

Spreekvaardigheid analytisch en globaal beoordeeld*

WIL KNIBBELER,

Instituut Toegepaste Taalkunde, Katholieke Universiteit, Nijmegen

Samenvatting

Er is een onderzoek ingesteld om na te gaan of en in hoeverre bij de beoordeling van spreekvaardigheid van het Frans als vreemde taal de resultaten van een analytische beoordeling overeenstemmen met de scores van een globale beoordeling. Onder analytische beoordeling wordt verstaan een kwantitatieve analyse van de transcripten van op geluidsband opgenomen spraakproducten. De globale beoordeling is uitgevoerd eerst door vijf Nederlanders en daarna door drie Fransen. De Franse beoordelaars zijn gekozen op grond van de resultaten van een uitgebreide selectieprocedure. Er werden belangrijke correlaties gevonden zowel tussen de scores van de analytische methode en die van de globale methode door de Nederlanders als tussen de globale scores van de Nederlanders en de globale scores van de Fransen.

1. Inleiding

In *Pedagogische Studiën* 51 (1974), 554-557, beschrijft ondergetekende een methode om spreekvaardigheid in het Frans te testen door middel van een zogenaamde Vrije Spreektoets die afgenomen is in het kader van het Evaluatieonderzoek van een Multimediale Cursus Frans (SVO project 0225).¹ Hierin schetst hij enerzijds de opzet en afname van de toets anderzijds de beoordeling van de resultaten. Met betrekking tot de beoordeling van de resultaten van vrije spreektoetsen wordt algemeen aangenomen dat het bijzonder moeilijk is een methode te ontwikkelen die aan de eisen van validiteit, objectiviteit en betrouwbaarheid voldoet.² De in de praktijk toegepaste methoden kan men onderscheiden in de globaal-impressionistische methoden en de analytische methoden. Bij beide methoden wordt de

beoordeling in kwantitatieve gegevens uitgedrukt. Bij de analytische methode echter wordt ernaar gestreefd het spreekgedrag in een zo groot mogelijk aantal componenten op te splitsen, waarvoor de beoordelaar aparte scores geeft. Voorbeelden hiervan vindt men o.m. bij Upshur (1971: 53-54), Savignon (1972), Beatens Beardsmore (1974: 323), en van Weeren (1975: 142-144).³ Als een ver doorgevoerde vorm van zo'n analytische beoordeling kunnen de formules van Dodson (1963) en de methode van Doyé (1974) genoemd worden. In geen van deze gevallen echter is men er toe overgegaan de resultaten van spreektoetsen te transcriberen alvorens ze te beoordelen.⁴ Het is in dit opzicht dat de analytische methode die door ons in het kader van het project Evaluatieonderzoek van een Multimediale Cursus Frans toegepast is, zich onderscheidt van andere analytische methoden. Onze methode bestaat in het transcriberen van de op geluidsband opgenomen producten van de vrije spreektoets, het coderen van alle fouten en het langs machinale weg kwantificeren van woorden en foutcategorieën. Voor meer bijzonderheden mogen we verwijzen naar de in de aanhef van dit artikel genoemde publicatie in *Pedagogische Studiën*. De kwantificering is hier zo ver doorgevoerd dat de methode als atomistisch-analytisch gekwalificeerd zou kunnen worden. Het lijkt geen twijfel dat deze methode een schat van concrete gegevens oplevert die voor diagnostische doeleinden benut kunnen worden. In het boven geciteerde artikel van *Pedagogische Studiën* hebben wij echter ook beschreven hoe wij op grond van deze analytische methode gekomen zijn tot het meten van spreekvaardigheid. Hierbij hebben we twee aspecten onderscheiden n.l. spreekcorrectheid en spreekgemak. Voor spreekcorrectheid werd als index het percentage goede woorden op het totaal aantal woorden genomen. Voor spreekgemak werd het aantal verschillende goede woorden genomen. Verder vindt men in Knibbeler (1974a: 12) een correlatieberekening van alle testvariabelen van de toetsen die in het kader van dit project afgenomen zijn. Hieruit blijkt dat de scores voor spreekgemak

* Deze studie is tot stand gekomen met medewerking van Ad Horsten, statisticus verbonden aan het Instituut voor Onderzoek van Wetenschappelijk Onderwijs van de Katholieke Universiteit van Nijmegen.

een niet onbelangrijke correlatie vertonen met de scores voor spreekcorrectheid: bij de pretests $r = 0.49$ ($N = 121$) en bij de posttests $r = 0.47$ ($N = 121$). Roy (1967) heeft eveneens een duidelijke samenhang gevonden tussen spreekcorrectheid en spreekgemak (delivery rate); dit laatste aspect van de spreekvaardigheid stelde hij vast op basis van het aantal lettergrepen.

