Taalgebruik beoordeeld.

Van elke beoordelingseenheid waren de schrijfprestaties van 80 - 90 leerlingen beschikbaar. De gesloten opdrachten en de SI-toetsen vallen buiten de beoordeling van jury's, omdat ze objectief gescoord werden.

4. Resultaten

Met de data die in het validatie-onderzoek waren verzameld, is nagegaan of de sturender opdrachtvormen en beoordelingsmethoden tot betrouwbaardere beoordelingen leiden. Er kan gesteld worden dat bij de beoordeling van teksten van bepaalde en gestructureerde opdrachten het effect van de opdrachtvorm niet of nauwelijks aanwezig is. Wel is gebleken dat scoringsvoorschriften betrouwbaardere beoordelingen opleveren dan geschaalde voorbeeldopstellen. Over het algemeen zijn beoordelingen over Inhoud & organisatie weer betrouwbaarder dan oordelen over Taalgebruik. Bij de correctie-opdrachten zijn de effecten tegengesteld aan de verwachtingen: de meest sturende correctie-opdracht wordt het minst betrouwbaar

beoordeeld. Hier wordt Taalgebruik betrouwbaarder beoordeeld dan Inhoud & organisatie.

Al in het vooronderzoek werd nagegaan of lekenbeoordelaars zonder betrouwbaarheids- en/of validiteitsverlies inzetbaar zijn voor de beoordeling van schrijfprestaties die met verschillende opdrachtvormen zijn verzameld. De inzet van lekenbeoordelaars is immers goedkoper dan die van deskundige beoordelaars. Gebleken is dat leken en deskundigen over het algemeen niet inwisselbaar zijn. Bij de minst sturende opdrachtvorm beoordeeld met geschaalde voorbeeldopstellen waren de verschillen in betrouwbaarheid tussen de soorten jury's het grootst in het voordeel van de deskundigen met name bij het aspect Taalgebruik. Bij de beoordeling van de gestructureerde opdrachten met scoringsvoorschriften en bij correctie-opdrachten met en zonder markering kunnen net zo goed leken ingezet worden als deskundigen.

Uit het validatie-onderzoek blijkt dat de vijf operationalisaties van schrijfvaardigheid positief maar niet hoog correleren (tussen .30 en .50) met het docentenoordeel over de schrijfvaardigheid van de leerlingen. Opmerkelijk is dat de meest en de minst gestructureerde opdrachtvormen gerelateerd aan het docentenoordeel het meest valide zijn. Verder zijn de geschaalde voorbeeldopstellen valider dan beoordelingen met scoringsvoorschriften.

Eveneens is nagegaan of de vijf operationalisaties van schrijfvaardigheid bepaalde subgroepen in de populatie bevoor- of benadelen. Als subgroepen gelden jongens en meisjes en hoog- en laagscorende leerlingen. Gebleken is dat de vijf operationalisaties onpartijdig zijn voor jongens en meisjes, maar niet voor hoog- en laagscorende leerlingen. De verschillen tussen de gemiddelde vaardigheid van hoog- en laagscorende leerlingen verkleinen naarmate de operationalisaties sturender worden. Het wijst erop dat goede schrijvers minder en zwakkere schrijvers meer profiteren van gestructureerdere opdrachtvormen.

Tevens is onderzocht hoe de relaties tussen de onderscheiden operationalisaties van schrijfvaardigheid te beschrijven zijn en hoe de relatie is tussen deze operationalisaties en enkele intellectuele (SI-toets) vaardigheden. Uit de analyses is gebleken dat de vijf schrijfvaardigheidsmetingen geen beroep doen op één latente vaardigheid. Er zijn extra vaardigheden in het geding met name bij de bepaalde en gestructureerde opdrachten. Deze extra vaardigheden, die aan te duiden zijn met het kunnen structureren en generen, komen sterker bij Inhoud & organisatie naar voren dan bij Taalgebruik. Verder is gebleken dat de vijf operationalisaties een verschillend beroep doen op de intellectuele vaardigheden. Er zijn geen intellectuele vaardigheden waaraan alle operationalisaties even sterk appelleren. Voor Inhoud & organisatie geldt dat de bepaalde opdrachten op meer intellectuele vaardigheden een beroep doen, dan de sturendere opdrachtvormen. Bij Taalgebruik is er geen structuur te ontdekken in de relaties tussen intellectuele vaardigheden en de operationalisaties van schrijfvaardigheid.

