

De periodieke peiling van het onderwijsniveau Balans van het taalonderwijs aan het einde van de basisschool

1. Inleiding

In dagbladen, tijdschriften en wetenschappelijke periodieken verschijnen met de regelmaat van de klok alarmerende berichten en klachten over het bedenkelijke taalvaardigheidsniveau van leerlingen. Die klachten zijn sterk aan mode onderhevig: werd zo'n twee decennia geleden voornamelijk de belabberde spelvaardigheid van 12-jarige leerlingen op de korrel genomen, en een tiental jaren geleden de wijdverbreide onkunde in het 'traditionele' ontleden en woordbenoemen, tegenwoordig is het *bon ton* het gebrek aan schrijfvaardigheid aan de kaak te stellen.

Karakteristiek voor althans een deel van die klachten is de al dan niet uitgesproken veronderstelling, dat 'het vroeger allemaal veel beter was'. Evenzeer karakteristiek is het feit dat de klagers in de regel verzuimen ook maar één enkel empirisch gegeven aan te dragen waaruit de houdbaarheid van hun claims zou kunnen blijken. Men volstaat met anekdotische bewijsvoering ('one-shot-one-case' studies) of beroept zich op eigen observaties, aangevuld met argumentaties waarbij tegenspraak bij voorbaat wordt uitgesloten ("Ieder weldenkend mens is het er over eens dat..."; "Het lijdt geen twijfel dat..."; enzovoort).

Is het nu werkelijk zo slecht gesteld met de taalvaardigheid van leerlingen aan het eind van de basisschool? Was het vroeger allemaal beter? Om deze en soortgelijke vragen bevredigend te kunnen beantwoorden, is empirische informatie onontbeerlijk. In 1986 gaf de toenmalige Minister van Onderwijs opdracht om periodiek de inhoud en het niveau van het onderwijs empirisch vast te stellen (te 'peilen'), een prijzenswaardig initiatief dat vermoedelijk (mede) geïnspireerd was op de al jaren bestaande praktijk in landen als de Verenigde Staten (National Assessment of Educational Progress), het Verenigd Koninkrijk (Assessment of Performance Unit) en Canada (i.c. British Columbia). In 1987 zijn in het kader van deze Periodieke Peiling van het Onderwijsniveau (voortaan: PPON) gegevens verzameld over de inhoud en het niveau van het rekenonderwijs aan het einde van de basisschool, in '88 gegevens over het taalonderwijs (eveneens eind basisonderwijs), en in '89 over het taalonderwijs medio basisonderwijs.

Om recht te doen aan het periodieke karakter ligt het in de bedoeling de peilingen om de vijf jaar te herhalen. De peiling van het taalvaardigheidsniveau einde basisschool, die is uitgevoerd door het Instituut voor Toetsontwikkeling (Cito) en waarvan de resultaten gerapporteerd zijn in *Balans van het taalonderwijs aan het einde van de*

basisschool (1990), wordt dus in 1993 herhaald. Dan ook wordt het hopelijk mogelijk om de zo vaak geuite klachten over de teloorgang van ons (taal)onderwijs op hun empirische merites te beoordelen.

De *Balans*, een lijvig boekwerk van dik 200 pagina's, bevat erg veel informatie, te veel om in kort bestek *en detail* samen te vatten. Ik beperk me tot een weergave van de voornaamste bevindingen; daarbij zal met name worden ingegaan op de uitgangspunten die ten grondslag hebben gelegen aan de ontwikkeling van de taalvaardigheidsinstrumenten om de prestaties van de leerlingen te meten. Over de vraag hoe de bevindingen *gewaardeerd* moeten worden (staat het er nu goed of slecht voor met het taalvaardigheidsniveau aan het eind van het basisonderwijs?), zal ik me niet uitlaten. Een gefundeerd antwoord op deze belangwekkende vraag kan helaas (nog) niet gegeven worden, aangezien de *Balans* voornamelijk 'inhoudelijke' informatie bevat, zonder uitgebreide en specifieke methodologische verantwoording (in nog te verschijnen technische rapportages zal het Cito op de methodologische aspecten afzonderlijk ingaan). Zonder zo'n verantwoording is het lastig, zo niet onmogelijk de status en reikwijdte van inhoudelijke uitspraken over het taalvaardigheidsniveau in te schatten.

2. Peilingsonderzoek

De *Balans* bestaat uit drie delen. In het eerste deel wordt het doel van peilingsonderzoek toegelicht, in het tweede deel wordt over de opzet van de taalpeiling en de uitkomsten (zowel het 'leeraanbod' als de taalprestaties) gerapporteerd en in het derde deel leveren drie deskundigen op het gebied van het taalonderwijs hun commentaar op de peilingsuitkomsten.

In het eerste deel van de *Balans* wordt het algemene doel van peilingsonderzoek omschreven als 'systematisch bij te dragen tot het verkrijgen van een beeld van het leeraanbod en de effecten van het onderwijs' (o.c. 3). Met die bijdrage beoogt men vier subsidiaire doelstellingen te verwezenlijken:

1. een referentiekader voor de school aan te bieden om de eigen doelstellingen en resultaten te beoordelen;
2. indicaties te leveren voor het verrichten van nadere (diepte)studies en evaluatie-onderzoek die van belang kunnen zijn voor de kwaliteitsbeoordeling en -bevordering van het onderwijs;
3. bij te dragen aan de ontwikkeling van toetsen voor het gebruik door scholen, die een vergelijking met landelijke trends mogelijk maken;
4. een (minimale) empirische basis aan te bieden voor een meer algemene maatschappelijke discussie over kwaliteit en niveau van het onderwijs.

Bij oppervlakkige beschouwing van de subsidiaire doelstellingen zou men zich kunnen afvragen wat de 'meerwaarde' van PPON is, vergeleken met de bestaande Cito-(taal)toetsen die aan het eind van de zesde klas worden afgenomen en die bedoeld zijn om de selectie voor

het voortgezet onderwijs van een empirische basis te voorzien. Een aanzienlijk percentage scholen neemt immers aan deze Cito-toetsen deel; het taalvaardigheidsniveau aan het eind van de basisschool wordt dus al door deze toetsen beschreven, en dat jaarlijks. Waarom dan nog een apart peilingsonderzoek?

De Cito-toetsen bestrijken, anders dan het instrumentarium ontwikkeld door PPON, slechts een beperkt gedeelte van de doelstellingen van het moedertaalonderwijs. In PPON wordt zowel de beheersing van de vier macro-vaardigheden lezen, schrijven, luisteren en spreken gepeild, als ook die van deelvaardigheden als stijl, zinsvorming, spelling, interpunctie, handschrift, enzovoort. Verder wordt in PPON ook het 'leer- en vormingsaanbod' gepeild en wordt getracht verschillen in taalvaardigheidsprestaties te relateren aan verschillen in leeraanbod, zij het (terecht) niet in causaal-deterministische, maar in tentatief-statistische zin.

