

De vorm van tekstbegripvragen en schrijfopdrachten

Een bespreking van: H. van den Bergh. Examens geëxamineerd. 's-Gravenhage, SVO (Selectareeks), 1988, dissertatie Universiteit van Amsterdam

1. Inleiding

In 1986 en 1987 werden twee belangrijke veranderingen ingevoerd in de examens Nederlands voor de MAVO en het LBO. Eerst werden in 1986 functionele schrijfopdrachten ingevoerd, aanvankelijk verplicht voor het LBO en naar keuze voor de MAVO. Nu kiest op beide schooltypen een kandidaat zelf de soort schrijfopdracht. Een jaar later werden de verschillende examens Nederlands tekstbegrip voor het lager beroepsonderwijs en voor de MAVO geharmoniseerd. Beide schooltypen kregen hetzelfde tekstbegripexamen, dat voor 70% bestond uit gesloten vragen en voor 30% uit open vragen. De meeste LBO-scholen waren gewend aan twee examenonderdelen voor tekstbegrip, één met gesloten en één met open vragen; de MAVO-kandidaten kenden tot op dat moment maar één tekstbegripexamen, en dat bestond uitsluitend uit open vragen (Van den Bergh & Sprengers, 1988).

Op voorstel van het Cito werd een onafhankelijk instituut belast met de evaluatie van deze veranderingen. Dit werd de SCO. Het proefschrift van Huub van den Bergh bestaat uit de artikelen waarin hij achtereenvolgens verslag heeft gedaan van deze door hem uitgevoerde evaluatie. In enkele andere tijdschriften is men al ingegaan op de onderzoekstechnische en statistische aspecten van zijn onderzoek. Mij gaat het hier vooral om algemene consequenties voor de examenproductie. Het feit dat er ondertussen enige tijd verstreken is sinds de resultaten van het onderzoek werden gepubliceerd, maakt het des te makkelijker aan te geven welke invloed ze al gehad hebben en wat er nog te doen staat.

In deze recensie geef ik eerst een korte samenvatting van zijn werkwijze en uitkomsten, vervolgens voorzie ik enkele belangrijke punten van commentaar, en tenslotte probeer ik het belang van de dissertatie duidelijk te maken¹.

2. Samenvatting

Van den Bergh stelt twee vragen:

- Doet het beantwoorden van open vragen en gesloten vragen tekstbegrip een beroep op dezelfde intellectuele vaardigheden?
- Wordt bij het maken van functionele schrijfvaardigheidsoopdrachten een beroep gedaan op hetzelfde complex van intellectuele vaardigheden als bij het schrijven van een opstel?

Kort gezegd komt de werkwijze van Van den Bergh erop neer, dat hij zich niet beperkt tot een onderlinge vergelijking van verschillende vraagvormen en opdrachten, maar dat hij de relevante vraagvormen en opdrachten vergelijkt met relevante algemene intellectuele vaardigheden. Om deze vergelijkingen uit te

kunnen voeren, heeft hij allereerst gezorgd voor een aantal betrouwbare en goed beoordeelbare toetsen.

Voor tekstbegrip heeft hij twee LBO-examenteksten elk uitgebreid met een aantal vragen. Hij zorgde ervoor dat hij dezelfde soorten vragen in deze nieuwe toets stelde als in de oude examens en ook in dezelfde verhoudingen. Hiervoor had hij eerst de vragen uit 19 (!) oude examens door docenten laten indelen volgens twee veelgebruikte beschrijvingsschema's. De nieuwe toets werd bovendien gepretest.

Voor de toetsing van de functionele schrijfvaardigheid heeft Van den Bergh bestaande functionele opdrachten geselecteerd uit toetsen die de SCO vaker heeft gebruikt, terwijl hij voor het schrijven van een opstel gekozen heeft voor een stelopdracht bij één van tien eindexamentitels (uit 1984) en een stelopdracht bij vijf pagina's citaten en meningen over het onderwerp 'de zin en onzin van huiswerk'. Voor de schrijfopdrachten stelde hij verschillende beoordelingsvoorschriften op, variërend van gedetailleerde (analytische) scoringsvoorschriften, tot beoordelingsschalen en een globale beoordeling door docenten. Voor de meeste opdrachten stelde hij meer dan één beoordelingsvoorschrift op. Daarnaast heeft Van den Bergh ook nog een efficiënte methode ontwikkeld om beoordelaars met elkaar te kunnen vergelijken zonder dat ze allemaal alle schrijfprodukten hoeven te corrigeren.

