

**HET BEOORDELEN VAN
(MONDELINGE) TAALVAARDIGHEID**
Een reactie op H. Wesdorp's 'Evaluatie-
technieken voor het moedertaalonderwijs'

**toetsings
problematiek**

Moedertaalonderwijs en (objectieve) toetsen. De discussie daarover kent al een hele Moer-traditie. In jaargang 1973 kruisten Bonset, Griffioen, Van Gils en Prick, Wesdorp de degen over het toetsen van luistervaardigheid. In jaargang 1974 en 1975 ging de discussie verder over het toetsen van tekstbegrip (Bonset en Luijten). In jaargang 1976 kwam Van Gils daar nog eens op terug. Nu, jaargang 1979, vatten Henk Lammers en Leo Lentz de draad van de discussie weer op naar aanleiding van het boek Evaluatietechnieken voor het moedertaalonderwijs van H. Wesdorp. In hun kritische bespreking van dit boek gaan Lammers/Lentz vooral in op de toetsingsproblematiek van de mondelinge taalvaardigheid. Naar hun idee botsen de normen voor goed (normaal functioneel) onderwijs met normen voor betrouwbare (in toetstechnisch opzicht) toetsing. Bij het normaal functioneel moedertaalonderwijs verkeer je vaker in besprekings- dan in beoordelingssituaties.

Inleiding

Augustus 1978 publiceert H. Wesdorp het eindverslag van een studie naar het meten van het taalvermogen met als ondertitel 'Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luisteren en discussievaardigheid'. Wesdorp is verbonden aan het Research Instituut voor Toegepaste Psychologie (RITP) en heeft zijn onderzoek kunnen verrichten dankzij een subsidie van de Stichting voor Onderzoek van het Onderwijs (SVO). Het onderzoek is nog niet voltooid.

Wesdorp zelf beschouwt zijn verslag als een tussenstand en spreekt van een materiaalverzameling die nog een lange weg moet gaan voor het gewenste eindproduct er is. Dat eindproduct zou dan 'een voor de onderwijspraktijk en/of de onderwijsresearch bruikbare handleiding betreffende de problematiek van de taalvaardigheid-evaluatie' moeten zijn (Wesdorp 1978: 1). Om nu tot een zo goed en bruikbaar mogelijk eindrapport te komen, legt Wesdorp deze tussenstand voor aan een aantal deskundigen: vakspecialisten en didactici aan Pedagogische Academies, Universitaire en Nieuwe Lerarenopleidingen. Opmerke-

lijk is dat in deze opsomming de leerkrachten ontbreken.

Januari 1979 stuurt Wesdorp een brief aan degenen die in eerste instantie het rapport toegestuurd kregen. Hij schrijft in deze brief dat bij nader inzien de doelgroep verder beperkt moet worden: het boek blijkt niet geschikt voor studenten van de opleidingsinstituten. Hij deelt mee aan een beknopt boekje voor studenten te werken. Het rapport zelf is op dit moment niet meer verkrijgbaar maar Wesdorp spreekt de hoop uit dat het binnenkort weer op ruimere schaal verkrijgbaar is.

Het rapport heeft vier, ongeveer even lange hoofdstukken die gewijd zijn aan schrijven, lezen, spreken/discussiëren en luisteren. Elk hoofdstuk begint met een algemene inleiding op de aard van de vaardigheid en de problemen die hieruit voortkomen voor de beoordeling.

Vervolgens worden verschillende beoordelingsmethoden aan de hand van literatuur besproken. Deze literatuur doet meestal verslag van wetenschappelijk onderzoek; weinig voorbeelden komen uit de lespraktijk.

Wesdorp beperkt zich tot de evaluatie van de taalvaardigheid van de leerlingen in het voortgezet onderwijs en in de hogere klassen van het basisonderwijs. Een aantal beoordelingsmethoden die aan bod komen zijn de cloze-techniek¹, de samenvatting, beoordelingsschalen, open vragen. In elk hoofdstuk wordt onderscheid gemaakt tussen globale en analytische beoordeling en komt de mogelijkheid van objectieve toetsing ter sprake. De laatste paragraaf van elk hoofdstuk gaat over de affectieve aspecten van de vaardigheid. Het eerste deel van hoofdstuk 1, over schrijfvaardigheid, kan beschouwd worden als een algemene inleiding. Wesdorp bespreekt hier bronnen van onbetrouwbaarheid, een bespreking waarop hij in de latere hoofdstukken regelmatig terugkomt.

