De effecten van taakuitvoering

Verbetert de spreekvaardigheid van T2'ers door het herhaaldelijk voeren van gesprekken met T1'ers?

Universiteit Utrecht
Nederlandse taal en cultuur: educatie en communicatie
Masterscriptie

Eerste begeleider: Jacqueline Evers-Vermeul
Tweede begeleider: Nivja de Jong

Jantine Marchal
Studentnummer: 3014843

10-05-2010

Samenvatting

Binnen dit scriptieonderzoek is onderzocht of het herhaald uitvoeren van gesprekstaken met moedertaalsprekers een positief effect heeft op de spreekvaardigheid van tweedetaalverwervers. Er zijn verschillende studies bekend waarin een effect van taakherhaling beschreven wordt, maar deze resultaten zijn niet eenduidig. Daarom is het belangrijk om het effect van taakherhaling nogmaals te onderzoeken. Daarnaast wordt onderzocht welke rol het onderwerp en eventuele hulpmiddelen hierbij invullen. De volgende onderzoeksvraag staat centraal: *wat is het effect van het herhaaldelijk uitvoeren van taken op de spreekvaardigheid van tweedetaalverwervers en welke rol spelen het onderwerp van de taak en mogelijke hulpmiddelen hierbij?* Deze vraag is onderzocht door spreekprestaties van tweedetaalverwervers te beoordelen tijdens de uitvoering van gesprekstaken met moedertaalsprekers.

De resultaten die naar voren komen in dit scriptieonderzoek zijn niet eenduidig. Het effect van herhaling geldt pas na de uitvoering van twee taken en alleen op complexiteit, vloeiendheid, uitspraak en algemene indruk. De proefpersonen worden niet steeds beter naarmate zij meerdere taken uitvoeren. Dit wordt in de literatuur wel gesuggereerd. De proefpersonen blijken pas profijt te hebben van hun ervaring met het uitvoeren van taken wanneer die ervaring minstens twee taken beslaat. Bij de uitvoering van een tweede taak speelt de ervaring van de uitvoering van de eerste taak nog geen rol. De ervaring speelt pas een rol vanaf taak drie. Het uitvoeren van gesprekstaken heeft dus alleen nut als er minstens drie taken uitgevoerd worden. Daarnaast zijn er geen hoofdeffecten van onderwerp gevonden, alleen twee interactie-effecten op het eindmoment. Het is dus niet het geval dat de scores op spreekvaardigheid bij bepaalde onderwerpen hoger zijn dan bij andere onderwerpen, maar dat dit samenhangt met het taaknummer. Elk taaknummer vraagt om een ander soort onderwerp om een zo hoog mogelijke score op spreekvaardigheid te halen. Dit is ook te zien bij het gebruik van hulpmiddelen, al zijn daar ook hoofdeffecten van hulpmiddelen te vinden. Op sommige aspecten van spreekvaardigheid wordt het hoogst gescoord wanneer er geen hulpmiddel gebruikt wordt en het laagst wanneer er afbeeldingen gebruikt worden. Het is dus belangrijk om een bewuste selectie te maken van te gebruiken hulpmiddelen.

1. Inleiding

Gesprekstaken worden in de onderwijspraktijk veelvuldig gebruikt voor de beoefening van spreekvaardigheid. Deze taken maken het mogelijk om in duo's (of met meerdere personen) in de doeltaal te communiceren. In het Moderne Vreemde Talenonderwijs is het gebruik van gesprekstaken dan ook niet meer weg te denken. Maar hoe moeten deze gesprekstaken er uitzien (qua onderwerp en hulpmiddelen) om zo effectief mogelijk te zijn? En hoe vaak moeten gesprekstaken herhaald worden om een positief effect te hebben op de spreekvaardigheid?

Deze vragen staan centraal in deze scriptie. Hierbij wordt aangesloten bij het onderzoeksproject NIFLAR. NIFLAR staat voor Networked Interaction in Foreign Language Acquisition and Research. Dit is een tweejarig project, gestart op 1 januari 2009, dat gecoördineerd wordt door de Universiteit Utrecht. Aan dit project werken acht verschillende universiteiten binnen en buiten Europa mee. Het doel van het project is, zoals het op de website van NIFLAR (www.niflar.eu) vermeld staat: 'make foreign language education more authentic and interactive through innovative e-learning environments'. Binnen het project wordt gewerkt met digitale omgevingen, zoals Second Life en Video Web Conferencing.