Naast de hier beschreven toetsen van tekstbegrip en schrijfvaardigheid waren er ook toetsen nodig om algemene intellectuele vaardigheden te meten. Van den Bergh heeft veel aandacht besteed aan de keuze van de relevante intellectuele vaardigheden en aan de toetsen die de beheersing daarvan moesten meten. Uit het model van Guilford (het zogenaamde Structure of Intellect model, zie Meuffels 1981) heeft hij zestien combinaties van operaties en semantische produkten gekozen. De 16 combinaties heeft hij geselecteerd op grond van het feit dat ze blijkens de literatuur van belang worden geacht voor lezen of schrijven, omdat hij ze zelf relevant achtte op grond van zijn eigen analyse van de tekstbegripvragen of omdat ze naar voren kwamen uit een analyse van hardop-schrijfprotocollen. Bij elk van deze zestien algemene vaardigheden heeft hij drie typisch talige toetsen geconstrueerd. Bijvoorbeeld voor de cognitie van semantische transformaties de toetsen 'andere betekenissen', 'dubbelzinnige zinnen' en 'woordplaatjes vertalen'. Na een pretest heeft hij één toets en verschillende minder geslaagde items verwijderd. Vervolgens heeft hij nagegaan of de afnamegegevens beter beschreven konden worden met het model van Guilford dan door andere modellen voor intellectuele vaardigheden. Dit bleek inderdaad zo te zijn. Hij nam daarom aan dat de samenhang tussen deze toetsen redelijk verklaard wordt door het model van Guilford.

Alle toetsen zijn afgenomen aan een grote groep leerlingen uit de derde klas LBO of MAVO. De open opdrachten werden meestal volgens twee beoordelingsvoorschriften beoordeeld en door meer dan één corrector. Vervolgens zijn verschillende analyses uitgevoerd om na te gaan waardoor verschillen in de scores op de diverse opdrachten veroorzaakt worden.

Een onderlinge vergelijking van open en gesloten tekstbegripvragen wees uit dat gesloten vragen tot hogere scores leiden dan open vragen met dezelfde inhoud, vooral bij moeilijker vragen. Voor LBO-leerlingen waren bijna alle vragen moeilijker dan voor MAVO-leerlingen. Bij LBO-leerlingen leverde de

gesloten vorm dan ook meer extra punten op dan bij MAVO-leerlingen. Van den Bergh vermoedt dat het voordeel van de gesloten vragen veroorzaakt wordt, doordat ze leerlingen die niet zeker zijn van het goede antwoord, de mogelijkheid bieden te raden en LBO-leerlingen door hun lager niveau van leesvaardigheid eerder genoodzaakt zijn te raden.

Uiteindelijk ging het Van den Bergh om een vergelijking van de scores op de open en gesloten vragen met de scores op de toetsen van de zestien verschillende intellectuele vaardigheden. De scores op de open vragen bleken inderdaad een andere samenhang te vertonen met deze intellectuele vaardigheden dan de scores op de gesloten vragen. Om de betekenis van dit verschil in samenhang duidelijk te maken heeft Van den Bergh ook de scores met elkaar vergeleken op vragen die dezelfde vraagvorm hebben, maar die bij verschillende teksten gesteld zijn. Nu werd duidelijk dat een verschil in tekst voor de combinatie van intellectuele vaardigheden waarop een vraag een beroep doet, veel meer uitmaakt dan een verschil in vraagvorm.

In verband met de theorievorming over intellectuele vaardigheden wijst Van den Bergh erop dat er relatief sterke overeenkomsten voorkwamen tussen de tekstbegripscores en geheugenvaardigheden, en dat de verschillende intellectuele vaardigheden uit het Guilford-model onderling niet onafhankelijk bleken te zijn, zoals steeds wordt aangenomen.