Na afname worden de Cito-toetsen openbaar gemaakt, zodat jaarlijks nieuwe opgaven geconstrueerd moeten worden. Deze omstandigheid sluit een adequate vergelijking van de taalprestaties over de tijd te enen male uit, want een geconstateerde voor- of achteruitgang in de prestaties kan niet meer ondubbelzinnig worden toegeschreven aan een verandering in vaardigheidsniveau; de moeilijkheidsgraad van de opgaven bijvoorbeeld kan veranderd zijn. De inhoud van de PPON-toetsen blijft daarom in principe geheim.¹

Bovendien dienen aan Cito-toetsen andersoortige beoordelingscriteria te worden aangelegd dan aan de toetsen die ontwikkeld zijn in het kader van PPON. Voor de jaarlijkse Cito-toetsen staat de predictieve validiteit centraal (dragen de toetsen substantieel bij aan een adequate voorspelling van het type vervolgopleiding?), voor de PPON-toetsen de inhoudsvaliditeit (dekken de toetsen de geformuleerde doelstellingen voor het moedertaalonderwijs aan het eind van het basisonderwijs?). Ook het betrouwbaarheids criterium krijgt voor beide typen toetsen een iets andere invulling. Cito-toetsen dienen, aangezien met behulp daarvan selectieve beslissingen worden genomen, erg betrouwbaar te zijn, maar voor PPON-toetsen is dit criterium minder relevant; de items in dit laatste type toetsen dienen immers zo geconstrueerd (en geanalyseerd) te worden dat het vaardigheidsniveau van de leerlingen over de hele score-range zo effectief mogelijk kan worden geschat, en dat op verschillende tijdstippen, ook wanneer zich relatief grote schommelingen in vaardigheidsniveau over de tijd zouden voordoen (met als mogelijk gevolg dat vele niet-discriminerende items met p-waarden van 1 of 0 verwijderd en vervangen moeten worden). In lijn met deze eisen worden de PPON-toetsen door het Cito dan ook (terecht) niet geconstrueerd en geanalyseerd in termen van het klassieke testmodel, maar in termen van (een van de varianten van) de zogenaamde latente trektheorie.²

Er zijn nog andere overwegingen die het unieke, van de jaarlijkse Cito-toetsen afwijkende karakter van peilingsonderzoek onderstrepen. In peilingsonderzoek gaat het, anders dan bij de Cito-toetsen waar de nadruk ligt op de prestaties van het individu, om de prestaties van de

groep leerlingen als geheel (het 'cohort'). Bovendien, en dat is een belangrijke overweging, is peilingsonderzoek *steekproef*-onderzoek, waarvan de uitkomsten (verkregen bij een relatief beperkte groep 12-jarigen) een adequate afspiegeling dienen te zijn van die van alle leerlingen aan het eind van de basisschool, ook van hen wier taalprestaties niet gepeild zijn. Een behoorlijk aantal basisscholen neemt jaarlijks deel aan de Cito-toetsen, maar gezien de vrijwillige deelname kan dit aantal, hoe aanzienlijk ook, niet als representatief beschouwd worden voor de populatie, i.c. alle basisscholen. Het is immers best mogelijk dat overwegend de 'goede' scholen aan deze jaarlijkse toetsen deelnemen. Generaliserende uitspraken over 'het' taalvaardigheidsniveau van 'de' scholier aan het eind van de basisschool zijn op basis van de uitkomsten van de jaarlijkse Cito-toetsen niet goed mogelijk.

Aan PPON hebben in totaal zo'n 4800 leerlingen van 225 scholen deelgenomen. Niet aan elk van die leerlingen zijn alle instrumenten ter meting van de taalvaardigheid voorgelegd; in het algemeen hebben in een klas van zo'n 20 leerlingen niet meer dan 3 à 4 leerlingen dezelfde toets gemaakt. Aan het merendeel van de toetsen is door 800 leerlingen deelgenomen, in een enkel geval (bijvoorbeeld bij de spreektoetsen) door slechts 130. Ook al lijkt dit een klein aantal, zeker wanneer het wordt afgezet tegen het aantal leerlingen dat aan de jaarlijkse Cito-toetsen deelneemt, niettemin zijn krachtens de specifieke wijze waarop de steekproef in het PPON-onderzoek is samengesteld, inductief-generaliserende uitspraken over 'het' taalvaardigheidsniveau van 'de' 12-jarige leerling mogelijk.

Gegeven het feit dat het PPON-onderzoek steekproef- en geen populatie-onderzoek is, zijn twee opmerkingen relevant. In de eerste plaats dient beseft te worden dat alle inhoudelijke uitspraken over de beheersing van een vaardigheid geen factisch-beschrijvende uitspraken zijn, maar schattingen met een zekere foutenmarge. In de tweede plaats, en dat is belangrijker, wordt de mogelijkheid tot generalisatie binnen het PPON-onderzoek gemitigeerd door de omstandigheid dat van de oorspronkelijk in de steekproef opgenomen scholen de bereidheid tot deelname verre van optimaal was - meer dan de helft van de aangeschreven scholen weigerde medewerking aan PPON te verlenen. Naar het waarom kan men slechts gissen. Het punt waar het met het oog op het steekproef-karakter van het PPON-onderzoek echter om draait is de vraag, of de scholen die weigerden deel te nemen op een of meerdere relevante attributen systematisch en substantieel afwijken van de scholen die wel bereid waren deel te nemen. Is dit het geval, dan is generalisatie een hachelijke zaak. Aan het hier geschetste probleem, dat in de literatuur bekend staat onder de naam 'non-respondenten'-probleem, wordt in de *Balans* geen expliciete aandacht besteed. Het lijkt me echter wenselijk om in de toekomst, via een aparte (diepte)studie (zie doelstelling 2 van PPON), althans een deel van de aard en omvang van dit probleem in kaart te brengen.

3. Het leer- en vormingsaanbod

Om inzicht te krijgen in het leer- en vormingsaanbod, is in het PPON-onderzoek onder meer nagegaan welke taalmethode de scholen gebruiken, en hoeveel van de beschikbare tijd er besteed wordt aan het taalonderwijs in het algemeen en aan macro- en deelvaardigheden in het bijzonder. Bovendien is aan de leerlingen een vragenlijst voorgelegd om na te gaan welke taalactiviteiten ze buiten de school ontlooiën.

Bijna elke school die medewerking aan het PPON-onderzoek verleende, maakt bij het taalonderwijs gebruik van in de handel verkrijgbare methoden. *Taal actief*, *Jouw taal*, *mijn taal* en *Taal totaal* worden tesamen op 70% van de scholen gebruikt. Vrijwel alle leerkrachten volgen hun methode in zijn geheel of in belangrijke mate. Vooral wat het begrijpend lezen betreft worden de in de handel verkrijgbare methoden aangevuld met hulpmethoden, deels omdat in de betreffende methoden dit onderdeel onvoldoende aan bod komt, deels uit ontevredenheid van de leerkrachten over de wijze waarop het begrijpend lezen behandeld wordt. Ook voor spelling, technisch lezen en ontleden zoeken leerkrachten hun heil in complementerende hulpmethodes.

De tijd die per week aan taalonderwijs besteed wordt (gemiddeld 411 minuten) vormt meer dan een kwart van de beschikbare onderwijstijd. Voor rekenen wordt veel minder tijd uitgetrokken: 255 minuten. In die 7 uur taalonderwijs wordt de meeste tijd besteed aan het oefenen in deelvaardigheden als spelling, ontleden, technisch lezen (tesamen 4 uur), 2 uur aan de macro-vaardigheden lezen, schrijven, luisteren en lezen en 1 uur aan drama en expressie.

Van de leerlingen geeft 4% te kennen nooit een boek te lezen, een even groot percentage zegt nooit TV te kijken. De leerlingen die wel lezen, doen dat regelmatig en tamelijk veel: gemiddeld bijna 4 boeken per maand. Stripboeken en jeugdboeken zijn het populairst. Ook blijven veel kinderen (89%) in hun vrije tijd wel eens wat te schrijven, voornamelijk verhalen en briefjes aan familie en vrienden. Dagelijks kijkt 70% van de leerlingen naar de televisie; de dag voor de toetsafname heeft 81% van de leerlingen ruim 7 kwartier gekeken.

De hier vermelde uitkomsten over het leer- en vormingsaanbod zijn verre van volledig, want zowel aan leerkrachten als aan leerlingen zijn veel meer en veel specifiekere vragen gesteld dan die welke hier, noodgedwongen in beknopte vorm, de revue zijn gepasseerd. Niettemin biedt de hier gegeven samenvatting inzicht in de wijze, waarop in het PPON-project getracht is het leer- en vormingsaanbod in kaart te brengen. Daar kleven enkele bezwaren aan.