Deze scriptie sluit aan bij het laatste punt: Video Web Conferencing. In deze omgeving kunnen studenten uit verschillende landen inloggen met een webcam en headset en met elkaar communiceren. NIFLAR gebruikt deze omgeving onder andere om gesprekken te voeren tussen docenten-in-opleiding uit Nederland en studenten Nederlands in Tsjechië. In deze scriptie wordt de uitvoering van deze gesprekstaken onder de loep genomen. Er wordt met name gekeken of een herhaling van taken effect heeft op de spreekvaardigheid. Daarnaast wordt onderzocht welke rol het onderwerp en mogelijke hulpmiddelen hierbij spelen.

In deze scriptie zal de volgende onderzoeksvraag centraal staan: *wat is het effect van het herhaaldelijk uitvoeren van taken op de spreekvaardigheid van tweedetaalverwervers en welke rol spelen het onderwerp van de taak en mogelijke hulpmiddelen hierbij?* Het onderzoek heeft dus als doel te onderzoeken op welke manier het gebruik van gesprekstaken een positief effect heeft op de spreekvaardigheid. Deze resultaten kunnen toegepast worden op de ontwikkeling van gesprekstaken in de onderwijspraktijk. Er zal op verschillende momenten dan ook verwezen worden naar de onderwijspraktijk.

Allereerst zal een theoretisch kader gegeven worden (hoofdstuk 2). Vervolgens wordt de opzet van het onderzoek besproken in hoofdstuk 3. De resultaten uit het onderzoek worden besproken in hoofdstuk 4, gevolgd door een conclusie (hoofdstuk 5) en een discussie (hoofdstuk 6).

2. Theoretisch kader

Om de onderzoeksvraag te kunnen beantwoorden, is het van belang dat er een definitie gegeven wordt van spreekvaardigheid. Want wat wordt er precies verstaan onder spreekvaardigheid? In paragraaf 2.1 wordt een antwoord gegeven op deze vraag. Daarnaast worden, in paragraaf 2.2, mogelijke methodes en opdrachten (taken) beschreven om spreekvaardigheid in de onderwijspraktijk te oefenen en er wordt voorgaand onderzoek beschreven waarin het uitvoeren van

taken centraal staat. Daarbij wordt beschreven wat volgens voorgaand onderzoek het effect is van het onderwerp van een taak en het belang van het herhalen van taken op de spreekvaardigheid van taalverwervers.

2.1 Spreekvaardigheid

Naast schrijfvaardigheid, luistervaardigheid en leesvaardigheid is spreekvaardigheid een belangrijk onderdeel van de verwerving van een (tweede) taal. Een goede beheersing van spreekvaardigheid is immers nodig om jezelf verstaanbaar te maken. Maar wat is nu precies spreekvaardigheid? Fulcher (2003: 23) drukt het op de volgende manier uit: "speaking is the verbal use of language to communicate with others". Spreken is dus het mondelinge gebruik van taal om te communiceren met andere personen. Hoe beter iemand zich uit kan drukken, des te beter de gesprekspartner begrijpt wat er bedoeld wordt. Het is dan ook belangrijk om spreekvaardigheid te trainen, zodat de spreekvaardigheid verbeterd kan worden.