Een visie is nodig

Wesdorp doet een poging een overzicht te geven van de beschikbare evaluatietechnieken voor taalvaardigheid. Dat moet positief gewaardeerd worden, ook omdat Wesdorp bereid is gebleken daarvoor een grote hoeveelheid literatuur in beschouwing te nemen. Daarbij willen we echter een paar opmerkingen maken. Door de hoeveelheid en de verscheidenheid van de literatuur wordt de in-

druk van objectiviteit gewekt. Die is er natuurlijk niet. Wesdorp maakt wel degelijk keuzes en doet dat vanzelfsprekend op grond van zijn ideeën over het toetsen van taalvaardigheid. Deze visie wordt echter niet expliciet gemaakt en niet ter discussie gesteld. De tekst werkt echter, tussen de regels door, wel degelijk opiniërend. Hoe die opinie er dan precies uitziet moeten we evenwel afleiden uit passages als de volgende.

Het is echter de vraag of de eis van de normale functionaliteit, die terecht aan onderwijssituaties gesteld mag worden, ook per se moet gelden voor toetsingssituaties (p. 255).

Het is naar onze mening duidelijk dat men bereid moet zijn in de toetsituatie enige concessies te doen ten aanzien van de natuurlijkheid, om op andere punten voordeel te behalen (p. 256).

Betrokkenheid van de leerkracht als beoordelaar kan negatieve effecten hebben. Sterker nog: kan tot meer interactie leiden tussen leraar en leerling, en deze interactie kan de beoordelingsbetrouwbaarheid negatief beïnvloeden (p. 256-257).

Uit deze passages kunnen we concluderen dat de betrouwbaarheid voor Wesdorp zwaar weegt en dat hij bereid is daarvoor offers te brengen.

De hamvraag is natuurlijk: onder welke omstandigheden is zo'n offer verantwoord en welke overwegingen moeten daarbij betrokken worden. Tot een afweging van de verschillende argumenten komt Wesdorp echter niet. Hij licht niet toe wanneer we die eis van normale functionaliteit² mogen laten vallen; hij geeft niet aan welke concessies acceptabel zijn. Juist een publikatie die in een handleiding moet resulteren zou aan de leerkracht duidelijk moeten maken dat het antwoord op dergelijke vragen afhangt van het *soort* toetsing. Het moet duidelijk worden dat een criterium als bijvoorbeeld betrouwbaarheid hoog genoteerd staat als het gaat om een centrale toets met het oog op selectie.

Waar Wesdorp een poging doet die duidelijkheid te verschaffen, wint zijn betoog aan inzichtelijkheid. Als voorbeeld noemen we het schema op pagina 254 waarin de samenhang tussen enerzijds de vrijheid van de beoordelaar en die van de leerling en anderzijds de beoordelingsmethode wordt aangegeven. Uit dat schema blijkt dat betrouwbaarheid en validiteit³ moeilijk te combineren zijn en dat er keuzen gemaakt moeten worden. Het schema helpt de leerkracht bij het maken van die keuze.

Twee visies

Wie, zoals Wesdorp, enige orde wil brengen in de problemen rond het toetsen van taalvaardigheid, die heeft een organiserend principe nodig: een visie op het taalvaardigheidsonderwijs, een theorie. Zo'n visie is een noodzakelijke voorwaarde om te kunnen kiezen tussen de verschillende beoordelingsmethoden. Maar er is nog een voorwaarde: het expliciteren van die visie. Als dat niet gebeurt, is een zakelijke discussie over beoordelingsmethoden onmogelijk. Het is dan vaak moeilijk vast te stellen of bezwaren zich richten tegen de methode zelf of tegen de achterliggende visie. Omdat Wesdorp zijn visie maar in beperkte mate duidelijk maakt, doen we hier een poging tot reconstructie. Daarbij dreigt het gevaar dat we Wesdorp onrecht aandoen, maar dat risico loopt een auteur die het beschrijven van zijn visie overlaat aan zijn critici, nu eenmaal.