Met betrekking tot bruikbaarheid van de instrumenten wordt opgemerkt dat de gestructureerde opdrachten en correctie-opdrachten moeilijker zijn te ontwikkelen dan bepaalde en gesloten opdrachten en datzelfde geldt ook voor geschaalde voorbeeldopstellen ten opzichte van scoringsvoorschriften.

Op basis van deze studie kan ten behoeve van periodiek taalpeilingsonderzoek een keuze gemaakt worden uit de onderscheiden opdrachtvormen, beoordelingsmethoden en typen beoordelaars. De auteur stelt dat de betrouwbaarheid en de validiteit van de metingen belangrijker zijn dan de praktische bruikbaarheid. Er wordt gesteld dat bij beoordelingsmethode de voorkeur uitgaat naar geschaalde voorbeeldopstellen boven de scoringsvoorschriften. Bij de keuze van de opdrachtvorm wordt onderscheid gemaakt naar het aspect dat men wil operationaliseren. Voor Inhoud & organisatie kan men het beste kiezen voor de bepaalde opdrachten en voor Taalgebruik geldt dat de gesloten opdracht en de correctie-opdracht bruikbaar zijn dan de vrijere opdrachtvormen. Over het algemeen geldt dat het beste deskundige beoordelaars ingezet kunnen worden. Lekenbeoordelaars kunnen aangezocht worden om de bepaalde opdrachten met geschaalde voorbeeldopstellen te beoordelen op het aspect Inhoud & organisatie. Aan het slot van zijn studie concludeert de schrijver dat de door hem getrokken conclusies voor een belangrijk deel overeenstemmen met de praktijk van de eerste taalpeiling van PPON in groep acht: de schrijfvaardigheid wordt met bepaalde opdrachten gemeten en de schrijfprestaties worden op het communicatieve aspect beoordeeld. De aspecten die betrekking hebben op Taalgebruik worden met meer gesloten opdrachtvormen beoordeeld.

5. Beoordeling

Het is verheugend te constateren dat de meting van taalvaardigheid voor een nieuw fenomeen als periodiek peilingsonderzoek zorgvuldig is voorbereid. Dit geldt in het bijzonder voor het evalueren van de betrouwbaarheid en validiteit van schrijfvaardigheidsmetingen, omdat dit ook in het kader van periodieke peilingen een veelzijdige en weerbarstige materie is. Uit de opzet van zijn onderzoek blijkt dat de auteur zich hiervan bewust is, want hij besteedt aandacht aan vrijwel alle aspecten die bij dergelijke keuzes een rol spelen. Vooral het validiteitsonderzoek wordt zorgvuldig aangepakt, hetgeen blijkt uit het feit dat de validiteit van de schrijfvaardigheidsmetingen op vier manieren is onderzocht. De analyse van de onderzoeksresultaten gebeurt op adequate wijze en de resultaten worden zorgvuldig en gedetailleerd gerapporteerd. Er wordt niet nagelaten om te wijzen op zwakke plekken in de onderzoeksopzet en moeilijk interpreteerbare uitkomsten worden niet verbloemd. De studie wordt gekenmerkt door degelijkheid en veelzijdigheid. De kritische kanttekeningen die bij de studie worden geplaatst, hebben dan ook betrekking op ondergeschikte punten.