Er worden verschillende soorten oefeningen gebruikt om spreekvaardigheid te trainen. In het onderwijs wordt vaak gebruik gemaakt van het uitvoeren van opdrachten in monoloogvorm. De meeste leerlingen/studenten hebben weleens individuele spreekopdrachten (spreektaken) uitgevoerd. Denk hierbij aan spreekbeurten of het voorlezen van een (stuk) tekst. Maar er zijn ook andere manieren om spreekvaardigheid te trainen. Hennephof (1971, in: Wesdorp, 1978) beargumenteert dat een gesprek (dialoog) als spreekvaardigheidoefening meer geschikt is dan een voordracht. Een gesprek is namelijk veel natuurlijker dan een voordracht en is dus een betere reflectie van een spreekprestatie in de werkelijkheid. Het zal eerder voorkomen dat mensen in de supermarkt in gesprek raken dan dat iemand een speech (monoloog) houdt. Ook kan een voordracht voor een publiek (docent en klasgenoten) zorgen voor spanningen bij een leerling. Iedereen let immers op het taalgebruik en de inhoud van de uitingen van de leerlingen. De leerling zal zich bij het voeren van een gesprek met een klasgenoot meer op zijn gemak voelen en gemakkelijker en vloeiender spreken dan wanneer hij een voordracht houdt en gespannen is (Van Gelderen, 1992). Daarnaast is een gespreksvorm geschikt voor het oefenen van spreekvaardigheid, omdat meerdere leerlingen dan tegelijkertijd bezig kunnen zijn met het verbeteren van hun spreekvaardigheid. Het is dus een efficiënte vorm van oefenen. Ook Rijlaarsdam (1982) legt de nadruk op het uitvoeren van spreekvaardigheidoefeningen in groepsverband. Volgens hem kun je spreekvaardigheid het best oefenen door het voeren van een discussie of debat (polyloog). Door het voeren van discussies kun je tegelijkertijd je spreekvaardigheid en je luistervaardigheid in de vreemde taal trainen. Je traint dus, net als bij het voeren van een gesprek, twee taalvaardigheden op hetzelfde moment (Wijgh, 1996). Dat draagt weer bij aan de efficiëntie die (wegens tijdgebrek) nodig is in de onderwijspraktijk.

In dit onderzoek wordt de spreekprestatie van studenten geanalyseerd wanneer zij een gesprekstaak uitvoeren. In de volgende paragraaf wordt een definitie gegeven van een gesprekstaak. Daarnaast komen verschillende typen gesprekstaken aan bod en wordt er een overzicht gegeven van studies naar het effect van het uitvoeren van een taak op de spreekvaardigheid.

2.2 Gesprekstaken

2.2.1 Definitie van een gesprekstaak

Over de definitie van een gesprekstaak wordt gediscussieerd in de literatuur (Ellis, 2003). In dit onderzoek is gekozen voor de definitie van Lee (2000, in: Ellis, 2003: 4), omdat deze definitie aansluit bij de definitie die binnen het NIFLAR-project gehanteerd wordt voor het begrip 'taak' (De Graaff, 2010). Een taak is volgens Lee:

(1) a classroom activity or exercise that has:
  (a) an objective obtainable only by the interaction among participants,
  (b) a mechanism for structuring and sequencing interaction, and
  (c) a focus on meaning exchange;
(2) a language learning endeavor that requires learners to comprehend, manipulate, and/or produce the target language as they perform some set of workplans.

Een taak geeft een richtlijn voor het gesprek en bevat vaak een communicatief doel dat aan het eind van het gesprek bereikt moet zijn. De taak geldt dus eigenlijk als leidraad voor het gesprek, zodat het gesprek gestuurd kan worden. Het biedt taalleerders een kans om met elkaar te communiceren in de doeltaal om tot een vastgesteld einddoel te komen.

Bij het uitvoeren van een gesprekstaak krijgen de deelnemers een beschrijving met de informatie die nodig is tijdens het gesprek en de verwachtingen die aan het gesprek gesteld worden. Dit kan betrekking hebben op het thema van het gesprek, maar het kunnen ook vragen zijn die gesteld moeten worden aan de gesprekspartner. Naast deze inhoudelijke informatie wordt ook vermeld hoeveel tijd er beschikbaar is voor het uitvoeren van de taak en hoe de taak beoordeeld zal worden. De sprekers weten dus wat van hen verwacht wordt bij het uitvoeren van de taak en dat is belangrijk voor een zo optimaal mogelijke uitvoering van de gesprekstaak (Luoma, 2004).