2. Probleemstelling

De vraagstelling die aan het onderhavige onderzoek ten grondslag ligt, is tweevoudig.

2.1. In de atomistisch-analytische methode moge dan wel enige garantie ingebouwd zijn dat de subjectiviteit van de beoordeling tot een minimum beperkt wordt, men kan zich toch afvragen of men op grond van aantallen woorden en aantallen fouten waarheidsgetrouwe uitspraken kan doen over een zo complexe vaardigheid als die van het spreekgedrag. Gorosch (1971: 4-5) zegt in dit verband: 'On pourrait dire: l'évaluation devrait être objective; c'est vrai, mais il faut définir l'objectivité en termes humains applicables à la maîtrise de la langue. Un outil imparfait qui mesure ce que l'on souhaite mesurer vaut mieux qu'un outil parfait, de type mécanique par exemple, qui mesure ce que l'on ne souhaite pas mesurer.' Ook Clark (1972: 41-42) is van oordeel dat: 'A Procrustean solution to the problem of scoring reliability in these cases would be to adopt some arbitrary and easily followed judging standard, such as counting the total number of words spoken or written. Extremely high scoring reliability could be obtained in this way, but at the expense of test validity, since under such a scoring system, the test would become a measure of the student's garrulousness rather than the linguistic quality of his speech or writing.' Waar het dus om gaat is de validiteit van onze op woorden- en foutentellingen gebaseerde conclusies te toetsen. Het meest voor de hand liggende toetsingscriterium is nu het via de globaal-impressionistische methode verkregen oordeel.

Om na te gaan in hoeverre eigenschappen van spreekgedrag afgeleid kunnen worden uit het percentage goede woorden en het aantal verschillende goede woorden, worden de uitspraken van globale beoordelaars als toetssteen genomen.⁵ Wanneer nu zou blijken dat er een overeenkomst bestaat tussen de scores van onze atomistisch-analytische beoordeling en de waarderingscijfers van een globaal-impressionistische beoordeling, zou men aan

deze bevinding een tweevoudige conclusie kunnen verbinden:

- a. analyses van spraakproducten op het niveau van het woord verschaffen ons niet alleen concrete gegevens voor diagnostische doeleinden, maar lenen zich ook om spreekvaardigheid te meten;
- b. als men uitsluitend spreekvaardigheid wil meten, kan men volstaan met de globaal-impressionistische methode. Deze laatste conclusie is des te belangrijker als men bedenkt dat de atomistisch-analytische methode zeer arbeidsintensief is. Het transcriberen van een test van 5 minuten en het toekennen van de foutcodes vergt gemiddeld ± 3 uur, terwijl voor de globaal-impressionistische beoordeling van dezelfde test niet meer dan ruim 5 minuten vereist zijn.

22. Zoals wij in de inleiding vermeld hebben, werd er bij de analytische methode een zekere samenhang gevonden tussen spreekcorrectheid en spreekgemak. Men kan zich nu afvragen of bij een globale beoordeling deze twee aspecten wel als onderscheiden componenten ervaren worden. Als dit niet het geval zou blijken te zijn, zou men met één enkele component kunnen volstaan, althans als het alleen maar gaat om het toetsen van de spreekvaardigheid. Dit zou de taakstelling voor de beoordelaar aanzienlijk vereenvoudigen.

