

Boekbespreking

Edith Vissers

Het beoordelen van spreekvaardigheid

Een bespreking van:

Amos van Gelderen, *De evaluatie van spreekvaardigheid in communicatieve situaties, globale beoordeling en gedetailleerde analyse van spreekprestaties van elf- en twaalfjarigen*. Amsterdam SCO-rapport, Universiteit van Amsterdam (proefschrift), 1992, ISBN 90-6813-347-0.

1. Inleiding

In de jaren '60 ontstaat voor het eerst aandacht voor het toetsen van mondeling taalgebruik. De eerste mondelinge productieve taalvaardigheidstoetsen worden afgenomen in een soort van laboratoria: in kleine 'studioachtige' ruimten wordt bijvoorbeeld van individuen verwacht dat zij woorden en zinnen herhalen die ze te horen krijgen (Shohamy, 1988).

Vanaf de jaren '70 begint de kritiek op een dergelijke vorm van het toetsen van spreekvaardigheid. Er zou bij het toetsen van spreekvaardigheid meer moeten worden uitgegaan van realistische communicatieve situaties. Canale en Swain wijzen er bijvoorbeeld later op dat het toetsen van spreekvaardigheid in een tweede taal zo vormgegeven moet worden, dat individuen hun mate van 'performance' kunnen aantonen in betekenisvolle contexten (Canale en Swain, 1980).

Sinds enkele jaren wordt met behulp van een dergelijke toetsvorm in Nederland binnen landelijk peilingsonderzoek, de zogenaamde Periodieke Peiling van het Onderwijsniveau (PPON), de spreekvaardigheid van leerlingen in het basis- en voortgezet onderwijs geëvalueerd.

Het peilingsonderzoek heeft een aantal doelen. In de eerste plaats is het bedoeld om informatie te verzamelen over de effectiviteit van het onderwijs. In de tweede plaats moet het peilingsonderzoek een empirische basis geven aan de voortdurende discussies over het onderwijsniveau en over de vraag die daaraan verbonden wordt: "Moet het onderwijsniveau verbeterd worden?" Ten derde is het onderzoek er op gericht onderwijsgevenden en onderwijsonderzoekers handreikingen te geven om het onderwijs te verbeteren.

2. Doel van de studie

Het doel dat in het proefschrift van Van Gelderen centraal staat, is een schema op te stellen voor het evalueren van spreekvaardigheid waarbij zo optimaal mogelijk tegemoet gekomen wordt aan de drie genoemde doelen van het periodieke peilingsonderzoek. Meer concreet probeert Van Gelderen voor de evaluatie van spreekvaardigheid een algemeen beoordelingsschema te ontwikkelen dat het mogelijk maakt om *betrouwbare* en *valide* oordelen over verschillende aspecten van de vaardigheid te ontwikkelen en dat tegelijkertijd tegemoet komt aan de eisen die vanuit het oogpunt van *efficiëntie* gesteld worden aan grootschalig onderzoek.

3. Opbouw van de studie

De evaluatie van spreekvaardigheid in communicatieve situaties is opgebouwd uit vier delen.

Deel I bevat een algemene inleiding waarin de achtergronden van het onderzoek beschreven worden, waarin de probleemstelling wordt uitgelegd en waarin een databeschrijving wordt gegeven.

In deel II wordt een theoretische grondslag gegeven voor het ontwikkelen van een beoordelingsschema en er wordt verslag gedaan van enkele kleinschalige experimenten en grootschalige studies die ten doel hebben een beoordelingsschema te ontwikkelen voor een betrouwbare, valide en efficiënte evaluatie van spreekvaardigheid. Het resultaat van wat beschreven is in deel II, is een verantwoording voor de beoordeling van spreekvaardigheid bestaande uit vier dimensies.

In deel III wordt verslag gedaan van een empirisch onderzoek om te bepalen in welke mate het oordeel over elke dimensie voorspeld kan worden door de *bedoelde* aspecten (de convergente predictie) en daarnaast of deze aspecten dat oordeel beter voorspellen dan de andere, *niet-bedoelde* aspecten (de divergente predictie). Deze informatie is van belang om vast te kunnen stellen in hoeverre de beoordeling van de vier dimensies zinvol is voor de evaluatie van spreekvaardigheid.