2.2.2 Typen gesprekstaken

Er zijn verschillende typen gesprekstaken. Een paar voorbeelden van mogelijke gesprekstaken (Luoma, 2004):

- neem deel aan een interview
- speel een rollenspel
- discussieer over een krantenartikel.

De meest gebruikte gesprekstaken in de onderwijspraktijk zijn 'een rollenspel spelen met behulp van een woordenlijst (denk bijvoorbeeld aan een gesprek in de supermarkt in het buitenland) en 'bespreek standpunten met betrekking tot een bepaald onderwerp' (Koet, 2007).

Er moet bewust nagedacht worden over de vorm van de gesprekstaak die ontwikkeld of geselecteerd wordt voor een spreekvaardigheidoefening. Niet alle typen gesprekstaken zijn geschikt wanneer specifieke vaardigheden getest of geoefend worden. Het belang van deze beslissing wordt

benadrukt door Van Gelderen (1992), Fulcher (2003) en Luoma (2004). Wanneer bijvoorbeeld getest wordt hoe goed iemand zich staande kan houden in een discussie, dan is het niet verstandig om een taak op te stellen waarbij van de spreker gevraagd wordt om een stuk tekst hardop voor te lezen (bijvoorbeeld als onderdeel van een rollenspel), maar dan kan er beter gekozen worden voor een polyloog.

Naast de vorm van een gesprekstaak (dialoog of polyloog) kun je ook de structuur van de taak bepalen. De taak geeft immers een bepaald kader aan de spreker waarbinnen hij moet reageren op zijn gesprekspartner. De taakontwerper kan bepalen hoeveel structuur hij aan het kader geeft, dus in welke mate het gesprek vastligt en hoeveel vrijheid de sprekers hebben. Wanneer een docent een taak ontwerpt waarbij sprekers een stuk tekst voor moeten lezen (bijvoorbeeld bij een uitgeschreven rollenspel), dan is de output van het gesprek van tevoren al bepaald. In dit geval spreken we van een *gesloten taak*. Wanneer er binnen de taak bepaalde vragen of stellingen besproken moeten worden, dan is er wel een duidelijke sturing in het gesprek, maar dan kunnen de sprekers zelf bepalen hoe ver ze op een onderwerp doorgaan en waar het gesprek dus naartoe gaat. Als dit het geval is, dan is er sprake van een *gestuurde taak*. Wanneer er slechts een thema/onderwerp verstrekt wordt voor het gesprek en het compleet overgelaten wordt aan de sprekers wat zij ermee doen, dan is er sprake van een *open taak*.

Naast de structuur van een taak is ook het feit of een gesprek gepland of ongepland is van belang (Luoma, 2004). Bij *planned speech* hebben de sprekers namelijk de mogelijkheid om zich voor te bereiden en standaardzinnen te bedenken. De sprekers kunnen voordat het gesprek aanvangt hun mening formuleren over het te bespreken onderwerp en de spreekvaardigheidsoefening daarmee goed voorbereiden. Dit is vooral van toepassing op monologen. Er kan ook gekozen worden voor *unplanned speech.* De sprekers weten hierbij niet welke kant het gesprek opgaat. De sprekers weten vooraf alleen het onderwerp van het gesprek. Ze vormen de tekst van het gesprek tijdens het spreken door op elkaar te reageren. *Unplanned speech* is moeilijker en kost meer cognitief vermogen dan *planned speech* omdat bij *unplanned speech* de tekst van het gesprek niet vastligt en er tijdens de uitvoering van de taak nagedacht moet worden over zowel vorm als inhoud. Het heeft als voordeel dat het gesprek natuurlijker verloopt en meer een reflectie van de werkelijkheid biedt dan bij *planned speech*. Dit komt vooral doordat het gesprek niet vooraf geconstrueerd is en het dus geen opsomming van zinnen is, maar een echt gesprek waarbij de sprekers moeten improviseren en op elkaar moeten reageren.