Een bezwaar betreft de rapportage. Bij de tijdsbesteding bijvoorbeeld worden enkel gemiddelden vermeld; het is genoegzaam bekend dat gemiddelden, wanneer daaraan geen spreidingsmaten zijn toegevoegd, een vertekend beeld van 'de werkelijkheid' kunnen opleveren. Een ander, fundamenteeler bezwaar betreft het feit dat de via questionnaires verkregen antwoorden van leerkrachten en leerlingen 'at face value' worden geaccepteerd. Maar wat mensen zeggen, hoeft uiteraard

niet overeen te komen met wat er in feite achter de deuren van het klasse-lokaal of buiten de school gebeurt. Hoe moeilijk is het niet om een accurate schatting te geven van de tijd, die je wekelijks aan de een of andere activiteit besteedt? Om enig inzicht te krijgen in de betrouwbaarheid van de beantwoording, had men er goed aan gedaan wanneer men (een deel van) de leerkrachten en de leerlingen (een deel van) de gestelde vragen opnieuw had laten beantwoorden. De ervaring leert immers dat de betrouwbaarheid (i.c. stabiliteit) van verbale reacties in de regel tegenvalt.

Niet alleen aan de betrouwbaarheid (de consistentie van de antwoorden over de tijd), maar ook aan de validiteit van de door leraren en leerlingen gegeven antwoorden (het waarheidsgehalte) wordt in PPON, ten onrechte, geen aandacht geschonken. Dit is des te verwonderlijker, aangezien aan de 'output'-kant (de prestaties op de taal-toetsen) de PPON-onderzoekers ernaar gestreefd hebben en de betrouwbaarheid en de validiteit van de meetinstrumenten te waarborgen. Diverse effecten kunnen de validiteit van verbale reacties contamineren, bij voorbeeld wanneer respondenten, bewust dan wel onbewust, middels hun antwoorden een sociaal wenselijk beeld van zichzelf presenteren. Zulk type 'sociale wenselijkheidstendenties' lijken een rol te hebben gespeeld wanneer men niet erg plausibele uitkomsten in de *Balans* aantreft als: "Van de leerlingen heeft 70% de dag voorafgaande aan de peiling gelezen, gemiddeld drie kwartier" en "Op de dag voorafgaande aan de peiling is door 46% van de leerlingen geschreven, ongeveer een kwartier", en schrijven voor schoolwerk is daarbij niet meegerekend (o.c. 38). Uit navraag bij de ouders van enkele leerlingen zou men indicaties hebben kunnen verkrijgen over het waarheidsgehalte van zulke antwoorden. Zowel ten aanzien van de betrouwbaarheid als ten aanzien van de validiteit ontbreekt het in de peiling naar het vormings- en leeraanbod aan controle. Het van dat aanbod gepresenteerde beeld bevat ongetwijfeld ruis, maar anders dan bij de instrumenten ter meting van de taalvaardigheden is onbekend in welke mate die ruis de werkelijkheid vertekent.

De 'input'-kant van het taalonderwijs komt er in PPON toch al enigszins bekaaid af, zeker wanneer die wordt afgezet tegen de 'output'-kant. Die onderbelichting wordt gereflecteerd in het aantal pagina's, dat in de *Balans* aan het leer- en vormingsaanbod wordt gewijd: nog geen 10 pagina's, vergeleken met dik 100 pagina's waarop de prestaties bij het lezen, luisteren, schrijven, spreken en verschillende deelvaardigheden worden beschreven.

Verder krijgt de lezer van de *Balans* wel een heel erg ruwe schets van het leer- en vormingsaanbod aangeboden - dat het merendeel van de leerkrachten een in de handel verkrijgbare methode op de voet volgt c.q. zegt te volgen is ongetwijfeld informatief, maar *hoe*, op welke wijze en in welk didactisch kader die methode vorm wordt gegeven blijft duister. Wat gebeurt er nu eigenlijk in de klas? Systematische observatie, in aanvulling op de nu gehanteerde vragenlijsten, zou een veel genuanceerder en rijker beeld van het leer- en vormingsaanbod kunnen opleveren. Als men daar een dieptestudie naar

zou willen verrichten, dan zou tevens de leerkracht als *variabele* in het onderzoek opgenomen moeten worden. Nu is het immers zo, dat er geen enkel substantieel verband kon worden aangetoond tussen enerzijds de aard van de gehanteerde taalmethode en anderzijds de prestaties op de instrumenten ter meting van de taalvaardigheid. Maar de leerkracht, zijn enthousiasme, zijn inzet en goede smaak, of juist het gebrek daaraan, het blijft een onmisbare schakel.

4. De prestaties

Het zou betrekkelijk eenvoudig zijn operationalisaties van verschillende taalvaardigheden te construeren, wanneer eindtermen voor het onderdeel Nederlandse taal voor het basisonderwijs beschikbaar zouden zijn. Liggen deze doelstellingen vast, dan kunnen instrumenten ontwikkeld worden die specifiek zijn toegesneden op de meting van die doelstellingen. Op dit moment echter ontbreken zulke eindtermen; of de mede door het SLO ontworpen voorstellen op dit vlak ooit politiek bekrachtigd zullen worden is nog zeer de vraag.

Het ontbreken van eindtermen stelt de onderzoekers van PPON uiteraard voor een lastige opgave. Er bestaat immers altijd het gevaar dat er taalvaardigheden worden gemeten waarvan de relevantie door 'het veld' wordt ontkend, of dat er taalvaardigheden worden gemeten waaraan in de taalmethoden geen aandacht wordt besteed, of dat er juist belangrijk geachte vaardigheden niet worden gemeten, enzovoort. Het commentaar van de drie deskundigen die in deel III van de *Balans* aan het woord komen, spreekt wat dit betreft boekdelen: de een is van oordeel dat er in de taalpeiling te veel de nadruk wordt gelegd op de formele aspecten bij taalbeschouwing, en juist te weinig op creatieve aspecten bij het schrijven, de tweede is van oordeel dat het lezen van argumentatieve teksten helemaal niet thuishoort op de basisschool, terwijl in de peiling het begrip van verhalende teksten juist ontbreekt, en de derde commentator mist onder andere toetsen voor het begrip van verhalen en gedichten, terwijl naar zijn smaak het kritisch lezen in de peiling wordt onderbelicht. Zonder deze drie commentatoren te willen desavoueren moet niettemin geconstateerd worden, dat hun opmerkingen een incidenteel karakter dragen en het structurele fundament waarop de hele onderneming van het operationaliseren van taalvaardigheden berust, in essentie onaangetast laten.

Het instrumentarium waarmee de verschillende taalvaardigheden zijn gemeten, is omvangrijk, divers, sluit in het algemeen aan bij wat genoemd wordt 'de leefwereld' van het kind, en kan in het algemeen gekarakteriseerd worden als leuk, aantrekkelijk; verder dekt dit instrumentarium het gros van de in de taalmethoden aan de orde gestelde activiteiten, ook die activiteiten (als spelling, interpunctie en ontleden) waarvan het belang door sommigen ten onrechte wordt gebagatelliseerd. Verder hebben de PPON-onderzoekers niet geschroomd om ook de moeilijk te operationaliseren vaardigheden als spreken en schrijven in de peiling te betrekken. Voeg daar nog bij dat voor het gros van al

die operationalisaties geldt dat deze zowel betrouwbaar (i.c. schaalbaarheid in termen van de latente trektheorie) als (inhouds)valide zijn, en het wordt duidelijk dat op het vlak van instrumentatie de PPON-onderzoekers ongetwijfeld een prestatie van formaat hebben geleverd.