In deel IV ten slotte wordt ingegaan op omstandigheden waaronder het onderzoek heeft plaatsgevonden en die de generaliseerbaarheid van de conclusies kunnen beperken. Bovendien wordt ingegaan op de vraag welke soorten van vervolgonderzoek wenselijk zijn om aan die beperkingen het hoofd te bieden. Ten slotte worden de toepassingsmogelijkheden van de gerapporteerde studie besproken. Er wordt besproken op welke wijze het beproefde schema functioneel kan zijn in de context van grootschalig peilingsonderzoek bij de beoordeling van spreekvaardigheid en op welke wijze het beoordelingsschema op micro-niveau kan worden ingezet om diagnostisch om te gaan met spreekvaardigheid in de klas.

4. Inhoud van de studie

In de theoretische onderbouwing van zijn voorgestelde beoordelingsschema gaat Van Gelderen allereerst in op verschillende vormen van analytische beoordeling. Voor grootschalig peilingsonderzoek beschouwt Van Gelderen een dergelijke manier van beoordelen echter niet geschikt omdat het maar de vraag is of in een grootschalig onderzoek de behoefte bestaat aan informatie over allerlei deelaspecten van de vaardigheid. Bovendien treedt met een analytisch beoordelingsschema het gevaar op dat de verschillende beoordelingscategorieën meettechnisch niet onderscheidbaar zijn. Van Gelderen onderschrijft hier de mening van Wesdorp die opmerkt dat beoordelaars niet in staat zijn meer dan drie of vier onderscheidingen te maken:

"Vaak bestaan analytische schema's uit véél subcategorieën, die elkaar overlappen, niet echt onafhankelijk kunnen worden beoordeeld, waardoor een schijnprecisie wordt gesuggereerd. Bij nadere analyse blijken dan bijvoorbeeld in het gunstigste geval slechts 3 of 4 factoren te onderscheiden waaruit volgt dat waarschijnlijk vaak overbodig werk wordt verricht: onder invloed van het halo-effect tot stand gebrachte herhalingen van reeds eerder gevelde oordelen." (Wesdorp, 1981)

Bovenstaande overwegingen zijn voor de onderzoeker reden om te zoeken naar mogelijkheden het grote aantal aspecten en criteria uit de analytische schema's samen te vatten in een meer globaal schema. Van Gelderen gaat in eerste instantie uit van het model van Becker (Becker, 1962) waarin drie dimensies onderscheiden worden. Hij komt echter tot de conclusie dat een dergelijke driedeling niet toereikend is voor een theoretisch verantwoorde indeling van beoordelingscriteria. Van Gelderen herformuleert daarom Beckers driedeling tot een vierdeling van dimensies. Hieronder zijn die dimensies met de bijbehorende beoordelingscategorieën schematisch weergegeven.

	Referentie	Overdracht	Spreekgemak	Verstaanbaarheid
Beoordelings-categorieën	relevantie organisatie verbanden	intonatie spreektrant oriëntatie	zinsbouw spreektempo woordkeus	articulatie

Referentie Hiermee wordt de symboolfunctie van taal bedoeld. Waarnaar verwijzen de uitingen van de spreker, wat is de betekenis van hun onderlinge samenhang in onderscheiden delen van de gesproken tekst en welk belang hebben ze voor de luisteraar?

Overdracht De overdrachtsdimensie heeft betrekking op al het taalgebruik en het non-verbale gedrag dat bedoeld is om de relaties tussen spreker enerzijds en luisteraar en onderwerp anderzijds uit te drukken.

Spreekgemak Van Gelderen verstaat hieronder het gebruik van middelen om de continuïteit van spraak te behouden. Sprekers moeten in het algemeen de indruk vestigen over voldoende beheersing van vocabulaire en syntax van de taal te beschikken om aan het woord te kunnen blijven, en aldus hogere functies van Referentie en Overdracht te kunnen vervullen.

Spreekgemak is niet substantieel met het criterium communicatieve effectiviteit verbonden maar meer voorwaardelijk. Van Gelderen onderscheidt drie functies van Spreekgemak: vlotheid, beheerstheid en vloeiendheid.

Verstaanbaarheid Verstaanbaarheid wordt bepaald door de mate waarin de luisteraar de uitingen van de spreker moeiteloos kan herkennen als betekenisvolle lexemen. Verstaanbaarheid is voorwaardelijk voor de effectiviteit van de communicatie.