3. Onderzoeksplan

De globaal-impressionistische beoordeling (voortaan globale beoordeling genoemd) is twee maal uitgevoerd:

- a. dezelfde vijf Nederlandse beoordelaars die analytisch beoordeeld hebben, beoordelen een aantal van de banden opnieuw, maar nu globaal;
- b. drie Fransen beoordelen globaal dezelfde banden die de Nederlanders globaal beoordeeld hebben. Ieder van beide beoordelingen heeft haar eigen doelstelling. Door middel van de globale beoordeling van de Nederlanders (GN) wordt getracht de vraag te beantwoorden of onze Nederlandse beoordelaars, als ze op hun spontane indruk waren afgestaan, tot dezelfde resultaten zouden zijn gekomen, m.a.w. of de resultaten van de analytische beoordeling als valide beschouwd mogen worden. Deze conclusie is echter dan pas houdbaar, als we enig zicht hebben op de beoordelingscompetentie van de door ons gekozen Nederlandse beoordelaars. Zoals reeds vermeld in *Pedagogische Studiën* 51, p. 547, zijn onze beoordelaars doctoraal studenten Frans. Het is namelijk niet ondenkbaar dat deze studenten

ofwel de Franse taal nog onvoldoende beheersen, ofwel ten gevolge van hun op perfectie gerichte opleiding de producten te streng beoordelen. Door middel van de globale beoordeling door Fransen (GF) wordt nu getracht de vraag te beantwoorden of de uitspraken van de Nederlandse beoordelaars als objectief beschouwd mogen worden.

Met betrekking tot de eerste toetsing (AN* - GN) dient het volgende opgemerkt te worden. Ten eerste: de vijf Nederlandse beoordelaars waren niet op de hoogte van de uiteindelijke resultaten van onze kwantitatieve analyse die geleid hebben tot de toekenning van een score voor spreekcorrectheid en spreekgemak. Zij wisten dus niet welke scores toegerekend waren aan de testpersonen van wie ze de transcripten gemaakt hadden en de fouten gecodeerd hadden. Ten tweede: het uitschrijven van bandopnames, het wikken en wegen van ieder woord in elke zin, het kiezen van de passende code uit een totaal van 86 mogelijkheden is een fundamenteel andere taak dan het geven van een globale impressie. Ten derde: er lag tussen het maken van de transcripten en de globale beoordeling een tussenruimte van 18 maanden. Men kan dan ook nauwelijks aannemen dat, wanneer van hen gevraagd wordt een spontane impressie weer te geven, deze taak beïnvloed zou worden door het feit dat ze zo lang geleden van deze banden een transcript gemaakt hebben.

Met betrekking tot de tweede toetsing (GN - GF) dient het volgende opgemerkt te worden. Ook de beoordeling door moedertaalsprekers mag niet als een onfeilbare norm beschouwd worden. Uit de literatuur over linguïstische intuïtie blijkt dat moedertaal-informanten niet eensluidend en consequent taaluitingen beoordelen. Om deze moeilijkheid enigszins op te lossen zijn de Franse beoordelaars gekozen op grond van een voorafgaande selectieprocedure. Het proefexperiment dat tot deze selectie geleid heeft, zal in de volgende paragraaf apart beschreven worden.

De voorschriften voor de globale beoordeling van de vrije spreektoets waren voor de Nederlandse en Franse beoordelaars gelijk en luiden aldus:

- Beluister iedere test eenmaal, zet daarbij de band niet stil en draai deze ook niet terug. Het gaat om de eerste spontane indruk.
- Geef onmiddellijk na het beluisteren van een test, dus alvorens met de volgende cursist te beginnen, 2 cijfers, een voor correctheid en een voor spreekgemak. Gebruik daarbij de schaal van 1-20.⁶
- Onder correctheid moet verstaan worden: de

juistheid van uitspraak, grammatica en woordgebruik.

N.B. het gaat om de eerste globale indruk.

- Onder spreekgemak moet verstaan worden: de vaardigheid taal te gebruiken als communicatiemiddel, d.w.z. de duidelijkheid waarmee een testpersoon zich uitdrukt en het gemak waarmee hij dat doet. Dit laatste komt o.a. tot uitdrukking in de gevarieerdheid van het taalgebruik (zinsbouw, woordgebruik).

N.B. het gaat om de eerste globale indruk.

Ten slotte dienen met betrekking tot de onderzoekprocedure nog twee algemene opmerkingen gemaakt te worden. Ten eerste: alle overeenkomsten tussen de verschillende beoordelingen zijn door middel van de product-moment correlatie berekend. Ten tweede: om de door onze Nederlanders uitgevoerde analytische beoordeling te toetsen is het uiteraard mogelijk een meer sluitend analyse-design toe te passen. Zo zou het denkbaar zijn om als tussenstap de analytische beoordeling door Fransen in te voegen. In de praktijk is deze stap echter nauwelijks haalbaar. Zoals uit *Pedagogische Studiën* 51 pp. 547-552 blijkt, vereist deze methode een langdurige instructieperiode en constante begeleiding. Een dergelijke onderneming zou onevenredig veel tijd, energie en geld gekost hebben.