Omdat eindtermen ontbreken, hebben de PPON-onderzoekers zich bij de constructie van de operationalisaties van de taalvaardigheden gebaseerd op wat genoemd wordt een 'domeinbeschrijving': het taal-domein is in eerste instantie opgesplitst in de vier bekende macro-vaardigheden lezen, luisteren, schrijven en spreken. Deze vier vaardigheden worden op hun beurt gesubcategoriseerd in verschillende subvaardigheden, afhankelijk van de specifieke taalgebruiksfunctie die aan een tekst kan worden toegeschreven: rapporterend, beschouwend, directief en argumentatief. Conform dit uitgangspunt worden er voor de meting van een macro-vaardigheid als begrijpend lezen (dat in PPON weer wordt onderscheiden van de deelvaardigheid 'technisch lezen' waarvoor twee aparte operationalisaties zijn ontwikkeld) vier verschillende meetinstrumenten geconstrueerd: een instrument ter meting van het begrip van rapporterende teksten, een instrument ter meting van het begrip van beschouwende teksten, enzovoort. Wat de macro-vaardigheid begrijpend lezen betreft blijft het daar niet bij: het begrip wordt zo ruim opgevat dat daaronder ook 'taal'activiteiten als het hanteren van informatiemiddelen (naslagwerken), het 'lezen' van tabellen en grafieken en het 'lezen' van kaarten vallen; en ook die activiteiten worden op hun beurt geoperationaliseerd. Het macro-begrip 'lezen' wordt dus door in totaal negen toetsen bestreken.

Het begrip 'taalgebruiksfunctie' dat bij de meting van het macro-begrip lezen zo'n fundamentele rol speelt, treffen we ook aan bij de meting van de drie andere macro-vaardigheden. Voor luisteren zijn vier toetsen ontwikkeld (i.c. luisteren naar rapporterende, beschouwende, directieve en argumentatieve teksten), voor het schrijven en voor het spreken zijn elk 21 taken geconstrueerd, waarbij drie taken steeds een specificatie vormen van één type taalhandeling die kenmerkend zou zijn voor een specifieke taalgebruiksfunctie. Zo zouden de taalhandelingen 'informatie doorgeven' en 'beschrijven' elk kenmerkend zijn voor taalgebruik waaraan een rapporterende functie kan worden toegekend, de taalhandelingen 'samenvatten' en 'gevoelens/evalueren' karakteristiek voor taalgebruik dat als 'beschouwend' kan worden getypeerd, enzovoort. Elke taalhandeling (er worden er in totaal zeven onderscheiden, bij het schrijven dezelfde als bij het spreken) wordt geconcretiseerd in drie taken. Een voorbeeld ter toelichting. De taalhandeling 'informatie doorgeven' bij het schrijven wordt instrumenteel uitgewerkt in de volgende drie situaties: a) een kattebelletje schrijven voor de overige gezinsleden; b) een memo schrijven, gericht aan de directeur van de school over een telefoontje; c) een briefje schrijven als reactie op een advertentie (NB: bij de twee eerstgenoemde taken krijgen de leerlingen de opdracht op een bandje aangeboden). Op het meest concrete niveau resulteert de instrumentele uitwerking van de hier geschetste principes in 55 verschillende instrumenten voor de meting van de macro-vaardigheden, op iets abstracter niveau (wanneer

de drie taken per taalhandeling worden geaggregeerd) in 27 instrumenten.

Daar blijft het overigens niet bij. Bij het schrijven worden ook de deelvaardigheden interactie, stijl, grammaticaliteit, spelling, interpunctie en handschrift beoordeeld, bij het spreken inhoud en organisatie, interactievaardigheid, formuleervaardigheid en spreektechniek (volume, intonatie, tempo en articulatie). Voor de deelvaardigheden stijl, grammaticaliteit, spelling en interpunctie zijn bovendien aparte toetsen ontwikkeld. Taalbeschouwing tot slot is met behulp van twee toetsen gemeten (één voor het ontleden en woordbenoemen, de andere voor semantische en pragmatische aspecten van schriftelijk taalgebruik).

Het hoeft weinig betoog dat de in het PPON-onderzoek gehanteerde domeinbeschrijving tot een uiterst gedifferentieerd en rijk geschakeerd meetinstrumentarium van de taalvaardigheid leidt; het hoeft evenmin veel betoog dat degene, die in kort bestek een informatieve samenvatting van al de uitkomsten betreffende al die taalvaardigheidsmaten wil verschaffen, voor een vrijwel ondoenlijke opgave wordt gesteld.

Hoe presteren de leerlingen op al die taalvaardigheidsmaten? Zijn de zo vaak geuite klachten over het bedenkelijke taalvaardigheidsniveau van de leerlingen nu wel of niet terecht? Om die vraag adequaat te beantwoorden, moeten deugdelijke maatstaven gespecificeerd zijn waaraan de kwaliteit van de door de leerlingen geleverde prestaties kunnen worden afgemeten. In het PPON-onderzoek worden twee verschillende typen maatstaven geëxpliciteerd, één met een absoluut en één met een relatief karakter. Een vaardigheid wordt door de leerlingen als groep voldoende beheerst, indien een gemiddelde leerling de helft van de opgaven goed beheerst en 90% van de opgaven tenminste matig, en indien een zwakke leerling ten minste 10% van de opgaven goed beantwoordt en minstens de helft matig beheerst, aldus de specificaties bij deze absolute norm.³ De relatieve norm die bij de beoordeling van de prestaties wordt gehanteerd, wordt geconstitueerd door de verwachtingen en wensen over het niveau, geuit door drie groepen respondenten: 1. ouderleden van medezeggenschapsraden 2. leraren van het basisonderwijs en 3. leraren Nederlands uit de brugklas van het voortgezet onderwijs. Deze respondentengroepen kregen bij elke toets/taak enkele opgaven voorgelegd, waarvan ze moesten aangeven welke daarvan een gemiddelde leerling goed beheerst, waarbij 'goed beheerst' werd gespecificeerd tot 'acht van de tien soortgelijke opgaven goed beheerst'. Verder moesten zij zich uitspreken over de vraag, welke opgaven naar hun oordeel goed beheerst zouden moeten worden. Een grote discrepantie tussen verwachtingen enerzijds en wensen anderzijds zou indicatief zijn voor de mate van 'onvrede' met het bereikte niveau.

Gewapend met deze twee verschillende typen maatstaven is het mogelijk uitspraken te doen over de kwaliteit van de prestaties van de leerlingen. Wat het lezen betreft wordt het onderdeel 'technisch lezen' voldoende beheerst, evenals het lezen van beschouwende teksten, het 'lezen' van tabellen en grafieken en het 'lezen' van kaarten. Ook

blijken de leerlingen in voldoende mate naslagwerken te kunnen hantieren, althans wanneer het gaat om de beheersing van vaardigheden die een noodzakelijke voorwaarde daarvoor vormen (alfabetiseren bijvoorbeeld). Maar in de toepassing van die kennis schieten de leerlingen tekort; ook wat het lezen van rapporterende en van argumentatieve teksten betreft moeten de prestaties als onvoldoende gekwalificeerd worden. Wat het luisteren betreft worden echter alle onderdelen in voldoende mate beheerst. Anders dan bij de receptieve vaardigheden lezen en luisteren presteren de leerlingen bij de productieve vaardigheden schrijven en spreken duidelijk onder de maat. Terwijl de drie respondentgroepen verwachten dat 60 tot 80% van de leerlingen aan de norm voldoet, blijkt dat in de praktijk slechts 2 tot 55% van de leerlingen hieraan voldoet.

Ook bij de deelvaardigheden treffen we een gevarieerd beeld in het prestatieniveau aan. Wat stijl, grammaticaliteit, spelling en taalbeschouwing betreft leveren de leerlingen voldoende prestaties, iets wat niet gezegd kan worden van de onderdelen interpunctie en woordbenoemen en zinsontleden.

Kunnen we nu op basis van deze resultaten voor bijvoorbeeld de productieve vaardigheden schrijven en spreken de conclusie trekken, dat massa's leerlingen functioneel analfabeet genoemd moeten worden, omdat zij in praktische, functionele situaties de betreffende vaardigheden nauwelijks adequaat kunnen aanwenden?⁴ Voordat een dergelijk vergaande conclusie getrokken kan worden, zal eerst het fundament waarop de hele empirische onderneming berust en de gehanteerde beoordelingscriteria kritisch onderzocht moeten worden.