Een kleinschalig experiment

In een eerste kleinschalig experiment gaat de onderzoeker na of uit de vier dimensies zinvolle categorieën voor beoordeling kunnen worden afgeleid die tot een bevredigende overeenstemming tussen beoordelaars leiden.

De vier taken die Van Gelderen voor zijn onderzoek gebruikt zijn de volgende:

taak 1: een verhaal navertellen;

taak 2: een ongeval melden;

taak 3: beschrijven hoe een spin zijn web weeft;

taak 4: een verhaal afmaken.

Uit steekproeven zijn voor iedere taak 40 prestaties geselecteerd. De beoordelaars kregen een instructie waarin de taken werden uiteengezet en de beoordelingscategorieën toegelicht. Alle categorieën zijn aan de beoordelaars gepresenteerd in de vorm van vijf puntsschalen. De extremen en het midden van deze schalen waren gemarkeerd door korte omschrijvingen. De categorie Relevantie bijvoorbeeld had als omschrijvingen: "hoofd- en bijzaken zijn goed onderscheiden" (5 punten), "vergeet wel wat" (3 punten) en "de grote lijn komt er niet uit" (1 punt).

Van Gelderen stelt na zijn eerste experiment dat de beoordelaarsovereenstemming bij de categorieën voor de taken 1 en 4 vrij hoog is. Voor het merendeel van de categorieën zou men bij het beoordelen van prestaties op de verschillende spreekttaken bij een gewenste betrouwbaarheid van .80 kunnen volstaan met een jury van drie beoordelaars. Bij de beoordeling van Overdracht en Spreekgemak bij de taken 2 en 3 zijn de overeenstemmingen over het algemeen lager: er zou in die gevallen een grotere jury nodig zijn.

Een grootschalige studie

Van Gelderen doet verslag van twee studies naar het gebruik van het schema voor beoordeling van prestaties op de vier spreektaken. Het belangrijkste doel van deze studies is te bepalen of (en in hoeverre) de vier beoordelingsdimensies meettechnisch onderscheiden informatie over de spreekvaardigheid bij elke taak geven. Uit het kleinschalig experiment is namelijk gebleken dat verschillende categorieën voor dezelfde dimensie niet of nauwelijks van elkaar onderscheiden werden door een jury.

Gebruikmakend van de inzichten die verkregen zijn in de kleinschalige beoordelingsexperimenten, is een jury van drie beoordelaars geïnstrueerd. Deze beoordeelde vervolgens de tweehonderd prestaties uit de landelijke steekproef op de taken 1, 2 en 3 (studie 1) en de honderd prestaties uit de Amsterdamse steekproef op de taken 1 en 4 (studie 2). Per dimensie werd één categorie beoordeeld, geselecteerd uit de categorieën die in de beoordelingsexperimenten uit hoofdstuk 4 gebruikt zijn. Om indicaties te krijgen over de beoordelaarsstabiliteit is op kleine schaal een herbeoordeling uitgevoerd van prestaties op de taken 1, 2 en 3.

Bij de selectie van de beoordelingscategorieën is rekening gehouden met de resultaten van de beoordelingsexperimenten. De verdeling was als volgt:

	Referentie	Overdracht	Spreekgemak	Verstaanbaarheid
taak 1	Relevantie (onderscheid hoofd-en bijzaken)	Spreektrant (levendigheid)	Spreektempo (weinig hapering, beheerst, vlot)	Articulatie (scherp, helder)
taak 2	Relevantie (vermeldt voornaamste zaken)	Spreektrant (ernstigheid)	Zinsbouw (vlot)	Articulatie (scherp, helder)
taak 3	Relevantie (juiste aanvullende informatie)	Intonatie (gevarieerdheid)	Zinsbouw (vlot)	Articulatie (scherp, helder)
taak 4	Relevantie (onderscheid hoofd- en bijzaken)	Spreektrant (levendigheid)	Spreektempo (weinig hapering, beheerst, vlot)	Articulatie (scherp, helder)

De correlaties tussen de oordelen zijn geanalyseerd met behulp van LISREL, een programma voor de analyse van lineaire structurele relaties. Per taak is een vierfactormodel getoetst. De vier factoren corresponderen met de vier dimensies van het beoordelingsschema.