In de studie heeft Taalgebruik betrekking op de vorm van de taal met het accent op de kwaliteit van het gebruikte idioom, op de gehanteerde grammaticale regels en op de mate waarin het taalgebruik effectief, passend en aantrekkelijk is. De spelling en de interpunctie behoren niet tot Taalgebruik. Bij de beoordeling van de schrijfproducten wordt de regel gehanteerd dat schrijffouten die ook in de gesproken taal fout zouden zijn onder Taalgebruik vallen en dat schrijffouten die in

de gesproken taal niet hoorbaar zijn onder Spelling vallen en dus buiten beschouwing gelaten worden. De schrijver geeft wel aan dat spelling en grammatica niet altijd haarscherp te onderscheiden zijn, maar er wordt niet ingegaan op de vraag waarom spelling en interpunctie buiten beschouwing gelaten worden en wat hiervan mogelijk de gevolgen zijn voor de procedure waarmee de schrijfproducten van leerlingen beoordeeld worden.

Volgens het afnamesdesign ontvangen de leerlingen vier schrijfopdrachten die oplopen in de mate van gestructureerdheid. Deze opdrachten gaan echter steeds over hetzelfde onderwerp, bijvoorbeeld over het uitleggen van de route die je moet volgen om jouw huis vanaf het station te bereiken. Deze aanpak wordt verdedigd door te stellen dat als er wederzijdse beïnvloeding zou zijn van de verschillende opdrachtvormen, dat dan verwacht mag worden dat die invloed eerder van de minst open naar de opener opdrachtvormen gaat dan omgekeerd. Als een leerling de vrijere opdrachtvorm nog tegoed zou hebben, bestaat er de kans dat hij leert van de gestructureerdere opdracht en dat hij deze 'extra' kennis (bijvoorbeeld over de opbouw van de tekst en over het onderwerp dat daarin aan de orde wordt gesteld) kan inzetten bij de vrijere opdrachten. Om elk effect, dat hetzelfde onderwerp op de resultaten van de verschillende opdrachtvormen kan hebben, uit te sluiten is het echter beter dat leerlingen a-select toegewezen worden aan de opdrachtvormen en onderwerpen. Uiteraard leidt dit tot een ingewikkelder afnamesdesign en wellicht tot de behoefte aan meer proefpersonen, maar het is dan wel mogelijk om de effecten van de opdrachtvormen zuiverder te schatten.

De studie richt zich ook op de bruikbaarheid van de verschillende operationalisaties van schrijfvaardigheid. De auteur stelt dat de bruikbaarheid slechts globaal beoordeeld wordt. Er wordt ingegaan op de inspanningen die geleverd moeten worden om de schrijfopdrachten en de beoordelingsmethoden te ontwikkelen, aan de tijd en de soorten proefleiders die nodig zijn voor de toetsafname en aan de tijd en de soorten beoordelaars ten behoeve van de beoordeling van de schrijfproducten. Bij de bruikbaarheid van bepaalde operationalisaties zijn wellicht ook andere aspecten belangrijk, zoals bijvoorbeeld de eisen die gesteld worden door het feit dat schrijfvaardigheidsmetingen herhaald moeten worden. Bij de uiteindelijke keuze voor de te hanteren operationalisatie blijkt dat de auteur dat de bruikbaarheid een aanzienlijk minder belangrijk criterium vindt dan de betrouwbaarheid en de validiteit van de metingen. De bruikbaarheid komt er in de studie zo bekaaid af en is vergeleken met validiteit en betrouwbaarheid dermate van een andere orde dat het wellicht beter zou zijn geweest dit onderdeel uit het proefschrift te schrappen.

Amsterdam: SCO, 1996.

A. Veen, A. Ouyman & K. de Groot. Eenheidsprijs Invalide. Onderzoek naar een project leerbeoordeling voor 0- tot 6-jarigen. Utrecht: SCO, 1996.

O. de Vries, Literatuuronderwijs als validatie feit. Inleiding voor het lezer lezen van literatuur op 1-100. Amsterdam: Stichting Amsterdamse Literatuur, 1996.