5. Een niet-eenduidige taxonomie

De manier waarop onderwijsresultaten worden gemeten, weerspiegelt vaak een impliciete of expliciete theorie, zo merkt één van de commentatoren in deel III van de *Balans* terecht op. Wat is de waarde van de taxonomie die ten grondslag ligt aan de domeinbeschrijving, op grond waarvan (onder meer) rapporterende, beschouwende, directieve en argumentatieve teksten worden onderscheiden? Impliciet wordt in het PPON-onderzoek aangenomen dat deze vier verschillend benoemde teksten typen vertegenwoordigen die elk gekenmerkt worden door een eigen, karakteristieke formele structuur. Het cruciale punt waar het bij de beoordeling van de waarde van de taxonomie om gaat, is de vraag of die typen teksten elk eigen, karakteristieke en met die formele structuur samenhangende probleemmoelijkheden constitueren, voor de oplossing waarvan de probleemoplosser moet appeleren aan telkens een *ander* type vaardigheid. Zou dat immers niet het geval zijn, dat wil zeggen zou ongeacht het type tekst telkens *dezelfde* vaardigheid gemobiliseerd moeten worden, dan verliest het onderscheid in typen teksten zijn informatieve waarde. Discriminatie, onderscheid heeft pas zin, als daardoor extra informatie wordt verschaft.

Relaties tussen de verschillende (sub)vaardigheden die gemobiliseerd moeten worden bij de vier onderscheiden typen teksten worden helaas niet vermeld, zodat we genoegen moeten nemen met andere, minder belangrijke criteria voor de beoordeling van de waarde van de taxonomie. Eén criterium betreft de mate van eenduidigheid. Zijn de onderscheiden categorieën eenduidig, kunnen we op basis van de gegeven omschrijvingen van de categorieën een concrete tekst eenduidig toeschrijven aan één van de vier onderscheiden klassen?

De gegeven omschrijvingen van de onderscheiden typen zijn ronduit vaag. In rapporterende teksten wordt een feitelijke weergave gegeven "van de werkelijkheid, zoals verslagen, mededelingen of beschrijvingen", in beschouwende teksten geeft een schrijver "een stand van zaken (weer) (...), maar (brengt) tegelijk ook zijn opvattingen tot uitdrukking (...), bijvoorbeeld boekbesprekingen, kritieken, discussies", directieve teksten zijn bedoeld "om het handelen van de lezer te begeleiden en te sturen. Voorbeelden: een gebruiksaanwijzing of een routebeschrijving", terwijl in argumentatieve teksten wordt getracht "de denkbeelden van de lezer te beïnvloeden. Voorbeelden zijn advertenties en protestbrieven". Waarom zou, bijvoorbeeld, een advertentie niet moeten worden ondergebracht in de categorie directieve teksten? Opstellers van een advertentie hebben toch primair de intentie het handelen van de lezer te sturen? En in een kritiek of boekbespreking, die in het PPON-onderzoek onder de categorie beschouwende teksten vallen, tracht een schrijver toch ook argumenten aan te voeren om zijn kwalificaties te staven teneinde de denkbeelden van een lezer te beïnvloeden? De categorieën in de taxonomie overlappen elkaar teveel en zijn te heterogeen van samenstelling, de voorbeeldteksten die bij de betreffende categorieën worden geleverd, voldoen nauwelijks aan de definitie die aan de betreffende categorie wordt gegeven en --als voornaamste bezwaar-- er is geen consequent indelingsprincipe.

Dat een dergelijke taxonomie tot spraak- en denkverwarring leidt, moge blijken uit de resultaten van een onderzoekje onder 11 doctoraalstudenten Nederlandse Taal- en Letterkunde. Hen werd gevraagd de zes teksten die in het voorbeeldenboek bij de *Balans* zijn opgenomen, in de vier onderscheiden categorieën onder te brengen (een asterisk in de tabel geeft aan tot welk type de voorbeeldtekst gerekend moet worden).

Tekst	Rapporterend	Beschouwend	Directief	Argumentatief
Kinderrechtswinkel	8*		5	
Telefoon	11*	1		
Schoolmuseum		11*	1	2
Jongens van Soweto	1	10*	1	3
Buitenbassins	5	3	4	5*
Kampioenen		5		7*

Geen enkele student bleek in staat alle zes de teksten 'goed' te categoriseren, dat wil zeggen conform de typeringen in PPON.⁵ Maar ook voor de schrijvers van de *Balans* is het kennelijk niet geheel duidelijk wat precies onder de vier onderscheiden categorieën moet worden verstaan: wordt er in de *Balans* bij het luisteren naar argumentatieve teksten gesproken over teksten waarmee een schrijver de *denkbeelden* wil beïnvloeden, in het voorbeeldenboek wordt diezelfde categorie getypeerd als teksten waarin getracht wordt *het gedrag en de denkbeelden* van de luisteraar te beïnvloeden. Het beïnvloeden van het gedrag hoorde toch thuis onder de noemer directief?

Men kan het hier geleverde commentaar op de taxonomie afdoen als 'wetenschappelijke haarkloverij', maar niet vergeten moet worden dat de consequenties van het gebrek aan eenduidigheid in de taxonomie verstrekkend zijn. Aan de leerkrachten is ook gevraagd naar de tijdsbesteding bij de vier onderscheiden typen teksten (zowel bij het lezen als bij het luisteren). Als verre van duidelijk is wat precies onder de onderscheiden categorieën moet worden verstaan, welke waarde moet er dan gehecht worden aan 'uitkomsten' als 'bij het lezen wordt relatief de meeste tijd besteed aan rapporterende teksten, en de minste aan argumentatieve'? Welke oefenteksten zou een leerkracht zijn leerlingen moeten aanbieden, wanneer hij, geconfronteerd met de PPON-uitkomsten, speciale aandacht aan het lezen van rapporterende en argumentatieve teksten zou willen schenken? Welke waarde moet er worden gehecht aan een 'uitkomst' als 'het lezen van rapporterende en argumentatieve teksten wordt niet voldoende beheerst'? En wat te denken van het commentaar van één van de drie commentatoren die het niet verwonderlijk acht dat kinderen weinig van argumentatieve teksten maken, want ze horen naar zijn oordeel nog niet in het basisonderwijs thuis?

Gezien het periodieke karakter van PPON is het uitgesloten fundamentele wijzigingen aan te brengen in de taxonomie die ten grondslag ligt aan de meting van de taalvaardigheidsprestaties. Om spraak- en denkverwarring te voorkomen verdient het echter aanbeveling de omschrijvingen van de categorieën aan te scherpen.

6. De uitwerking van het pragmatisch uitgangspunt

Als *fundamentum divisionis* om verschillende soorten teksten van elkaar te scheiden is, zoals eerder opgemerkt, gekozen voor het type *functie* dat aan taalgebruik kan worden toegekend. Dit uitgangspunt, ontleend aan de pragmatiek, komt er kort gezegd op neer dat in de verbale communicatie niet het woord, woordgroep of zin centraal staat, maar de *taalhandeling* (i.c. de illocutionaire handeling). Maar de potentiële vruchtbaarheid van dit uitgangspunt wordt in het PPON-onderzoek aangetast door een onzorgvuldige en idiosyncratische uitwerking, waarin voorbij wordt gegaan aan de theoretische inzichten en opvattingen die binnen de pragmatiek als gemeengoed bestempeld kunnen worden.

In de standaardtheorie worden vijf typen illocuties onderscheiden: beweerders ('assertives'), stuurders ('directives'), binders ('commissives'), uitdrukkers ('expressives') en verklaarders ('declarations') waarmee vijf verschillende soorten basisfuncties van taalgebruik corresponderen: we vertellen mensen hoe dingen in elkaar zitten, we trachten hen ertoe te bewegen dingen te doen, we verplichten onszelf dingen te doen, we geven uiting aan onze gevoelens en opinies en we brengen door middel van onze uitingen veranderingen in de werkelijkheid tot stand. Vergelijken we deze in de pragmatiek vigerende classificatie, die betrekkelijk consistent en eenduidig is, met de in het PPON-onderzoek gehanteerde taxonomie, dan valt ogenblikkelijk op dat daarin zowel functies zijn geschrapt als één functie is toegevoegd. Zo komt de categorie 'rapporterend' min of meer overeen met de klasse van beweerders (zoals beweren, mededelen, rapporteren, informeren), de categorie 'beschouwend' met die van de uitdrukkers (bedanken, feliciteren, condoleren, betreuren) en de categorie 'directief' met die van de stuurders (verzoeken, bevelen, uitdagen), terwijl de klasse van de binders (beloven, zweren, verzekeren) en die van de verklaarders (de oorlog verklaren, ontslag nemen, in de echt verbinden) ontbreken. Waarom wordt niet duidelijk gemaakt.