De resultaten maken duidelijk dat het model redelijk past voor drie van de vier taken; voor taak 3 lijkt het model minder goed te passen. De geschatte correlaties tussen de beoordelingsdimensies per taak zijn soms tamelijk hoog. De correlaties tussen Overdracht en Spreekgemak blijken in het algemeen het hoogst te zijn, die

tussen Referentie en Verstaanbaarheid het laagst. In alle gevallen kan volgens de onderzoeker gesteld worden dat de oordelen meettechnisch onderscheiden informatie over spreekvaardigheid geven.

De predictieve oordeelsvaliditeit

Van Gelderen vindt deze meettechnische verantwoording van zijn beoordelings-schema niet voldoende bevredigend. In een predictief valideringsonderzoek probeert hij antwoord te krijgen op de vraag in hoeverre de gegeven jury-oordelen geprediceerd kunnen worden door meetbare aspecten die aan de oordelen ten grondslag kunnen liggen.

Het antwoord voor elk van de vier dimensies wordt in twee stappen gegeven. In de eerste stap wordt nagegaan in hoeverre het jury-oordeel geprediceerd wordt door de bedoelde aspecten. In de tweede stap wordt bepaald of het jury-oordeel over elke dimensie beter geprediceerd wordt door de bedoelde dan door de niet-bedoelde aspecten.

De predictieve validering voor drie van de vier dimensies - te weten Referentie, Overdracht en Verstaanbaarheid - is goed. De fonetische en linguïstische predictoren die in de definitie van elk van deze dimensies zijn opgenomen blijken *die* dimensie beter te verklaren dan *een van de andere* dimensies.

Voor Referentie en Overdracht kan bovendien vastgesteld worden dat grote percentages variantie van de jury-oordelen verklaard kunnen worden door een relatief klein aantal predictoren. De Referentieoordelen bij drie verschillende spreektaken zijn sterk gerelateerd aan het aantal genoemde belangrijke inhoudselementen; Overdrachtsoordelen bij één (narratieve) spreektaak hebben een sterk verband met aspecten die aan expressiviteit, boeiendheid, betrokkenheid en gepastheid van de spreektrant gekoppeld zijn (intonatievariatie en adequaatheid van register). Het Verstaanbaarheidsoordeel (bij één taak) is meer dan de andere oordelen - gerelateerd aan het aantal moeilijk verstaanbare uitingen.

Wat betreft Spreekgemak zijn de conclusies over de predictieve validering van de oordelen minder positief. Lang niet alle aspecten die in de definitie van Spreekgemaksfuncties zijn opgenomen, zijn in het jury-oordeel meegewogen. In de tweede plaats blijkt een groot gedeelte van de variantie van het Spreekgemaksoordeel verklaard te kunnen worden door predictoren voor Overdracht.

De bruikbaarheid van het beoordelingsschema

In het laatste deel van zijn proefschrift beschrijft Van Gelderen de bruikbaarheid van het beoordelingsschema in twee verschillende contexten: in grootschalig peilingsonderzoek en in diagnostische evaluatie in de onderwijspraktijk.

Van Gelderen ziet twee belangrijke voordelen van het voorgestelde beoordelings-schema voor grootschalig peilingsonderzoek:

- er zijn relatief veel validiteitsgegevens over de oordeelsdimensies bekend zodat nauwkeuriger kan worden aangegeven wat precies bij elke dimensie relevant is voor de beoordeling.
- het instrument kenmerkt zich door een relatieve eenvoud: het is voldoende om per spreekprestatie vier oordelen te geven. Vanuit het oogpunt van efficiëntie een duidelijk voordeel.

Van Gelderen gaat ten slotte in op de diagnostische functie die het beoordelings-schema kan hebben in de onderwijspraktijk. Hij merkt op dat het beoordelings-schema met name bruikbaar is voor de beoordeling van monologische 'beurten' met een zekere interne samenhang en een duidelijk begin en eind, gericht op een definieerbaar communicatief doel (denk bijvoorbeeld aan: het vertellen van verhalen, het geven van een referaat of voordracht voor de klas, het geven van een instructie of uiteenzetting). Het feit dat in het spreekvaardigheidscurriculum voor basis- en voortgezet onderwijs dit soort taken een prominente plaats innemen, beschouwt Van Gelderen als een argument voor de relevantie van zijn beoordelingsschema. Andere argumenten die hij noemt als voordeel voor de onderwijspraktijk zijn:

- bij een afzonderlijke beoordeling van de dimensies blijft de informatie over de onderscheiden aspecten van de vaardigheid behouden zodat gericht gewerkt kan worden aan verbetering van die aspecten;
- een afzonderlijke beoordeling vereist een veel explicietere definitie van de beoordelingscriteria dan een globale beoordeling. Het zal voor de leraar mogelijk worden het oordeel concreet te onderbouwen waardoor de leerlingen meer informatie krijgen over de specifieke moeilijkheden verbonden aan een bepaalde spreesituatie;
- de beoordelingscriteria die afgeleid kunnen worden uit het beoordelingsschema, kunnen van taak tot taak verschillen. Een verdere bewustwording van deze specificiteit zal er volgens Van Gelderen toe leiden dat in het leerplan meer aandacht besteed wordt aan een gemotiveerde selectie van spreektaken voor systematisch spreekonderwijs.
- het is mogelijk de oordeelscategorieën niet alleen door de leraar maar ook door de leerlingen zelf te laten toepassen.

5. Enkele kritische kanttekeningen

Hoewel ik veel waardering heb voor de studie waarvan Van Gelderen verslag doet in zijn proefschrift, wil ik toch enkele kritische kanttekeningen maken.

Domein voor het peilen van spreekvaardigheid

In zijn proefschrift merkt Van Gelderen in het eerste hoofdstuk op dat bij de peiling van taalvaardigheid het domein bepaald wordt door realistische communi-

catieve situaties, waarin verschillende componenten van bedoelde vaardigheden op een geïntegreerde wijze getoetst worden. Het probleem waarvoor Van Gelderen zich gesteld ziet is hoe uit alle mogelijke taalgebruikssituaties een selectie kan worden gemaakt op grond waarvan een gegronde uitspraak kan worden gedaan over het taalvaardigheidsniveau van leerlingen.

In zijn onderzoek gaat Van Gelderen uit van vier taken waarop de spreekvaardigheid en daarmee het beoordelingsschema worden getoetst. In hoofdstuk 2 van het proefschrift wordt beschreven op welke wijze deze vier taken voor het onderzoek zijn geselecteerd. Uit een lijst van 129 soorten taken hebben respondenten (ouders, leerkrachten en specialisten) aangegeven hoe wenselijk en haalbaar zij deze taken als doelstelling van het onderwijs beschouwden. Voor een eerste steekproef in het kader van een voorstudie zijn de volgende zes taken geselecteerd:

- iets vertellen om iemand te amuseren (het navertellen van een gehoord verhaal);
- rapporteren over gebeurtenissen (het snel melden van een ongeval aan de politie);
- uitleggen aan leeftijdgenoten hoe een concreet proces zich voltrekt (hoe een spin zijn web weeft);
- beschrijven van een concrete zaak (nauwkeurig een huis beschrijven om een medeleerling te helpen het te vinden);
- verzoeken om informatie over procedures (vragen om reisinformatie bij de NS);
- uitleg geven aan leeftijdgenoten over een concrete zaak (hoe je een nieuw spel moet spelen).

Vanuit budgettaire overwegingen zag Van Gelderen zich genoodzaakt uit deze zes taken een selectie te maken. Voor zijn onderzoek selecteerde hij daarom de eerste drie taken.

Wanneer peilingsonderzoek er op is gericht om vast te stellen in hoeverre het onderwijs zijn doelen bereikt, dan vraag ik mij af of er sprake is van een optimale keuze van taken. Van Gelderen is er met deze selectie niet in geslaagd om voor de beoordeling van spreekvaardigheid zoveel mogelijk uit te gaan van realistische communicatieve situaties.

'Iets vertellen om iemand te amuseren' bijvoorbeeld is een taak die niet alleen een beroep doet op iemands spreekvaardigheid maar zeker ook op de mate waarin iemand in staat is per definitie amusant te zijn. Een aspect dat hopelijk geen doelstelling van het spreekonderwijs zal zijn.

Het 'verzoeken om informatie over procedures' en 'het beschrijven van een concrete zaak' zouden in het spreekvaardigheidsonderwijs belangrijker onderwijsdoelen moeten zijn dan 'iets vertellen om iemand te amuseren'.

Een ander probleem waarvoor Van Gelderen zich geplaatst zag, was dat hij van mening was dat de genoemde zes taken met elkaar gemeen hadden dat zij aan de spreker inhoudelijk weinig vrijheid boden. De taken waren zo gestructureerd dat de sprekers verplicht waren een aantal concrete inhoudselementen aan de luisteraar mee te delen of te vragen. Omdat in realistische communicatieve situaties de

spreker in de gelegenheid is de inhoud van de tekst zelf te genereren en te structureren, bedacht hij een taak waarin de spreker een dergelijke vrijheid wél kreeg. De kandidaat moet bij deze taak een fictief verhaal dat hij op band te horen krijgt, afmaken.