Ten aanzien van de schrijfvaardigheid komt Van den Bergh tot verontrustende uitkomsten. Voordat hij de verschillen tussen de functionele schrijfo opdrachten en de stel opdrachten onderzocht, heeft hij namelijk eerst nagegaan in hoeverre de schrijftoetsen eigenlijk één gemeenschappelijke vaardigheid in het schrijven toetsen. Dat valt tegen. De scores op de schrijftoetsen bleken in sterke mate afhankelijk te zijn van de specifieke opdracht (het onderwerp in de opsteltitels of het schrijfprodukt bij de functionele schrijfo opdracht, bijvoorbeeld de sollicitatiebrief, een verslag of een formulier). Hij constateert dat er wel 14 verschillende opdrachten nodig zijn om een algemene en betrouwbare uitspraak over "de" schrijfvaardigheid van een leerling te kunnen doen, los van de specifieke examenopdracht.

Bij de verdere analyse van het verband tussen de verschillende soorten schrijfo opdrachten en de algemene intellectuele vaardigheden speelde de beoordeling een belangrijke rol. Daarbij ging het om het verschil tussen globale en analytische scoring. Opsteltitels werden globaal gescoord, dat wil zeggen dat de docent een totaal oordeel over het hele opstel moest geven. Functionele schrijfo opdrachten werden analytisch gescoord: de beoordelaar moest per inhoudselement of per beoordelingsaspect punten toekennen of aftrekken. Van den Bergh had verschillende opdrachten op beide manieren laten scoren. Uit zijn analyses bleek dat de prestaties op functionele schrijfo opdrachten alleen dán met andere intellectuele vaardigheden samenhangen dan de prestaties op stel opdrachten, wanneer de functionele opdrachten gescoord en de stel opdrachten globaal beoordeeld werden. Maar wanneer beide typen opdrachten op dezelfde manier beoordeeld werden, bleef er van dit verschil weinig over. Met andere woorden: dat traditionele opstel opdrachten iets anders meten dan functionele schrijfo opdrachten, komt niet doordat de leerling er anders op reageert, maar doordat de corrector ze anders beoordeelt.

Helaas bleken de intellectuele vaardigheden maar een gering deel van de resultaten op de verschillende schrijfopdrachten te kunnen verklaren. De resultaten bleken voor het grootste deel samen te hangen met de specificiteit van de gegeven opdrachten, dat wil zeggen met het specifieke opstelonderwerp of met de kenmerken van het verlangde schrijfproduct.

3. Commentaar

De bovenstaande samenvatting kan nauwelijks recht doen aan het proefschrift Van den Bergh, dat op zeer leesbare wijze veel gegevens en gecompliceerde analyses behandelt. De algemene indruk die na lezing bij mij is achtergebleven, is er een van grote zorgvuldigheid in de verslaggeving, maar zeker en vooral in de opzet en uitvoering van het onderzoek. De constructie van de toetsen is systematisch en gedegen opgezet, de keuze van de te onderzoeken intellectuele vaardigheden is ruim en weloverwogen, de analyses zijn waar mogelijk dubbel uitgevoerd, op verschillende subsets van de data, en de resultaten worden waar nodig gerelativeerd. Toch blijft er genoeg stof voor discussie over. Niet omdat Van den Bergh onverantwoorde conclusies trekt, maar omdat sommige conclusies betrekking hebben op materiaal dat niet helemaal (meer) representatief is voor de huidige examens. Andere conclusies hebben daarentegen grote consequenties ook voor andere toetssituaties dan de examens.

Het spreekt vanzelf dat examenmakers met meer dan gebruikelijke belangstelling kijken naar de toetsen die Van den Bergh heeft geconstrueerd voor de meting van tekstbegrip en schrijfvaardigheid. Dát hij nieuwe toetsen moest samenstellen, is duidelijk. De betrouwbaarheid, de beoordeelbaarheid en de variantie van de examens waren relatief laag, hetgeen betekende dat hij op basis van dat materiaal nauwelijks samenhang met andere vaardigheden zou kunnen vaststellen, ook al zou die er wel zijn. Maar het valt te betwijfelen of zijn tekstbegriptoetsen en in mindere mate zijn stelopdrachten inhoudelijk nog representatief zijn voor de huidige eindexamens.