Evenmin wordt duidelijk, waarom van de scherp omschreven definities binnen de pragmatiek afgeweken is en waarom deze vervangen zijn door vage omschrijvingen. Zo wordt de oorspronkelijke klasse van de beweerders (onder meer) getypeerd door de omstandigheid dat de leden van deze klasse, anders dan die van de overige klassen, op waarheid of onwaarheid beoordeeld kunnen worden. Van een bewering kun je immers zeggen, of deze waar of onwaar is, maar dat kan niet bij een belofte, of een verzoek -hier is niet zozeer het begrip 'waarheid' relevant, maar het begrip 'geslaagd'. De versmalling van deze klasse tot rapporterende teksten waarin een *feitelijke* weergave van de werkelijkheid wordt gegeven, zoals in het PPON-onderzoek gebeurt, heeft onder meer tot gevolg dat (fictieve) verhalende teksten niet meer onder deze klasse kunnen worden ondergebracht - en het zijn juist verhalende teksten die door de commentatoren node worden gemist. Het onderbrengen van verhalende teksten onder de noemer 'beweerders' (om het even hoe men deze klasse verder wil benoemen) zou wel lukken, wanneer men zich meer aan de oorspronkelijke definitie gehouden had.

Door het toevoegen van de categorie 'argumentatief' wordt het indelingsprincipe aangetast: immers, bij deze categorie staat niet meer de illocutie, maar de perlocutie (i.c. het met de illocutie verbonden effect, het overtuigen) centraal! Een advertentie hoort, qua illocutie, thuis in de klasse van de stuurders, een protestbrief in de klasse van uitdrukkers - maar in het PPON-onderzoek worden deze heterogene elementen onder dezelfde categorie 'argumentatief' geschaard, een categorie die op een heel ander niveau gesitueerd moet worden dan de overige drie.

Toegegeven, een volstrekt eenduidige, uitputtende teksttypologie bestaat niet, en het is twijfelachtig of deze op korte termijn ontwik-

keld zal kunnen worden. Ook wanneer men in het PPON-onderzoek de uitgangspunten van de pragmatiek consequenter en zorgvuldiger had toegepast, dan nog zouden niet alle classificatieproblemen opgelost zijn. Enerzijds is dat het gevolg van het feit, dat binnen de pragmatiek de nadruk op de intenties en bedoelingen van de spreker/schrijver ligt⁶ - welke intentie of intenties aan een schrijver van een tekst kunnen worden toegeschreven is niet zelden een kwestie van interpretatie. Anderzijds levert de uitwerking van de pragmatische uitgangspunten problemen op, doordat taalhandelingen gedefinieerd zijn op zinsniveau, en niet op tekstniveau. Zo kan op tekstueel niveau een stuuder (een advertentie bijvoorbeeld) gerealiseerd worden door louter beweerders op elementair zinsniveau.

7. Communicatieve effectiviteit

Waarom vallen de prestaties bij de produktieve taalvaardigheden schrijven en spreken zo tegen? Anders dan de prestaties bij de receptieve vaardigheden zijn de schrijf- en spreekprestaties *beoordeeld*, en wel door 180 leraren uit groep 7 en 8. Bij hun beoordeling van de 21 taken maakten ze gebruik van een voorbeeldtekst en van teksten die daar in positieve dan wel negatieve zin van afweken. De beoordeling richtte zich primair op de mate waarin de transactie is geslaagd die in de taak is bedoeld: is de tekst zodanig dat de lezer zonder veel moeite kan doen of laten wat de schrijver van hem vraagt. We spreken in dit verband van communicatief effectief (o.c. 86-87). Dit begrip, communicatieve effectiviteit, is weliswaar gebaseerd op het binnen de pragmatiek vigerende begrip illocutionair of communicatief effect (is de lezer/luisteraar in staat te achterhalen welke propositie de spreker/schrijver met welke communicatieve strekking (een bewering?, een belofte? een waarschuwing?) beoogt over te brengen?), maar heeft toch een wat andere, meer uitgebreide betekenis. Immers, ook de *kwaliteit van de stilistische middelen* waarmee het illocutionaire doel tot uitdrukking wordt gebracht, speelt in het PPON-onderzoek een niet te onderschatten rol bij de beoordeling van de schrijfprestaties (omslachtige formuleringen, het weglaten van zinsgrenzen en dergelijke worden de schrijver aangerekend, bondigheid, duidelijkheid en dergelijke leveren pluspunten op). De bewering dat 12-jarigen niet of nauwelijks in staat zijn zich schriftelijk communicatief effectief uit te drukken, moet dus in aanzienlijke mate worden genuanceerd.

In de tweede plaats valt er heel wat af te dingen op de normen, die worden aangelegd bij de beoordeling van een communicatief effectieve tekst. Het uitgangspunt bij de beoordeling wordt gevormd door voorbeeldteksten die aan 'de norm' voldoen - maar deze norm is veel te hoog gesteld, en strookt bovendien niet met de betekenis van het begrip 'communicatieve effectiviteit' zoals dat binnen PPON figureert. Een voorbeeld. De leerling heeft op school telefoondienst. Hij krijgt een telefoontje van een zekere mevrouw De Jong van de afdeling Onderwijs van de gemeente; zij deelt hem mee dat de leerling de

directeur van de school moet berichten dat een vergadering die vandaag om 8 uur zou plaatsvinden en waarbij de directeur aanwezig zou zijn, geen doorgang kan vinden vanwege ziekte van twee mensen. De vergadering wordt uitgesteld en 14 dagen verplaatst, zelfde aanvangstijd, zelfde plaats. Is de directeur dan verhinderd, dan kan hij contact opnemen met de opbelster, tel.nr. 522631. Deze opdracht krijgt de leerling per telefoon aangeboden; hij mag tijdens het telefoongesprek aantekeningen maken. Vervolgens moet hij het bericht in een memo aan de directeur vastleggen. De volgende tekst voldoet, aldus PPON, niet aan de norm: Wegens ziekte is vanavond de vergadering uitgesteld voor twee weken als u niet kunt moet u opbellen naar het nummer 522631. Het onderwerp blijft hetzelfde.

Deze tekst is naar mijn oordeel wel degelijk communicatief effectief: de directeur weet immers dat de vergadering vanavond niet doorgaat, maar over twee weken wel. Hij weet ook, waarom. Verder wordt de directeur op de hoogte gesteld van het telefoonnummer dat hij moet bellen, mocht hij over twee weken verhinderd zijn. Daarmee is het primaire illocutionaire doel bereikt. Het feit dat de naam van de opbelster, mevrouw De Jong, niet wordt vermeld, wordt door de onderzoekers van PPON 'minder handig' gevonden, maar die vermelding is voor het bereiken van het communicatieve effect van dit memo volstrekt niet relevant. Als 'de norm', ten onrechte overigens, zo extreem hoog wordt gesteld, dan hoeft het ook geen verbazing te wekken dat vervolgens massa's leerlingen die norm niet halen.

Een derde 'verklaring' voor de tegenvallende prestaties bij het schrijven en spreken kan gevonden worden in de manier, waarop de schrijftaken zijn geoperationaliseerd. Het hierboven gegeven voorbeeld van 'De telefoondienst' maakt duidelijk dat de schrijfprestaties niet alleen afhankelijk zijn van de schrijfvaardigheid, maar evenzeer van de luistervaardigheid en van het geheugen. Bij weer andere schrijftaken (samenvatten bijvoorbeeld) moet een sterk beroep op de leesvaardigheid gedaan worden. Zijn er bij het 'schrijven' ook andere vaardigheden in het geding, dan worden de 'schrijf'prestaties noodzakelijkerwijs gedrukt, tenzij aannemelijk wordt gemaakt dat er geen individuele verschillen in die andere vaardigheden optreden.