4. Selectie van de Franse beoordelaars

Tien in Frankrijk woonachtige Fransen van verschillende leeftijd en vooropleiding kregen de opdracht tien door Nederlanders ingesproken zinnen globaal op de uitspraak te beoordelen met behulp van een 5-punts-schaal. Er waren in totaal 112 ppn die beoordeeld moesten worden. De bandopnamen van deze 112 ppn waren samengebracht op een verzamelband. Op deze verzamelband kwamen de bandopnamen van 12 ppn tweemaal voor. Deze dubbel exemplaren waren willekeurig over de verzamelband verspreid. Verder zijn de tien zinnen ook nog ingesproken door vijf in Nederland verblijvende Fransen. De duur van dit verblijf varieerde van 2 maanden tot 5 jaar. Deze Fransen zijn gekozen, omdat zij naar het oordeel van de proefleider als model van standaarduitspraak zouden kunnen fungeren. Ook deze Franse opnamen zijn op willekeurige plaatsen ingevoegd tussen de andere prestaties. De beoordelaars waren onkundig van het feit dat in de hele verzameling dubbel exemplaren en opnamen van *native speakers* voorkwamen. De selectie van de beoordelaars heeft plaats-

* AN = analytische beoordeling door Nederlanders.

gevonden op grond van vier criteria: constantheid, nuance-gevoeligheid, validiteit en objectiviteit.

De *constantheid* werd getoetst op grond van de gelijkheid in de beoordeling bij de dubbelexemplaren, uitgaande van de gedachte: hoe kleiner het totale verschil, des te constanter het beoordelings-

gedrag. In Tabel 1 vindt men de scores die de beoordelaars aan de dubbelexemplaren gegeven hebben, links van het streepje bij de eerste beoordeling en rechts bij de tweede beoordeling. In de onderste rij vindt men het totaal van de absolute verschil-scores.

Tabel 1: Scores dubbelexemplaren

	Beoordelaars									
	0	1	2	3	4	5	6	7	8	9
Proef- persoon										
1	4-3	2-3	3-3	3-3	2-3	4-3	3-2	2-3	3-3	3-3
2	3-2	3-2	3-4	2-4	3-3	4-3	2-2	2-3	3-3	3-4
3	3-4	3-3	4-3	3-3	3-3	4-4	4-3	3-3	4-3	4-4
4	3-3	3-3	3-2	2-2	3-3	3-3	4-2	2-2	3-3	4-3
5	3-3	3-5	3-3	3-2	3-4	4-3	2-3	2-3	3-3	4-3
6	2-3	2-3	3-3	3-2	3-1	3-2	3-3	2-2	2-2	4-4
7	3-2	2-3	3-2	2-2	2-2	3-3	3-2	2-2	2-2	3-3
8	4-4	4-4	4-4	4-4	4-4	4-4	3-3	3-4	4-3	3-4
9	4-2	3-4	3-3	3-2	3-2	4-3	3-2	5-2	3-2	3-3
10	3-3	2-2	3-2	3-2	3-2	4-3	3-3	3-3	3-3	3-3
11	4-3	3-4	4-4	4-4	4-4	4-4	4-4	4-3	4-4	4-4
12	2-3	2-4	3-3	3-3	3-3	3-4	3-4	2-4	3-3	3-3
Totaal Verschil	9	10	5	6	6	7	8	10	3	4

De *nuance-gevoeligheid* werd vastgesteld aan de hand van de standaarddeviaties behorende bij de gemiddelden van de per beoordelaar toegekende

scores, hierbij uitgaande van de gedachte: hoe hoger de spreiding, des te gevarieerder heeft de beoordelaar van de beoordelingsschaal gebruik gemaakt.