Naast het kiezen van de juiste gespreksvorm en structuur is ook de moeilijkheidsgraad van een taak belangrijk. Wanneer de taak (te) moeilijk gemaakt wordt, dan zal het spreken waarschijnlijk moeizamer gaan dan wanneer de taak eenvoudig uit te voeren is. Een taak moet dus niet te ingewikkeld zijn, want dat komt de spontaniteit van het gesprek niet ten goede. De sprekers zullen lang na moeten denken over een adequate reactie, waarbij zij zowel aan de inhoud als aan de vorm veel tijd moeten besteden. Hierdoor zal het gesprek niet vloeiend verlopen. Aan de andere kant moet een taak ook niet te eenvoudig zijn, want dan is de uitdaging voor de sprekers simpelweg verdwenen. De moeilijkheidsgraad van een gesprekstaak wordt door verschillende factoren beïnvloed. Allereerst speelt het onderwerp van het gesprek een grote rol. Als dit thema dicht bij de spreker staat en heel concreet is, dan is het gemakkelijker om erover te spreken dan wanneer het thema ver van de spreker af staat (Skehan, 1998; Council of Europe, 2001). Ook speelt de kennis van de spreker over het gespreksonderwerp een rol. Hoe meer iemand van een bepaald onderwerp afweet, des te gemakkelijker hij erover zal kunnen spreken.

Bij het ontwikkelen van gesprekstaken zijn er dus veel keuzes die gemaakt moeten worden. Deze keuzes zullen immers van invloed zijn op de spreekprestatie. In dit onderzoek is ervoor gekozen om te onderzoeken of het herhaaldelijk uitvoeren van gesprekstaken überhaupt zorgt voor een verbetering van de spreekvaardigheid en of het onderwerp van het gesprek en eventuele hulpmiddelen hieraan bij kunnen dragen. Het onderwerp en eventuele hulpmiddelen bepalen immers deels de moeilijkheidsgraad van een gesprekstaak en dit kan een effect hebben op de spreekvaardigheid. In dit onderzoek wordt dan ook gekeken of het onderwerp van de taak een effect heeft op spreekvaardigheid en of de proefpersonen steeds beter gaan scoren op spreekvaardigheid naarmate zij meer gesprekstaken uitvoeren (effect van herhaling). Deze onderzoeksthema's zijn al in verschillende studies naar voren gekomen. De resultaten van deze studies worden in de volgende subparagraaf besproken.

2.2.3 Het effect van taakuitvoering op de spreekvaardigheid

Ellis (2003) beschrijft in zijn werk twee onderzoeken waarbij het effect van onderwerp getest werd op het begrip en taalvaardigheidsniveau van de sprekers. Het eerste onderzoek is uitgevoerd door Gass en Varonis (1984). Bij het onderzoek waren 142 proefpersonen betrokken die moesten luisteren naar twee sets met op band ingesproken zinnen. Deze zinnen zijn ingesproken door vier non-native speakers. De ene set zinnen ging over een verhaal dat de 142 proefpersonen net gehoord hebben en de andere set zinnen ging over dingen die in de wereld gebeuren. De proefpersonen moesten opschrijven wat ze begrepen uit de zinnen. Uit het onderzoek is gebleken dat de zinnen over het verhaal beter begrepen worden dan de zinnen over wereldse dingen. Gass en Varonis hebben daaruit de volgende conclusie getrokken: bekendheid met het onderwerp heeft effect op begrip. Deze conclusie wordt onderbouwd door een recent en grootschalig onderzoek, namelijk een studie van Lange (2000). Hij heeft zijn proefpersonen taken laten uitvoeren over twee verschillende onderwerpen. Taak één hield het voeren van een discussie over gevangenen in en taak twee het voeren van een discussie over harttransplantaties. Uit deze studie is gebleken dat proefpersonen meer gemotiveerd zijn om over een onderwerp te spreken als het onderwerp dichter bij hen ligt. Bij taak één werd er namelijk hoger gescoord op spreekvaardigheid dan bij taak twee. Harttransplantaties worden als abstracter beschouwd dan gevangenen.

Deze studies laten zien dat het onderwerp van een taak effect kan hebben op de spreekvaardigheid van degenen die de taak uitvoeren. Omdat er maar twee onderzoeken bekend zijn naar het effect van het onderwerp op spreekvaardigheid is het belangrijk om na te gaan of in dit scriptieonderzoek vergelijkbare resultaten gevonden kunnen worden. Dit kan namelijk een verklaring geven voor een verschil in scores op spreekvaardigheid bij de verschillende taken.