In de meeste onderzoeken waar Wesdorp zijn studie op baseert, wordt veel belang gehecht aan een zo groot mogelijke betrouwbaarheid van een toets. De achterliggende filosofie daarbij is: als we een betrouwbare toets hebben kunnen we eindelijk eens meten wat het effect van ons onderwijs is en zo onderscheid maken tussen goed en slecht onderwijs. Binnen deze traditie is het organiserend principe dus: betrouwbaarheid terwille van de effectmeting. Wij beschouwen Wesdorp als een aanhanger van deze filosofie, mede op grond van zijn *Voorwoord* waar hij schrijft: 'Zonder een acceptabel evaluatie-instrumentarium is elke ontwikkeling, vernieuwing of verbetering van het moedertaalonderwijs een imaginair zaak' (p. 3). Dit citaat leidt niet onherroepelijk tot onze conclusie, maar onze interpretatie wordt ondersteund door het volgende. Wie het effect van een bepaalde taalvaardigheidsmethode wil meten, moet allerlei storende elementen tot een minimum beperken. Zowel de te toetsen vaardigheid als de toetssituatie moeten zo 'zuiver' mogelijk zijn. In dit licht bezien is het niet verwonderlijk dat Wesdorp zich behoudend opstelt ten aanzien van een tweetal vernieuwingen in het moedertaalonderwijs die afbreuk doen aan deze 'zuiverheid':

... het gebruik van informatiebronnen, een vaardigheid die met veel recht een moedertaalvaardigheid genoemd zou kunnen worden, valt buiten het bestek van dit rapport (p. 2).

Met onze indeling naar de vier taalvaardigheids-

aspecten sluiten wij ons aan bij de traditie; maar menig auteur zet vraagtekens bij die traditie en vraagt zich af of de geconstateerde verschillen – bijvoorbeeld tussen de ontwikkeling van de spreek- en schrijfvaardigheid – geen artefacten zijn van de gehanteerde meetmethoden (kanaalar-tefacten) of van het onderwijssysteem (didactische artefacten) (p. 1).

Wesdorp vraagt zich in het voorwoord af of de keuzen die hij gemaakt heeft de bruikbaarheid van het rapport niet hebben aangetast. Wij zijn geneigd die vraag met ja te beantwoorden. Wij denken dat zijn visie niet het juiste kader biedt voor de verbetering van het taalvaardigheidsonderwijs.

Dat onderwijs is te zeer procesmatig om alleen het produkt als criterium voor de kwaliteit van het onderwijs te kunnen nemen. Wij denken dat het niet zozeer Wesdorp is die faalt als wel de onderzoekstraditie waarop hij zich baseert. We vinden steun voor deze mening bij Meuffels (1978: 58-59) die er in een artikel over effectmeting op wijst dat van breed opgezet empirisch onderzoek weinig te verwachten valt als het theoretisch kader voor dat onderzoek ontbreekt. Ook Diegritz (1977: 1) komt tot een dergelijke conclusie. Hij wijst erop dat we ondanks een onmetelijk aantal empirisch-analytische onderzoeken nu nog niet meer weten over bijvoorbeeld het verloop van communicatieve processen in het onderwijs, dan twintig jaar geleden. Hij geeft aan dat veel verschijnselen die niet meet- en kwantificeerbaar zijn buiten beschouwing bleven zodat we nu voornamelijk de beschikking hebben over een onsamenvattend geheel van resultaten uit het effect-onderzoek.