Tabel 2: Standaarddeviaties per beoordelaar

	Beoordelaars									
	0	1	2	3	4	5	6	7	8	9
Standaard- deviatie	0.88	0.97	0.84	1.03	0.85	0.75	0.91	0.93	0.86	0.54

De *validiteit* werd getoetst aan de hand van de beoordeling van de *native* prestaties. Hierbij is uitgegaan van de gedachte dat in vergelijking met de Nederlandse ppn de Franse ppn de hoogste score verdienen. Het is uiteraard mogelijk dat ook moedertaalsprekers uitspraakfouten maken. Hiertegen kan echter ingebracht worden dat, zoals hierboven

vermeld, de Franse ppn met enige zorg gekozen zijn en dat het bovendien hier gaat om een globale impressie. Een beoordelaar beantwoordt dus aan het validiteitscriterium, als hij aan de vijf Franse ppn de maximum score van 5 toekent. Tabel 3 geeft de som van de scores die iedere beoordelaar aan de vijf *native* prestaties toegekend heeft.

Tabel 3: Som native prestaties

	Beoordelaars									
	0	1	2	3	4	5	6	7	8	9
Scores	24	25	25	23	25	25	25	22	25	21

De *objectiviteit* werd getoetst aan de hand van de overeenstemming tussen de beoordelaars. Tegen de keuze van dit criterium kan uiteraard als bezwaar aangetekend worden dat het beoordelingsgedrag van een groep niet als objectiviteitsnorm gehanteerd mag worden. In dit verband dient echter opgemerkt te worden dat het objectiveren van uitspraakcorrectheid door de taalkundigen als een uiterst delicate zaak gezien wordt en dat men bij gebrek aan beter doorgaans de communis opinio als leidraad neemt. Om nu de overeenstemming tussen de beoordelaars te berekenen zijn eerst de scores van iedere individuele beoordelaar vergeleken met de scores van alle andere beoordelaars. Iedere vergelijking is uitgevoerd in de vorm van een tabel, in totaal 45 tabellen waarvan men een voorbeeld vindt in Tabel 4.

Tabel 4: Vergelijking beoordelaar 0 met beoordelaar 1

	Beoordelaar 1					
	Scores	1	2	3	4	5
Beoordelaar: 0	1	1	3	1	0	0
	2	0	7	8	10	0
	3	2	11	32	20	3
	4	0	1	6	11	5
	5	0	0	0	1	7

De cijfers in de tabel staan voor de aantallen ppn aan wie de beoordelaars een bepaalde score toegekend hebben. In de hoofddiagonaal vindt men het aantal ppn aan wie beoordelaar 0 en beoordelaar 1 dezelfde score toegekend hebben. Om het verschil in beoordeling te wegen zijn de aantallen in de 2 eerste bijdiagonalen later vermenigvuldigd met factor 1, de aantallen in de 2 tweede bijdiagonalen vermenigvuldigd met factor 2 en de aantallen in de 2 derde bijdiagonalen vermenigvuldigd met factor 3. Op grond van deze bewerking konden uit de 45 tabellen twee indices afgeleid worden n.l. het aantal gelijke beoordelingen en het aantal verschilpunten (= gewogen verschil). Een totaal overzicht van deze twee vergelijkingsindices vindt men in Tabel 5. Voor iedere beoordelaar zijn hier twee rijen vermeld: de

bovenste rij geeft de maat van de gelijke beoordeling aan van iedere individuele beoordelaar t.o.v. iedere andere individuele beoordelaar, de onderste rij de maat van de gewogen verschillen van iedere beoordelaar t.o.v. iedere andere individuele beoordelaar.

Tabel 5: Onderlinge vergelijking individuele beoordelaars

	1	2	3	4	5	6	7	8	9	
0	G	58	66	52	81	53	62	59	78	47
	V	88	67	91	51	80	76	75	54	97
1	G		57	45	60	63	55	51	51	56
	V		79	111	79	72	88	93	88	85
2	G			65	68	73	75	63	70	64
	V			68	64	59	61	74	61	76
3	G				62	38	57	50	59	44
	V				76	113	89	100	83	122
4	G					56	64	55	83	45
	V					77	71	80	49	96
5	G						56	54	55	77
	V						82	87	78	59
6	G							53	54	62
	V							87	80	83
7	G								71	48
	V								67	102
8	G									45
	V									97

Om nu vast te stellen in hoeverre iedere individuele beoordelaar in zijn beoordeling overeenkomt en verschilt met de *totale* groep van beoordelaars, zijn

van iedere beoordelaar de aantallen gelijke beoordelingen en de aantallen verschillen gesommeerd. In Tabel 6 vindt men een overzicht van deze totalen.