Van den Bergh is uitgegaan van een tekst met 17 meerkeuzevragen uit het LBO-examen van 1981 en van een tekst met 17 open vragen uit het LBO-examen van 1982. Dit materiaal heeft hij bewerkt en aangevuld. De bewerking hield in dat bij meerkeuzevragen overeenkomstige open vragen gemaakt zijn door de alternatieven weg te laten, en dat bij open vragen overeenkomstige meerkeuzevragen zijn gemaakt door antwoorden van leerlingen als alternatief op te nemen. Ter aanvulling werden ook extra vragen gemaakt, zowel in open- als in meerkeuzevorm. Na de pretest zijn vragen die in psychometrisch opzicht niet voldeden, verwijderd en is uit de overige vragen een toets samengesteld die qua inhoud overeenkwam met een gemiddeld examen. Dit laatste betekent dat het resultaat een gemiddeld percentage vragen bevatte naar bijvoorbeeld expliciet vermelde informatie, naar de hoofdgedachte van een passage, of naar de structuur van de tekst. Deze hele procedure is een schoolvoorbeeld van hoe je een onderzoekstoets zou moeten maken. Alleen wijkt het resultaat op drie punten af van de examenpraktijk.

Ten eerste heeft Van den Bergh zich door zijn wijze van constructie bij voorbaat beperkt tot vragen die in beide vraagvormen bruikbaar zijn. Dat is

heel begrijpelijk, omdat er anders moeilijk zuivere vergelijkingen te maken zijn, maar het doet af aan de overeenkomst met de eindexamens, waarin juist die vraagvorm voorkomt die de beste of de enig juiste lijkt te zijn. De examenconstructeurs doen hun best een open vraag te gebruiken in die gevallen waarin de kandidaten zelfstandig een redenering, conclusie, verklaring of uitleg moeten leveren. In dat geval wordt bij voorkeur géén meerkeuzevraag gesteld. De meerkeuzevraag wordt zoveel mogelijk gereserveerd voor vragen naar de hoofdgedachte, het schrijfdoel en de structuur, en voor die gevallen waarin het antwoord op een open vraag ofwel extreem duidelijk is (bijvoorbeeld een citaat), ofwel zo sterk uiteen kan lopen dat een uitputtende opsomming niet doenlijk is. Uit de beschrijving die Van den Bergh geeft, krijg ik de indruk dat deze specifieke combinaties van vraagvorm en inhoud in zijn toets niet voorkomen, omdat ze niet in beide vraagvormen gerealiseerd kunnen worden. Zo levert het schrappen van alternatieven bijvoorbeeld onduidelijke open vragen op, omdat de alternatieven het probleem mede definiëren.

Ten tweede valt het op dat Van den Bergh in zijn toets de vragen niet aanbiedt in de volgorde van de tekstdelen waarop ze betrekking hebben. In de examens gebeurt dat wel. Door de andere aanbiedingswijze is meer zoekwerk (en een beter geheugen?) nodig dan in de examens.

Ten derde heeft het grote aantal vragen per tekst (25) de toetsconstructeurs op het Cito nogal verbaasd. Bij een dergelijk aantal is het bijna onmogelijk te voorkomen dat verschillende vragen over hetzelfde gaan (hetgeen een kunstmatig hoge interne consistentie oplevert) of dat ze betrekking hebben op details.

Bij de schrijfvaardigheid is niet helemaal duidelijk waarom de functionele schrijftoetsen uit het eigen PPON-onderzoek zijn verkozen boven oude functionele examenopdrachten. Verder is de stelopdracht die vergezeld gaat van 5 pagina's informatie, niet representatief voor de stelopdrachten in het examen, die immers bestaan uit een titel met korte toelichting. De beoordeling van opstellen wordt in de examenpraktijk niet meer volledig overgelaten aan het globale oordeel van de docent. Deze dient rekening te houden met de aanwijzingen die in toenemende mate in de opdracht worden gegeven. Maar de opsteltitels uit 1984 die Van den Bergh heeft geselecteerd zijn redelijk representatief voor de huidige examens, omdat ze enkele aanwijzingen bevatten over het gezichtspunt van de schrijver en over de doelgroep. In 1984 was dat nieuw, nu is het een minimum.