In onze visie⁴ is taalvaardigheid het kunnen verwerven en verwerken van informatie. Dat betekent dat een leerling in staat moet zijn aan de hand van een concrete vraag informatiebronnen (personen, instanties, encyclopedieën, boeken, tijdschriften, kranten) te raadplegen; daaruit informatie te verzamelen en deze informatie zo te ordenen en te presenteren dat ze opnieuw toegankelijk is voor anderen. Bij vaardigheden als lezen en luisteren ligt het accent dan op het verwerven van de informatie en op het beoordelen van het belang/de kwaliteit van de informatie. Bij vaardigheden als spreken en schrijven ligt het accent op de ordening en de presentatie van de informatie waarbij de leerling bijvoorbeeld in

staat moet zijn rekening te houden met zijn publiek.

In onze benadering van het moedertaalonderwijs staat het omgaan met informatie in reële taalgebruikssituaties centraal. Dit betekent dat de informatie werkelijk van belang is voor de leerlingen en dat er praktische consequenties aan verbonden zijn. Enkele voorbeelden.

Het schrijven van een brief aan de ouders (doel: informatie geven over de plannen voor een klasavond; belang: toestemming van de ouders; praktische consequentie: het al dan niet doorgaan van de avond). Een artikel voor de schoolkrant (doel: standpunt uiteenzetten over de censuur van de rector in de vorige editie; belang: persvrijheid op school; praktische consequentie: een op gang gebrachte discussie). Het lezen van een artikel uit de krant en een samenvatting maken als je het voor je werkstuk wilt gebruiken. Een gesprek met de leraar wiskunde voorbereiden om een herkansing voor een proefwerk te bepleiten.

In deze taalactiviteiten komen vele 'technische' aspecten van taalvaardigheid aan de orde. Deze kunnen zeker apart behandeld en beoordeeld worden mits duidelijk is met welk doel de activiteiten verricht worden. Als dat duidelijk is, wordt de beoordeling al een stuk eenvoudiger: is mijn brief helder, plezierig leesbaar; heb ik dat artikel uit de krant zodanig weergegeven dat de samenvatting de informatie bevat die ik ook voor mijn werkstuk nodig heb? Kortom, de beoordeling richt zich op de vraag: 'Heb ik mijn doel bereikt en heb ik dat op een adequate manier aangepakt?' Het is duidelijk dat hier niet langer sprake is van taalvaardigheid in engere zin, maar eerder van communicatieve vaardigheid. Welk etiket we er ook op plakken, het is belangrijk dat deze visie een aantal consequenties heeft die verschillen van die van Wesdorp. Ter illustratie bespreken we er één, de betrouwbaarheid.

De betrouwbaarheid

We gebruiken hier het woord betrouwbaar in een bijzondere betekenis: een toetstechnische. Betrouwbaar betekent dan: in hoeverre krijg je in meerdere toetsituaties met dezelfde toets dezelfde resultaten. Gewoonlijk kan de betrouwbaarheid vergroot worden door de toetssituatie te structureren. Doen we dit dan dreigt het gevaar dat we alleen minder relevante vaardigheden kunnen toetsen. Wesdorp schrijft over dit probleem

bij het toetsen van luistervaardigheid:

Perren (1968) constateert in dit verband, dat we vaak wél in staat zijn te discrimineren (betrouwbaar te meten) t.a.v. zaken die minder belangrijk zijn, terwijl we slechts onbetrouwbare oordelen kunnen vellen over zaken die echt wél belangrijk zijn. Het dilemma is dat tussen betrouwbare meting van irrelevante vaardigheden en onbetrouwbare 'meting' van relevante vaardigheden (p. 256).

Dit dilemma blijkt voor Wesdorp zwaar te wegen. Bij elke vaardigheid zien we opnieuw de worsteling om een zo eerlijk mogelijk evenwicht te vinden tussen betrouwbaarheid en relevantie.