Een vierde 'verklaring' kan naar mijn oordeel gevonden worden in de intrinsiek cognitieve complexiteit van de taken waarvan een groot gedeelte domweg niet geschikt is voor leerlingen van dit leeftijdsniveau. Een 12-jarige is geen miniatuur-volwassene, de structuur en inhoud van zijn denken moeten niet gezien worden als een in kwantitatief opzicht imperfecte afspiegeling van dat van een volwassene, maar die structuur en inhoud zijn in kwalitatief opzicht essentieel anders, zo leert de cognitieve ontwikkelingspsychologie ons. Niemand zal verbaasd opkijken wanneer een 12-jarige met zijn beperkte abstractievermogen niet in staat blijkt logaritmische rekenopgaven op te lossen. Is een 12-jarige, met zijn 'werkelijkheidsgebonden', concretistische denken überhaupt in staat op adequate wijze een studietekst samen te vatten ten behoeve van een bijschrift voor een tentoonstelling? De bewering dat de leerlingen onvoldoende effectief (schriftelijk)

communiceren, of zo men wil dat de leerlingen functioneel analfabeet zijn, komt in een heel ander daglicht te staan wanneer men beseft dat de leerlingen, zelfs na intensieve oefening, niet voldoende effectief *kunnen* communiceren. Een adequate beoordeling van de taalprestaties van 12-jarigen is alleen mogelijk wanneer rekening wordt gehouden met de cognitieve mogelijkheden en beperkingen van het ontwikkelingsstadium waarin die leerlingen verkeren.

Er kunnen nog meer verklaringen gegeven worden voor de tegenvallende prestaties bij de schrijf- en spreektaken, maar de hier genoemde mogen volstaan om duidelijk te maken dat boude beweringen over de aanwezigheid van functioneel analfabetisme onder leerlingen weliswaar snel zijn uitgesproken, maar bij nadere beschouwing veel minder snel kunnen worden beargumenteerd dan in eerste instantie lijkt.

8. Normering

Zoals eerder opgemerkt, worden er in het PPON-onderzoek twee criteria gehanteerd om te bepalen, of de prestaties op de verschillende taalvaardigheidsmaten het predicaat 'voldoende' dan wel 'onvoldoende' verdienen. Van het eerste (absolute) criterium (de groep als geheel beheerst een vaardigheid voldoende indien 1. de gemiddelde leerling de helft van de opgaven goed beheerst en 90% ten minste matig en 2. de zwakke leerling de helft van de opgaven matig en ten minste 10% van de opgaven goed beheerst) wordt in de *Balans*, terecht, opgemerkt dat de gegeven specificaties betrekkelijk arbitrair zijn: een strikt wetenschappelijke basis voor de caesuur voldoende/onvoldoende bestaat nu eenmaal niet. Een andere, even arbitraire invulling van dit criterium als 'Een gemiddelde leerling beheerst de helft van de opgaven goed en een groot deel, 75%, tenminste matig' leidt uiteraard tot andere, i.c. betere taalprestaties. Bij de beoordeling van de prestaties op de verschillende taalvaardigheidsmaten dient men uiteraard rekening te houden met het relatief arbitraire karakter van dit criterium.

Het relatieve criterium (de verwachtingen en wensen van drie respondentgroepen) leidt evenmin tot een dwingende typering van de taalprestaties in termen van de dichotomie voldoende/onvoldoende. Ten eerste is het zeer de vraag, of de 46 ouderleden van medezeggenschapsraden een goede inschatting kunnen maken van wat de 'gemiddelde' leerling zoal presteert. Door de bank genomen zijn het de beter opgeleiden, en meer betrokken ouders die zitting nemen in een medezeggenschapsraad. Representatief voor 'de ouders' zijn ze geenszins. Ten tweede ontbreken gegevens over de mate van overeenstemming tussen de leden van de respondentgroepen, en ten derde ontbreken gegevens over de mate van stabiliteit. Verwachtingen en wensen zijn snel uitgesproken, maar juist op het vlak van stabiliteit onderscheiden de kenners zich van de knoeiers.

Veel taalvaardigheidsprestaties moeten op basis van de gehanteerde criteria als onvoldoende worden bestempeld. Het is goed om daarbij te

beseffen, dat beide criteria relatief arbitrair zijn, temeer daar bij velen de neiging bestaat de onvoldoende prestaties toe te schrijven aan degenen die verantwoordelijk zijn voor het onderwijs.

9. Slotopmerkingen: balans van de *Balans*

In het voorafgaande is hier en daar wat commentaar geleverd op de opzet, uitvoering en resultaten van wat beschouwd mag worden als het meest grootscheepse project dat ooit in Nederland is uitgevoerd om de inhoud en opbrengsten van het taalonderwijs eind basisschool in kaart te brengen. Dat commentaar, vaak voorzien van expliciete aanbevelingen, werd niet geleverd om het project als zodanig te discrediteren, integendeel, maar met de intentie de kwaliteit ervan te optimaliseren. In een 'Ten geleide' bij de *Balans* merkt de projectleider van PPON op dat zonder overdrijving (...) gezegd (kan) worden dat deze "Balans" een indrukwekkend werkstuk is geworden. Dat is vooral te danken aan de uitvoerders van het onderzoek, die zich met grote inzet en deskundigheid van hun taak hebben gekweten. Deze opmerking kan ik onderschrijven, op één punt na. De rapportage is bepaald niet indrukwekkend te noemen. Niet alleen is de *Balans* slordig geschreven (bijvoorbeeld de niet-consistente omschrijving van de taalgebruiksfuncties) en vaak storend redundant (zo wordt keer op keer herhaald dat ouders, leraren basisschool en vwo verwachtingen en wensen uitspreken), terwijl in andere gevallen noodzakelijke informatie ontbreekt (bijvoorbeeld over de afnamecondities, over de motivatie van de leerlingen, enzovoort), ronduit ontsierend echter zijn contaminaties als 'spellingfouten' of pleonastische uitdrukkingen als 'zoals bijvoorbeeld', en dat in een rapport waarin de kwaliteit van het taalonderwijs centraal staat.

Tot slot. De *Balans* is primair bedoeld voor een inhoudelijk geïnteresseerd publiek. Tal van passages die doorspekt zijn met statistisch en/of psychometrisch jargon (logistisch itemresponsmodel, meergroepenanalyse in LISREL, variantie-analyses op logaritmisch getransformeerde scores, homogeniteitsanalyse, enzovoort) zijn uitsluitend te begrijpen door een kleine groep specialisten. Dergelijke passages kunnen naar mijn oordeel gemakkelijk gemist worden, want ze zijn helemaal niet relevant voor een goed begrip van 'de grote lijn' in de uitkomsten. Het valt te hopen dat over vijf jaar, wanneer opnieuw gerapporteerd zal worden over de inhoud en opbrengsten van ons taalonderwijs, een veel stringentere scheiding zal worden aangebracht tussen inhoud en technische details, zodat de tekst ook toegankelijk wordt voor een breed publiek. Bij die herhaalde meting zal de waarde van periodiek peilingsonderzoek, meer dan nu het geval is, pas goed tot zijn recht kunnen komen. Dan ook is de tijd aangebroken om een meer gefundeerd eindoordeel uit te spreken.