Behalve deze verschillen tussen de examens en het toetsmateriaal uit het onderzoek, doet ook het feit dat de proefpersonen (derde klas LBO/MAVO) nauwelijks examentraining hebben gehad, afbreuk aan de mate waarin de uitkomsten op de examenpraktijk betrokken kunnen worden.

De bovenstaande opmerkingen doen wellicht iets af aan de generaliseerbaarheid naar de huidige examens van de conclusies over de verschillende intellectuele vaardigheden waarop de gebruikte opdrachten een beroep doen. Andere bevindingen zijn echter zo algemeen en duidelijk, dat ze behalve voor de examens ook voor andere toetsen in onderwijs en onderzoek consequenties moeten hebben. Dit geldt met name voor de bevinding dat de tekst, de specifieke schrijf- of stelopdracht en de beoordelingsmethode voor een verontrustend

groot deel de verschillen tussen scores bepalen. In examens of toetsen die pretenderen tekstbegrip, schrijfvaardigheid of desnoods alleen maar taalvaardigheid te meten, is het bepaald ongewenst dat de score zo sterk afhankelijk is van de specifieke opgave of beoordelingsmethode. Het is dan ook hoogst noodzakelijk op dit punt verbeteringen aan te brengen.

Voor de tekstbegripexamens C/D betekent dit dat er een nog groter aantal teksten moet worden aangeboden, die niet over één thema gaan, maar veel verschillende onderwerpen behandelen. Daarmee wordt al een begin gemaakt. Ik denk dat dit ook consequenties heeft voor het HAVO-examen waarin nu maar één lange tekst wordt bevraagd.

De eis dat een examen veel meer opdrachten moet bevatten, geldt ook voor schrijfvaardigheid en helemaal voor de toetsing van het traditionele opstel. Daarbij komt nog het effect van de beoordelingsmethode. De bevinding van Van den Bergh dat de specifieke opdracht en de wijze van beoordeling de scores sterker bepalen dan de keuze tussen een opstel of een functionele opdracht, stemt tot nadenken. Dit betekent dat voor het toetsen van schrijfvaardigheid de keuze tussen opsteltitels en functionele opdrachten niet erg relevant is, minder relevant dan een goed gevarieerd aanbod van opdrachten en van beoordelingsmethoden. Misschien is het een oplossing om één examen aan te bieden, waarin een behoorlijk aantal korte opdrachten analytisch, en een langere opdracht meer globaal beoordeeld of zelfs geschaald wordt.

Opvallend is dat in het onderzoek voldoende beoordelaarsovereenstemming bereikt is bij alle beoordelingsmethoden, terwijl de beoordelingsinstructies niet essentieel afwijken van hetgeen in de examenpraktijk gebruikelijk is. Daaruit put ik de hoop dat met een uitgebreider systeem van opleiding en voorlichting op het punt van het beoordelen ook in de examenpraktijk een bevredigende objectiviteit haalbaar zou kunnen zijn.