Een bruikbare handleiding voor de onderwijspraktijk dient uit te gaan van de keuzes die de leerkracht gewoonlijk in dit dilemma maakt. Wij denken dat leerkrachten bij het oefenen van een vaardigheid proberen evaluerende opmerkingen te maken die een leerling zo veel mogelijk inzicht geven in de manier waarop hij met een vaardigheid omgaat. Pas in tweede instantie zal hij zich bekommeren om de betrouwbaarheid van zijn oordeel. Als er in het onderwijs voortdurend concessies worden gedaan aan de betrouwbaarheid, dan is het niet juist om bij het samenstellen van een handleiding concessies te doen aan de relevantie en de natuurlijkheid. Er ontstaat dan een wrijving tussen toetsing en onderwijs. Dat dit gevaar dreigt, blijkt uit Wesdorps standpunt ten aanzien van een aantal bronnen van onbetrouwbaarheid. Deze zijn de variabiliteit van de beoordelaar, de verschillen in de testsituatie en de variabiliteit van de beoordeelde. Wesdorp waardeert deze bronnen negatief en doet voorstellen hoe ze te elimineren (Wesdorp 1978: 252-255). Overigens leiden deze voorstellen tot een verlies aan validiteit. Een betrouwbaar en valide oordeel over zulke complexe vaardigheden als bijvoorbeeld het luisteren, blijkt een illusie.

Naar een handleiding

Wesdorp heeft met dit rapport een eerste stap willen zetten naar een handleiding voor de evaluatie van taalvaardigheid. Welke informatie heeft een onderzoeker nodig om zo'n handleiding te maken?

Uiteraard dient hij op de hoogte te zijn met de evaluatiemethoden die in de vakliteratuur beschreven worden. Maar hij dient óók een visie op het taalvaardigheidsonderwijs te hebben die aan-

sluit bij wat er in de praktijk van het onderwijs gebeurt. Kennelijk moet daarvoor meer gebeuren dan het enquêteren van leerkrachten. Want van de problemen zoals die duidelijk werden uit een door Wesdorp (1974) gehouden enquête, vinden we in dit rapport opmerkelijk weinig terug. De leerkracht moet dan ook van ondervraagde tot deelnemer worden. Dit vereist dat de deskundigheid van de leerkracht gewaardeerd en gebruikt wordt. Wij denken dat dit nu nauwelijks het geval is. Zelfs in een 'progressief' boek als *Taaldidactiek aan de basis* treffen we nog een passage als de volgende aan:

Aan kritiek op het gebruik van (objectieve) toetsen in het onderwijs heeft het de laatste jaren niet ontbroken. Voor een deel is ze zeker gerechtvaardigd: het samenstellen van een goede (is: betrouwbare en valide) toets is een moeilijk karwei, dat een grote deskundigheid vraagt. En zelfs toetsen die door deskundigen (bijvoorbeeld door het Centraal Instituut voor Toetsontwikkeling (CITO) in Arnhem) gemaakt zijn, voldoen nog niet altijd aan de te stellen eisen. Niettemin geloven we dat de critici zich in veel gevallen beter zouden kunnen richten tegen de toetsgebruikers, die zich lang niet altijd voldoende rekenschap geven van wat wel en wat niet met een bepaalde toets kan worden gemeten en die aan de uitkomst soms conclusies verbinden die de toetsontwerpers zeker niet voor hun rekening zouden willen nemen. Wat ooit is opgemerkt ten aanzien van een scheermes, namelijk dat het een nuttig instrument is voor wie er verstandig mee omgaat, maar een gevaarlijk wapen kan worden in handen van een dwaas, geldt mutatis mutandis ook voor een toets.

We zijn het hier maar gedeeltelijk mee eens omdat de zaak in werkelijkheid natuurlijk niet zo simpel ligt. Uit Wesdorp (1974) en Driest e.a. (1975) blijkt hoe invloedrijk de luistertoetsen in de praktijk zijn. Het gaat ons wat te ver om al die leraren als dwazen te bestempelen. Wij geloven dat de leraren helemaal niet zo'n vrije keus hebben: de door het CITO vervaardigde toetsen zijn nu eenmaal geen scheermessen die iedere leerkracht willekeurig kan laten liggen of oppakken. Volgens ons bevindt de leerkracht zich in de volgende situatie. Enerzijds kan hij zelf bepalen wat er gebeurt; anderzijds is er pressie (tijdgebrek, sectiebeleid, schoolleiding) om de bestaande toetsen te gebruiken. Hoe de uiteindelijke beslissing dan uitvalt, hangt in grote mate af van het zelf-