Het streven om bij het toetsen van spreekvaardigheid een zo groot mogelijke thematische vrijheid voor de spreker te creëren is zeker van groot belang. Bovenstaande taak waarmee Van Gelderen daaraan meent tegemoet te komen, voldoet mijns inziens echter niet. Vanuit het oogpunt van het zoveel mogelijk gebruik maken van realistische communicatieve situaties is de taak namelijk slecht gekozen. De taak is weinig realistisch en bovendien is het afmaken van een fictief verhaal niet erg communicatief. De mate van fantasie en creativiteit zullen hier interfereren.

Bij de opmerking van Van Gelderen dat hij vindt dat er voldoende variatie aan taken is voor de beproeving van een algemeen beoordelingsschema, kunnen dan ook vraagtekens geplaatst worden.

Beschrijving van de dimensies

Van Gelderen vindt in zijn onderzoek meettechnisch onvoldoende onderscheid in de vier dimensies die hij heeft gedefinieerd. Gezien de definities die hij zelf geeft in hoofdstuk 3 kan dat niet echt verwonderlijk genoemd worden.

Bij de beschrijving van de dimensie Referentie merkt Van Gelderen onder andere op: "Op het locale niveau gaat het vooral om het gebruik van (qua betekenis) duidelijke woorden en verwijzingen tussen zinnen. Aspecten als inhoud, structuur, zinsbouw en woordkeus zijn relevant voor de beoordeling van referentie." Bij de definiëring van de dimensie Overdracht zegt Van Gelderen: "Vele gedragsaspecten blijken voor de beoordeling van overdracht relevant: niet alleen stijl, spraaktechnische en non-verbale aspecten maar ook woordkeus, zinsbouw en inhoudelijke aspecten."

Van Gelderen noemt voor de beoordeling van deze twee dimensies dezelfde aspecten: inhoud, woordkeus en zinsbouw. Hoewel Van Gelderen aangeeft dat deze aspecten bij de verschillende dimensies op een ander niveau beoordeeld worden, vraag ik me af of het onderscheid aan beoordelaars gemakkelijk is duidelijk te maken. Aan de instructie voor de beoordelaars zal zeer veel aandacht besteed moeten worden. Zeker wanneer het gaat om grootschalig peilingsonderzoek, zal nader bekeken moeten worden of het beoordelingsschema van Van Gelderen op een goede manier overgedragen kan worden aan een groot aantal beoordelaars.

Lengte van de beoordelingsschalen

Van Gelderen hanteert voor iedere taak en voor iedere beoordelingscategorie dezelfde lengte voor de beoordelingsschaal: een 5-puntsschaal. Gezien de twijfels die er zijn over de mogelijkheid om bij kortdurende prestaties, zoals bij de taken

2 en 3, nauwkeurig verschillende categorieën (simultaan) te beoordelen, lijkt het Van Gelderen verstandig het aantal te beoordelen categorieën tot een minimum te beperken. Het zou voor de hand liggen om de lengte van de beoordelingsschaal aan te passen aan de lengte van het te verwachten spreekprodukt.

Status van de dimensies

In zijn definitie van de verschillende dimensies noemt Van Gelderen de dimensies Referentie en Overdracht *substantieel* en de dimensies Spreekgemak en Verstaanbaarheid *voorwaardelijk*. Het is mijns inziens logisch om aan die dimensies die een substantiële bijdrage leveren aan het oordeel over de spreekvaardigheid een groter gewicht toe te kennen dan aan de voorwaardelijke dimensies. Een mogelijkheid om dat te doen zou zijn om de lengte van de schalen voor de voorwaardelijke dimensies korter te laten zijn dan de lengte van de schalen voor de substantiële dimensies.

Een homogeen jury-oordeel om onderscheid vast te stellen?

In het eerste kleinschalige experiment gaat Van Gelderen na of hij een onderscheid kan vinden in de verschillende beoordelingscategorieën. Hij merkt daarbij op dat het vinden van onderscheid tussen beoordelingscategorieën is gediend met een zo homogeen mogelijk jury-oordeel. Een beoordelaar die afwijkende criteria hanteert, vermindert de kans op het vinden van jury-oordelen die meettechnisch onderscheiden zijn. Daarom heeft Van Gelderen er in zijn eerste experiment voor gekozen om de scores van één beoordelaar niet mee te laten tellen.