Tabel 6: Vergelijking individuele beoordelaars met totale groep

Beoordelaar	Gelijk	Vershil (gewogen)
0	556	679
1	496	783
2	601	609
3	472	853
4	574	643
5	525	707
6	538	717
7	504	765
8	566	657
9	488	817

Conclusie:

De hoogste overeenkomst en het laagste verschil in beoordeling komen voor bij de beoordelaars 2, 4 en 8. Deze drie beoordelaars scoren eveneens het hoogst bij de beoordeling van de *native* prestaties (validiteitscriterium). Bovendien blijken deze drie beoordelaars op het vlak van de constantheid en de nuance-gevoeligheid dusdanig te scoren (zie tabellen 1 en 2) dat er geen reden tot verwerping gevonden werd. Dit alles was voor ons aanleiding om de beoordelaars 2, 4 en 8 te kiezen voor de globale beoordeling van de vrije spreektoets.

5. Resultaten vergelijking analytische scores globale cijfers

De vergelijking tussen de verschillende beoordelingen heeft betrekking op drie variabelen n.l.: de werkwijze: analytisch versus globaal; de spreekvaardigheidscomponenten: correctheid en gemak; de beoordelingshoedanigheid: inheems versus uitheems.

Deze drie variabelen geven aanleiding tot een groot aantal correlatieberekeningen. Iedere correlatieberekening heeft tot doel een bepaald aspect van het probleem onder de loupe te nemen. Aangezien een belangrijk probleem bij iedere beoordeling de betrouwbaarheid van de beoordelaar is⁷, zal ook dit aspect bij onze analyses betrokken worden. Langs deze weg wordt de totale probleemstelling in een aantal deelvragen opgesplitst.

5.1. Welke overeenkomst bestaat er tussen de vijf Nederlandse beoordelaars bij de globale beoordeling zowel voor wat betreft spreekcorrectheid als spreekgemak?

Tabel 7: Globale beoordeling alle Nederlandse beoordelaars

N = 54

Correctheid				Gemak			
0.73				0.72			
0.61	0.53			0.60	0.65		
0.65	0.54	0.46		0.75	0.75	0.63	
0.62	0.59	0.43	0.41	0.66	0.78	0.68	0.73

Conclusie:

De correlaties tussen de globale beoordelingen van de vijf Nederlanders variëren tussen 0.41 en 0.78. Dus de criteria die de vijf Nederlandse beoordelaars hanteren, vertonen een belangrijke onderlinge overeenkomst. Deze overeenkomst kan als indicatie van betrouwbaarheid beschouwd worden.

5.2. Welke overeenkomst bestaat er bij iedere Nederlandse beoordelaar tussen het globale beoordelingscijfer dat hij geeft voor spreekcorrectheid en het globale beoordelingscijfer dat hij geeft voor spreekgemak?

Tabel 8: Globale beoordeling iedere Nederlandse beoordelaar correctheid en gemak

r =	N =
0.82	70
0.84	66
0.85	70
0.78	69
0.68	70

Conclusie:

Er bestaat een niet onaanzienlijke overeenkomst tussen de waardering die elke Nederlandse beoordelaar geeft voor spreekcorrectheid en spreekgemak. Op grond van deze bevinding zijn we geneigd aan te nemen dat het weinig zin heeft tussen deze twee parameters een onderscheid te maken, althans op de wijze zoals ze door ons geoperationaliseerd zijn.

5.3. Welke overeenkomst bestaat er bij iedere Nederlandse beoordelaar tussen de analytische scores die uit zijn transcripten afgeleid zijn en zijn globale waarderingscijfers zowel voor wat betreft spreekcorrectheid als spreekgemak?

Tabel 9: Nederlanders analytisch en globaal

Correctheid		Gemak	
r =	N =	r =	N =
0.59	32	0.57	32
0.12	15	0.65	15
0.51	32	0.58	32
0.68	27	0.81	27
0.65	22	0.20	22

Conclusie:

Met uitzondering van beoordelaar 2 bij spreekcorrectheid en beoordelaar 5 bij spreekgemak is de overeenkomst tussen de analytische scores en de globale waarderingscijfers aanzienlijk te noemen. Deze bevindingen verschaffen dus enige indicatie voor het feit dat aan de analytische scores een zekere validiteit toegekend kan worden.