Noten

1. In principe, want krachtens het feit dat in het PPON-onderzoek gebruik wordt gemaakt van (een van de varianten van) de latente trektheorie, kan het beginsel van geheimhouding gedeeltelijk worden geschonden: per toets worden er steeds, ter informatie van de lezer, enkele (voorbeeld)opgaven verstrekt. Over vijf jaar, wanneer de peiling herhaald wordt, zullen deze openbaar gemaakte opgaven vervangen moeten worden door nieuwe; lukt het om die nieuwe opgaven zo te construeren dat ze samen met de andere, niet-openbaar gemaakte opgaven één schaal vormen, dan blijft het mogelijk om prestatieverschuivingen in de tijd na te gaan, ook al zou de moeilijkheidsgraad van de nieuw geconstrueerde opgaven verschillen van die van de openbaar gemaakte opgaven. Dit is één van de grote voordelen van de latente trektheorie, een theorie waarvan in PPON terecht gebruik is gemaakt. (Zie verder noot 2, voor een summierere uitleg van latente trektheorie.)
2. Latente trekmodellen zijn meetmodellen die ontwikkeld zijn als alternatief voor het zogenaamde klassieke meetmodel. Eén van de bezwaren die men tegen dit klassieke testmodel kan aanvoeren is dat de meettechnische kernbegrippen als betrouwbaarheid en validiteit gedefinieerd zijn als correlatiecoëfficiënt. Betrouwbaarheid en validiteit zijn daarom afhankelijk van de populatie, waarvoor zij gedefinieerd zijn, zodat elke test voor elke populatie een andere betrouwbaarheid en validiteit bezit.
Een ander fundamenteel bezwaar tegen het klassieke testmodel is, dat in dit model voorbij gegaan wordt aan de manier waarop test-scores tot stand komen. Een testscore wordt zonder meer gepresenteerd als het resultaat van een meting, zonder dat er enig meettheoretisch onderzoek heeft plaatsgevonden naar bijvoorbeeld de (schaal) eigenschappen van die scores. Latente trekmodellen proberen een aantal problemen waarmee het klassieke testmodel te kampen heeft te ondervangen door een (modeltheoretische) specificatie te geven van wat er precies gebeurt wanneer een leerling een item in een test tracht te beantwoorden. Die situatie is potentieel zeer gecompliceerd, omdat een groot aantal factoren het uiteindelijke testresultaat (item goed of fout) kan beïnvloeden (factoren als het vaardigheidsniveau van de persoon, de moeilijkheid van het item, de mate waarin er gegokt wordt, enzovoort). Om het testgedrag van een persoon te beschrijven wordt in alle latente trekmodellen een onderscheid gemaakt tussen de geobserveerde testscore en de te meten psychologische variabele (bijvoorbeeld leesvaardigheid). Aangenomen wordt dat zo'n psychologische variabele niet direct waarneembaar is, maar op de een of andere manier tot uitdrukking komt in de geobserveerde testscore. De psychologische variabele is met andere woorden 'latent', het is een latente trek die wordt voorgesteld als een continuüm waar elke persoon zijn positie op heeft.

Latente trekmodellen verschillen onderling in de mathematische specificatie van de wijze, waarop de latente trek in de geobserveerde testscore tot uitdrukking komt. Zoals elk model zijn ook latente trekmodellen eenvoudige en vereenvoudigde concepties van de werkelijkheid. Om de vraag te kunnen beantwoorden of deze modellen een redelijke beschrijving van de empirische gegevens opleveren, zijn statistische toetsen ontwikkeld. Zeker niet elke verzameling items zal aan de (doorgaans stringente) eisen van een specifiek type latente trekmodel voldoen; is dat echter wel het geval, dan vertoont de resulterende test een aantal bijzonder aantrekkelijke eigenschappen: zowel de schattingen van de moeilijkheid van de items als de schattingen van het vaardigheidsniveau van leerlingen zijn populatie-onafhankelijk. Gelet op het periodieke karakter van PPON is die populatie-onafhankelijkheid een *conditio sine qua non*.

Uit de *Balans* wordt overigens niet echt duidelijk van welk type latente trektheorie gebruik is gemaakt. Ook al wordt er gesproken over het algemene één-parametermodel (?), op grond van de summiere uitleg (p.27-28) kan worden afgeleid dat niet het conventionele één-parameter model wordt bedoeld (het zogenaamde Rasch-model), maar het twee-parameter model (het zogenaamde Birnbaum-model).

3. Een opgave goed respectievelijk matig beheersen betekent, dat de kans op succes 80% respectievelijk 50% bedraagt.
4. De 'feasibility study', het haalbaarheidsonderzoek dat vooraf ging aan de periodieke peiling en dat werd uitgevoerd door het SCO, haalde de landelijke pers met de wel zeer vergaande conclusie dat 5 tot 7% van de 12-jarigen functioneel analfabeet genoemd moet worden. Verderop in dit artikel probeer ik aan te tonen, dat de wetenschappelijke status van dergelijke uitspraken gering is.
5. De studenten kregen de zes voorbeeldteksten in willekeurige volgorde aangeboden. Hen werd verzocht elk van de zes teksten in één categorie onder te brengen. De omschrijvingen van de categorieën waren letterlijk overgenomen uit de *Balans* (ook die van de categorie Directief, hoewel geen enkele voorbeeldtekst tot dat type behoorde). De studenten vonden de taak moeilijk tot zeer moeilijk; volgens hun zeggen was het bij vele teksten niet mogelijk deze aan slechts één categorie toe te wijzen. Dat verklaart ook, waarom er in de tabel bij de voorbeeldteksten per rij meer dan elf antwoorden staan.
6. Het pragmatisch uitgangspunt is dan ook bij uitstek geschikt voor het operationaliseren van taken bij het schrijven en spreken. De bedoelde intentie, i.e. de illocutie, kan immers via de instructie aan de leerlingen worden opgelegd. Maar ook bij het schrijven en spreken is de uitwerking van het pragmatisch uitgangspunt onzorgvuldig en inconsequent. Voor deze twee produktieve vaardigheden worden

zeven taalhandelingen onderscheiden: 1. informatie doorgeven 2. beschrijven 3. samenvatten 4. gevoelens/evalueren 5. instrueren 6. informatie vragen 7. argumenteren/overtuigen. Al deze zeven onderscheiden handelingen zijn helemaal geen taalhandelingen, althans niet in de betekenis die daaraan binnen de pragmatiek wordt gegeven. In de pragmatiek worden taalhandelingen opgevat als elementaire, abstracte handelingen die corresponderen met het niveau van de zin, en niet met dat van de tekst. Binnen het PPON-onderzoek worden taalhandelingen echter opgevat als taalhandelingscomplex, en dat compliceert de notie 'taal-handeling' in aanzienlijke mate. Het fraaie van die elementaire notie is immers, dat er (constituerende) regels, semantische conventies geformuleerd kunnen worden ('felicity conditions') voor een geslaagde uitvoering van de betreffende taalhandeling (bijvoorbeeld: het uiten van een belofte geldt alleen als een belofte indien de luisteraar liever zou hebben dat de spreker de in de propositionele inhoud tot uitdrukking gebrachte handeling wel doet dan dat hij dat niet doet en wanneer het niet al voor spreker en luisteraar bij voorbaat duidelijk is dat de spreker bij een normaal verloop van de gebeurtenissen de betreffende handeling zou uitvoeren). Het elementaire begrip 'taalhandeling' is onlosmakelijk verbonden met dit type regels (propositionele inhoudsregel, voorbereidende regel, oprechtheidsregel en essentiële regel); zonder specificatie van dit type regels verliest het begrip 'taalhandeling' zijn betekenis. Tot nu toe is het binnen de pragmatiek niet gelukt om voor taalhandelingscomplexen zulke regels te formuleren, uitgezonderd voor het complex 'argumentatie'. Verder vallen de 'taalhandelingen' nr. 3 en 4 (samenvatten en gevoelens/evalueren) onder een aparte categorie 'beschouwend taalgebruik', maar daar horen ze helemaal niet thuis - ze horen bij de categorie 'rapporterend'. En bovendien wordt het indelingsprincipe aangetast door de 'taalhandeling' nr. 7 (argumenteren/overtuigen), waarbij de perlocutie, en niet meer de illocutie centraal staat.

Bibliografie

Zwarts, M. (red.), *Balans van het taalonderwijs aan het einde van de basisschool. Uitkomsten van de eerste peiling einde basisonderwijs*. Arnhem: Cito, 1990.

(manuscript aanvaard: 18 maart 1991)