Gezien het geringe aantal beoordelaars (vijf voor het beoordelen van prestaties op de taken 1, 2 en 3 en vier voor het beoordelen van prestaties op taak 4), lijkt mij dit een onverantwoorde beslissing. Het is immers goed mogelijk dat de beoordelaar afwijkende criteria hanteert omdat het onderscheid tussen de verschillende beoordelingscategorieën niet duidelijk was voor deze beoordelaar. Door uitschakeling van deze beoordelaar kan daarmee voorbij gegaan worden aan het doel van deze studie.

Spreekgemak, een bijzondere dimensie

De dimensie Spreekgemak blijkt uit de resultaten van het onderzoek een bijzondere positie in te nemen. Niet alleen meettechnisch is deze dimensie moeilijk te onderscheiden van de andere maar ook in het predictieve validiteitsonderzoek blijkt dat bedoelde aspecten voor deze dimensie in het jury-oordeel nauwelijks worden meegewogen. Van Gelderen zou hieruit hebben moeten concluderen dat deze dimensie het niet verdient om in een analytisch beoordelingsschema als aparte dimensie te worden opgenomen. Van Gelderen handhaaft echter zonder verdere argumentatie zijn vierdeling.

De diagnostische functie die een analytisch schema voor de onderwijspraktijk kan hebben, is niet relevant voor deze dimensie. Wanneer blijkt dat een leerling niet over voldoende spreekgemak beschikt, dan zal dat naar alle waarschijnlijkheid meer te maken hebben met de persoon zelf dan met het spreekvaardigheidsonderwijs dat deze leerling heeft gehad. Remediërend onderwijs om spreekgemak te verbeteren zal dan ook weinig effect hebben.

Beoordeling: globaal of analytisch

Van Gelderen noemt de resultaten van zijn studie tamelijk bemoedigend voor gebruik van het geteste beoordelingsschema. Gezien de resultaten van zijn experimenten en die van het predictieve valideringsonderzoek echter zou hij hebben moeten concluderen dat *voor de beoordeling* van spreekvaardigheid een meer globaal beoordelingsschema zeker een nader onderzoek zou verdienen.

Meettechnisch immers geeft hij zelf al aan dat het onderscheid tussen zijn voorgestelde dimensies niet in bevredigende mate onderscheidbaar is. Er moet volgens Van Gelderen zelfs getwijfeld worden of de jury-oordelen over verschillende categorieën überhaupt verschillende informatie verschaffen. In zijn predictieve validiteitsonderzoek vindt Van Gelderen het onderscheid tussen de verschillende dimensies voldoende als de aspecten die hij voor die dimensies definieert de bedoelde dimensie beter voorspellen dan een van de andere dimensies. Voor het handhaven van de verschillende dimensies in het beoordelingsschema vind ik dit een zwak geformuleerde norm.

Ondanks deze kritische kanttekeningen ben ik van mening dat de studie van Van Gelderen een zinvolle kan zijn. Ik denk dan met name aan de diagnostische functie die het beoordelingsschema *voor de onderwijspraktijk* kan hebben: 'spreekvaardigheid in de klas'. Zoals Van Gelderen zelf ook opmerkt in een eerder verschenen artikel (van Gelderen en Oostdam 1992), kan een analytische benadering, zoals van Gelderen die in zijn studie voorstelt, bij het ontwikkelen van leerplannen en leermiddelen voor spreekvaardigheidsonderwijs een goed uitgangspunt zijn.

Bibliografie

- Becker, S.L., The rating of speeches; scale independence, *Speech Monographs* 29 (1962), 38-44.
- Canale M. & M. Swain, Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics* 1 (1980), 1-47.
- Gelderen A. van & R. Oostdam, Het gebruik van beoordelingsschema's voor spreekvaardigheid in de lessen Nederlands. *Levende talen* 469 (1992), 124-132.

Shohamy E., A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition* 10 (1988), 165-179.

Wesdorp, H., *Evaluatietechnieken voor het moedertaalonderwijs*, Den Haag: Staatsuitgeverij, RITP, SVO, 1981.

(manuscript binnengekomen 20 oktober 1994)

(manuscript aanvaard 3 november 1994)