5.4. Welke overeenkomst bestaat er tussen de drie Franse beoordelaars onderling zowel voor wat betreft spreekcorrectheid als spreekgemak?

Tabel 10: Alle Franse beoordelaars

Correctheid		Gemak	
N = 54			
0.67		0.74	
0.71	0.58	0.83	0.77

Conclusie:

De positieve correlatie tussen de globale beoordelingen van de drie Franse beoordelaars is significant. Evenals bij de vijf Nederlandse beoordelaars vertoont hun beoordelingsgedrag een belangrijke onderlinge overeenkomst. Deze onderlinge overeenkomst mag als een indicatie van betrouwbaarheid gezien worden.

5.5. Welke overeenkomst bestaat er bij iedere Franse beoordelaar tussen het cijfer dat hij geeft voor

spreekcorrectheid en het cijfer dat hij geeft voor spreekgemak?

Tabel 11: Iedere Franse beoordelaar correctheid en gemak

r =	N =
0.88	108
0.88	108
0.96	104

Conclusie:

De significante overeenkomst die bij de Nederlandse beoordelaars bij de globale beoordeling gevonden werd tussen de waarderingscijfers voor de beide componenten van het spreekgedrag, blijkt nog groter te zijn bij de Franse beoordelaars. Er is dus de duidelijke indicatie gevonden dat spreekcorrectheid zoals door ons gedefinieerd, nauwelijks onderscheiden wordt van spreekgemak. Bijgevolg lijkt het weinig zinvol om naar deze differentiatie in beoordelingsparameters te streven.

5.6. Welke overeenkomst bestaat er tussen het globale waarderingscijfer van de Nederlandse beoordelaars en dat van de Fransen zowel voor wat betreft spreekcorrectheid als spreekgemak?

Tabel 12: Globale beoordeling Nederlanders en Fransen

Fransman	Correctheid Nederlander				
	r	0.71	0.64	0.45	0.64
n	N = 70	N = 58	N = 70	N = 66	N = 60
r	0.63	0.52	0.53	0.41	0.50
n	N = 54	N = 58	N = 54	N = 57	N = 60
r	0.71	0.67	0.56	0.59	0.37
n	N = 54	N = 57	N = 54	N = 54	N = 58

Fransman	Gemak Nederlander				
	r	0.75	0.79	0.57	0.81
n	N = 70	N = 58	N = 70	N = 66	N = 60
r	0.67	0.77	0.69	0.66	0.64
n	N = 54	N = 58	N = 54	N = 57	N = 60
r	0.70	0.79	0.62	0.74	0.64
n	N = 54	N = 57	N = 54	N = 54	N = 58

Conclusie:

Bij de globale beoordeling bestaat er een niet te onderschatten overeenkomst tussen de cijfers van iedere Franse beoordelaar en die van iedere Nederlandse beoordelaar. Hieruit mag afgeleid worden dat de door ons gekozen Nederlandse beoordelaars, toen zij globaal beoordeelden, valide uitspraken gedaan hebben.

6. *Algemene conclusie*

Wanneer men spreekvaardigheid in een vreemde taal wil beoordelen, heeft het weinig zin onderscheid te maken tussen spreekcorrectheid en spreekgemak, althans volgens de definities zoals die in dit experiment zijn gehanteerd. Zowel bij de atomistisch-analytische beoordeling als bij de globaal-impressionistische beoordeling blijken beide parameters een belangrijke correlatie te vertonen.

Wanneer men uitsluitend de spreekvaardigheid beoordelen wil, kan men volstaan met de globaal-impressionistische methode, aangezien de resultaten van deze methode overeenkomen met die van de arbeidsintensieve analytische methode. Wanneer men echter met het oog op diagnostische doeleinden er naar streeft meer concrete gegevens ter beschikking te krijgen omtrent spraakproducten van vreemde-taalleerders, dan verdient de atomistisch-analytische methode die toegepast is bij het project Evaluatieonderzoek van een Multimediale Cursus Frans, alleszins aanbeveling. De conclusies die omtrent spreekvaardigheid getrokken werden uit deze detail-analyses, blijken bevestigd te worden door de globale beoordelingsmethode.