Ten aanzien van het gebruik van het Guilford model heb ik enige twijfel. Aan de ene kant is het zeer te prijzen dat Van den Bergh de concurrerende vraagvormen heeft vergeleken met een extern criterium. Maar het model van Guilford mag dan aanzienlijk duidelijker zijn dan andere bestaande en door Van den Bergh besproken modellen, het blijft een model voor algemene intellectuele vaardigheden. Deze bieden slechts nu en dan bruikbare verklaringen voor de bevindingen ten aanzien van tekstbegrip en schrijfvaardigheid. Het dichtst bij de praktijk van het toetsen van tekstbegrip komt Van den Bergh met de constatering dat leerlingen met een hoge vaardigheid in de divergente productie van semantische relaties of eenheden (dus leerlingen die goed zijn in het produceren van alle mogelijke betekenisvolle relaties of betekenissen), alleen hoge leesscores behalen wanneer zij ook goed zijn in het evalueren van semantische produkten (dus in het beoordelen van de bruikbaarheid van hun eigen produkten). De veronderstelling is dat leerlingen die wel goed zijn in divergente productie, maar niet in tekstbegrip, hun antwoorden ontlenen aan hun kennis van de wereld, zonder ze te evalueren aan de gestelde vraag en aan de tekst. Dit lijkt me aannemelijk. Maar op dit punt had ik toch graag gezien dat de divergente productie getoetst werd aan de hand van specifieke taaltoetsjes, in plaats van door de toetsen 'getals-associaties' (noem zoveel mogelijk woorden die te maken hebben met een gegeven getal) of 'gebruiken genereren' (bedenk zoveel mogelijk manieren waarop een gegeven voorwerp gebruikt kan worden).

Tenslotte verbaasde het me dat voor elke intellectuele vaardigheid drie toetsen geconstrueerd zijn, maar dat het merendeel van deze toetsen de meerkeuzevorm had. Zeker waar het gaat om de invloed van vraagvormen ligt een minder eenzijdig gebruik van vraagvormen voor de hand.

4. Conclusie

De kritische toon van sommige van de hierboven gemaakte opmerkingen neemt niet weg dat ik grote bewondering heb voor het verrichte werk en erken dat veel van de resultaten van groot belang zijn voor de examenproblematiek bij het vak Nederlands op C-niveau.

Van groot belang is dat Van den Bergh onmiskenbaar duidelijk heeft gemaakt hoe sterk de invloed van een specifieke tekst of schrijfpodracht is in vergelijking met de vorm van de vragen of van de schrijfpodracht. Voor de opbouw van examens en andere toetsen heeft deze bevinding grote consequenties. Het is een extra impuls voor al eerder op gang gekomen pogingen om in tekstbegripexamens (Nederlands C/D, maar ook moderne vreemde talen) meer gevarieerd tekstmateriaal op te nemen. Daarnaast maakt hij behatenswaardige opmerkingen over beoordelingsmethoden.

Dat Van den Bergh zich niet heeft willen beperken tot een eenvoudiger evaluatie, maar deze heeft verbonden met een bijdrage aan het onderzoek naar intellectuele vaardigheden, is hem allerminst kwalijk te nemen. Het onderzoek heeft in dit opzicht ook resultaat gehad. Het zou mooi zijn als die resultaten aanleiding zijn om het verband tussen lees- en schrijfonderwijs, lees- en schrijfprocessen en lees- en schrijftaken verder in kaart te brengen.

Noten

1. Bij het schrijven van deze recensie werd dankbaar gebruik gemaakt van commentaar van Marie-Christine Sprengers, Harjet Robben, Ted Sanders en Ton Hendrix, allen Neerlandici, en van Hans Kuhlemeier. Het resultaat blijft een persoonlijk oordeel.

Bibliografie

- Bergh, H. van den & M.C. Sprengers, Veranderingen in het examen Nederlands tekstbegrip LBO/MAVO. *Levende Talen* 428 (1988), 87-90.
- Jong, J.H.A.L. de, Taakcomponenten en vaardigheidsfactoren. Reactie op H. van den Bergh, 'De correlatieve structuur van enkele schriftelijke taalvaardigheden'. *Tijdschrift voor taalbeheersing* 11 (1989), 3, 215-224.
- Meuffels, B., Taalvaardigheden en processen: een poging tot theoretische en empirische integratie. *Tijdschrift voor Taalbeheersing* 3 (1981), 1, 14-31.
- Meuffels, B., Boekbeoordeling H. van den Bergh, Examens geëxamineerd. *Tijdschrift voor Taalbeheersing* 12 (1990), 2, 151-159.
- Robben, H., Functionele schrijfvaardigheid in het examen LBO en MAVO. *Levende Talen* 419 (1987), 188-195.

Zwarts, M., Boekbespreking H. van den Bergh, Examens geëxamineerd.
Tijdschrift voor onderwijs research 15 (1990), 3, 196-198.