vertrouwen van de leerkracht, van het geloof in eigen kunnen. Maar als we nagaan hoe je als leerkracht gewaardeerd wordt, dan lijkt het ons erg moeilijk om nog veel geloof in eigen kunnen over te houden. Ten aanzien van de toetsingsproblematiek wordt er een situatie gecreëerd waarbij de verantwoordelijkheid en de deskundigheid van de leerkracht in de verdrinking komt. Wordt in de eerst alinea van bovenstaand citaat niet uitgelegd dat het maken van een goede toets een moeilijk karwei is? Het vraagt zo'n grote deskundigheid dat zelfs de deskundigen van het CITO het (nog) niet kunnen. Wat wil je als leerkracht dan nog? Natuurlijk zijn leraren wel deskundig. Het is alleen een ander soort deskundigheid. Gewoonlijk zullen leraren niet weten hoe ze een item-analyse moeten uitvoeren maar ze zullen wel degelijk kunnen aangeven of, bijvoorbeeld, een toets aansluit bij het onderwijs dat ze (willen) geven. De eerste soort deskundigheid is belangrijk voor degenen die een toets met een hoge graad van betrouwbaarheid voorstaan. De tweede soort deskundigheid is belangrijk voor degenen die hun onderwijs zo veel mogelijk normaal functioneel willen houden. Wij constateren dat deze twee soorten deskundigheid nogal eens verward worden en dat leerkrachten wegens het ontbreken van de eerste soort deskundigheid ten onrechte als ondeskundig worden behandeld.

Wesdorp zelf is niet optimistisch over de mogelijkheid zijn rapport nog te bewerken tot een bruikbare handleiding. We delen zijn pessimisme. Niet alleen omdat er nog veel moet gebeuren en het een bijzonder moeilijke taak is. Maar ook omdat het probleem eerst opgelost kan worden als Wesdorp het roer omgooit en voor een principieel andere aanpak kiest.

Noten

- 1 Bij de cloze-techniek wordt uit een tekst bijvoorbeeld elk tiende woord weggelaten. De leerlingen moeten deze open plekken in de tekst invullen. Hoe hoger het aantal juiste invullingen, des te hoger het tekstbegrip.
- 2 Normale functionaliteit: 'Dit principe komt erop neer dat je je onderwijs zo inricht, dat je leerlingen er iets van leren dat ze, naar eigen oordeel, a. óf op korte termijn praktisch kunnen gebruiken, b. óf 'alleen maar' boeiend vinden' (Ten Brinke 1977).
- 3 Een toets is betrouwbaar als we bij herhaalde afnames onder vergelijkbare omstandigheden bij dezelfde

leerlingen dezelfde resultaten krijgen.

Een toets is valide als we inderdaad meten wat we willen meten.

- 4 'Onze' betekent hier natuurlijk dat we aansluiten bij en een verdere uitwerking proberen te geven aan de ideeën zoals we die vinden in een project als Gericht Schrijven, Griffioens *Zeggenschap* en Ten Brinkes *The complete mother tongue curriculum*.

Literatuur

S. ten Brinke *The complete mother tongue curriculum. A tentative survey of all the relevant ways of teaching the mother tongue in secondary education* Groningen 1977

T. Diegritz en H. Rosenbusch *Kommunikation zwischen Schülerns. Schulpädagogische und linguistische Untersuchungen. Didaktische Konsequenzen* München 1977

J.P. Diest e.a. 'Luistervaardigheid in het moedertaalonderwijs' in: *Levende Talen* 312 (1975), p. 242-250

B. Meuffels 'Effectonderzoek taalvaardigheid' in: *Spektator* 1978-1979, nr 1/2, p. 47-61

Nijmeegse Werkgroep Taaldidactiek *Taaldidactiek aan de basis* 2e druk, Groningen 1978

H. Wesdorp *Schoolonderzoek-methoden voor het vak Nederlands op het Havo. Deel I: enquête-resultaten* Amsterdam, RITP, 1974

H. Wesdorp *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid* Amsterdam, RITP, 1978

STERKE STAALTJES?

STEVIGE STAPPEN?

Naar:

Fie van Dijk

Straatweg 222, Breukelen