Naast deze aandacht voor een mogelijk effect van het onderwerp van een taak ligt er ook een focus op het effect van het herhalen van taken op de spreekvaardigheid van taakuitvoerders. Het kan in de onderwijspraktijk natuurlijk voorkomen dat een spreker meerdere keren in aanraking komt met dezelfde taak. Dit kan exact dezelfde taak zijn, maar een taak van hetzelfde type ('vertel iets na') geldt ook als herhaling. Docenten kunnen ervoor kiezen om hun leerlingen telkens eenzelfde type taak (het voeren van een gesprek met een klasgenoot over onderwerp *x*) uit te laten voeren, omdat slechts het onderwerp aangepast hoeft te worden om een nieuwe taak te laten ontstaan. Dit is

tijdbesparend voor de docent (hij hoeft geen compleet nieuwe taak te ontwerpen) en voor de student (de vorm van de taak is bekend). De sprekers kunnen echter ook gewend raken aan het uitvoeren van een bepaald type taak. Hierdoor kunnen bijvoorbeeld verveling en demotivatie optreden. Er zijn verschillende onderzoeken gepubliceerd waarin het effect van herhaling van taken onderzocht wordt. Deze onderzoeksresultaten zijn niet zo eenduidig als de resultaten bij de studies naar het effect van het onderwerp. Hierdoor is het noodzakelijk om vervolgonderzoek naar het effect van herhaling van taken te doen.

Yule, Powers en McDonald (1992) laten in hun studie een effect van herhaling zien. Bij het onderzoek waren enerzijds 20 internationale onderwijsassistenten betrokken en anderzijds 50 taalleerders. De taalleerders werden aan onderwijsassistenten gekoppeld en beide partijen kregen een plattegrond in handen. De plattegronden weken iets van elkaar af. Het was de bedoeling dat de onderwijsassistent de taalleerder een bepaalde route op zijn kaart moest laten afleggen en dan 'onderweg' moest kijken tegen welke problemen de taalleerders aan zouden lopen. Zodra de route op de kaart ingetekend was, werd er gewisseld van partner en werd er een soortgelijke opdracht uitgevoerd (in andere duo's). De opdrachten waren dus niet exact gelijk. Wanneer de onderwijsassistenten een dergelijke opdracht voor de tweede keer uitvoerden, waren ze beter in staat de taalleerder te helpen en verliep het gesprek vloeiender. De taalleerder vroeg op zijn beurt (bij de uitvoering van de tweede taak) meer om opheldering en bevestiging. De conclusie van Yule, Powers en McDonald was dan ook: herhaling van een bepaalde taak zorgt voor een verhoging van de communicatieve efficiëntie.

Ook Bygate (1996) heeft laten zien dat het herhaaldelijk uitvoeren van een bepaalde taak positief uitpakt. Aan de proefpersoon werd gevraagd om 'een fragment van Tom en Jerry na te vertellen'. Na drie dagen werd deze taak weer uitgevoerd door de proefpersoon (zonder dat hij vooraf wist dat hij de taak twee keer zou gaan doen). Hetzelfde Tom en Jerry-fragment werd bekeken en de proefpersoon moest weer vertellen wat hij gezien had. De proefpersoon vertoonde bij de tweede uitvoering van de taak een hoge mate van zelfcorrectie en maakte gebruik van een meer gevarieerde woordenschat en complexere zinsbouw. Hij scoorde hoger op 'accuracy', 'repertoire' en 'fluency'. Dit komt doordat de proefpersoon bij de eerste uitvoering van de taak vooral bezig was met betekenisoverdracht, maar bij de tweede keer ook aandacht aan de vorm van zijn taaluitingen besteedde. Deze resultaten zijn niet alleen te verklaren door ervaring met het uitvoeren van een bepaalde taak. Het is logisch dat de proefpersoon meer en vloeiender kan vertellen over een tekenfilmfragment wanneer hij het fragment voor de tweede keer ziet. De resultaten kunnen behaald zijn doordat de taak letterlijk herhaald is.