Noten

1. Over dit project is o.m. gerapporteerd in Knibbeler (1974a, 1974b en 1975).
2. Zie o.m. Perren (1968: 115), Brings (1972: 20) en Bachman (1973: 42).
3. Bij de beoordeling van de schrijfvaardigheid kunnen soortgelijke analytische methoden toegepast worden, zie Wesdorp (1974).
4. Baetens Baerdsmore (1974: 323) heeft een poging in deze richting ondernomen, maar vermeldt dat hij deze methode al spoedig verlaten heeft.
5. Ook Robinson (1969: 9) is van oordeel dat beide beoordelingsmethoden toegepast dienen te worden. Hij ziet echter de taalkundige kwantificering van spraakproducten als de toetsing van het impressionistisch oordeel.

6. Deze schaal is gekozen omdat ze in het Franse schoolstelsel nog steeds gebruikt wordt en bijgevolg voor de Franse beoordelaars het meest vertrouwd is.
7. Valette (1969: 36) zegt hierover: 'The matter of scorer reliability is the subject of continuing investigation, one which must be taken into account for each set of test papers and each team of new scorers.'

Bibliographie

- Bachman, Lyle F., 'Testing Oral Production', *Bulletin of the English Language Center of Bangkok* III/1 (1973), 41-58.
- Baetens Beardsmore, H., 'Testing Oral Fluency', *IRAL* XII/4 (1974), 317-326.
- Brings, Friedhelm, 'Der Leistungstest. Ein Literaturüberblick', *Praxis des neusprachlichen Unterrichts*, 19/1 (1972), 14-25.
- Clark, John L. D., *Foreign Language Testing: Theory and Practice* (Language and the Teacher: A Series in Applied Linguistics nr. 15), Philadelphia 1972.
- Dodson, C. J., *Oral Examinations* (Faculty of Education, University College of Wales, Pamphlet nr. 12), Aberystwyth 1963.
- Doyé, Peter, 'Verfahren zur Feststellung von Schülerleistungen im Englischunterricht und ihre Eignung für die verschiedenen Lernbereiche', *Praxis des neusprachlichen Unterrichts* 21/3 (1974), 251-264.
- Gorosch, M., 'Exposé introductif', in: Comité de l'Education extrascolaire et du Développement culturel, *Symposium organisé à Rüscklikon 3 au 7 mai 1971*, Strasbourg 1971.
- Knibbeler, Wil, *Rendement van een Taalprogramma Frans verwerkt door volwassenen zonder hulp van een leraar*, Paper gelezen op de Eerste Onderwijsresearchdag 29 maart 1974, Nijmegen 1974a.
- Knibbeler, Wil, *Kwantitatieve Analyse van een cursus Frans*, Nijmegen 1974b.
- Knibbeler, Wil, *Persoonsvariabelen bij het leren van een vreemde taal: Onderzoek naar het Leraarloze Taalleren bij volwassenen*, Nijmegen 1975.
- Perren, G. E., 'Testing Spoken Language: Some unsolved Problems', in: Alan Davies (ed.), *Language Testing Symposium*, London 1968, 107-116.
- Robinson, Peter, *Oral Expression Tests*, Based on a paper presented at the Canadian Linguistic Association Congress, Toronto, June 10-13, 1969.
- Roy, Robert R., *Oral French Proficiency, Identification and Evaluation*, Edmonton 1967.

Savignon, Sandra J., 'Teaching for Communicative Competence: A Research Report', *Audio-Visual Language Journal* X/3 (1972), 153-162.

Upshur, John A., 'Objective Evaluation of Oral Proficiency in the ESOL Classroom', *TESOL Quarterly* V/4 (1971), 47-59.

Valette, Rebecca M., *Directions in Foreign Language Testing*, New York 1969.

Weeren, Jan van, 'Spreekvaardigheid en Vreemdetalen-onderwijs (II)' *Levende Talen* 311 (1975), 141-145.

Wesdorp, H., *Het meten van de produktief-schriftelijke taalvaardigheid. Directe en indirecte methoden: 'Opstelbeoordeling' versus 'schrijfvaardigheidstoetsen'*, Purmerend, 1974.

Curriculum vitae

Zie Pedagogische Studiën 51 (1974) 557.