In een tweede studie toont Bygate (2001) echter soortgelijke resultaten aan: de herhaling van een taak heeft effect op de spreekvaardigheid. De taken waarop dit onderzoek gebaseerd is, zijn 'vertel een Tom en Jerry fragment na' (narratieve taak) en 'houd een interview' (interviewtaak). Bij het onderzoek werden 48 proefpersonen betrokken. Zij waren verdeeld over drie condities. De proefpersonen in conditie één kregen in tien weken tijd zes verschillende narratieve taken (vertel een Tom en Jerry fragment na) en slechts twee interviewtaken. De proefpersonen in conditie twee kregen in diezelfde tien weken tijd zes verschillende interviewtaken en slechts twee narratieve taken. De controlegroep deed aan het begin en het einde van de tien weken slechts twee interviewtaken en twee narratieve taken. De begin- en eindtaken waren voor elke conditie gelijk. Één taak werd dus letterlijk herhaald. Uit het onderzoek bleek dat groep één aan het eind veel beter scoorde op vloeiendheid en complexiteit voor de narratieve taak die herhaald werd. Hetzelfde gold voor groep twee bij de herhaalde interviewtaak. Maar dit resultaat werd niet gevonden voor de andere taken,

die niet herhaald werden. Hier zien we dus dat de resultaten alleen gelden wanneer proefpersonen twee keer exact dezelfde taak uitvoeren. Wanneer dat het geval is, dan wordt er een grotere complexiteit en vloeiendheid gevonden in de spreekvaardigheid van de proefpersonen bij de tweede uitvoering ten opzichte van de eerste uitvoering van de taak. Wanneer een nieuwe taak van hetzelfde type uitgevoerd wordt, is er geen effect te vinden.

Dit effect vinden we wel bij Gass, Mackey, Fernandez en Alvarez-Torres (1999) in hun studie. Aan dit onderzoek hebben 103 studenten Spaans deelgenomen. Zij werden verdeeld over drie condities, waaronder één controlegroep. In elke conditie kregen de proefpersonen geluidloze fragmenten uit Mr. Bean te zien en werd hen gevraagd om direct na het kijken te vertellen wat ze gezien hebben. De proefpersonen in conditie één kregen één fragment drie keer te zien (met telkens twee dagen ertussen) en één fragment één keer. De proefpersonen in conditie twee kregen elke keer een ander fragment te zien (met ook telkens twee dagen ertussen). En de proefpersonen in conditie drie (controlegroep) kregen alleen aan het begin van de week en aan het eind van de week een filmpje te zien. Zij zagen dus maar twee fragmenten. De navertellingen van de proefpersonen werden beoordeeld op lexicon, morfosyntaxis en algemene spreekvaardigheid. De resultaten zijn als volgt: de proefpersonen in de eerste conditie scoorden bij de eerste drie navertellingen (van hetzelfde fragment) steeds beter op alle drie de vlakken. Maar wanneer zij een nieuw fragment te zien kregen, dan was alleen de score van de algemene spreekvaardigheid hoger dan bij de voorgaande fragmenten. Dit was ook te zien bij de tweede conditie. Er is dus een effect van herhaling (de algemene score voor spreekvaardigheid gaat omhoog). Dit effect geldt niet voor morfosyntaxis en lexicon.

Tot slot laten ook Lynch en McLean (2000) zien dat proefpersonen bij herhaling van een taak hoger scoren op accuraatheid en vloeiendheid ten opzichte van de eerste uitvoering van de taak. Aan dit onderzoek namen veertien proefpersonen deel. Lynch en Mclean hebben het effect van taakherhaling onderzocht met behulp van twee verschillende taken, namelijk ‘het maken van een postercollage naar aanleiding van een artikel’ en ‘het beantwoorden van vragen naar aanleiding van de posters’. Taak één werd uitgevoerd in paren. Daarna zijn de paren opgesplitst. De ene helft van het koppel bleef bij de poster staan en beantwoordde vragen over de inhoud van de poster, terwijl de andere helft van het stel rondliep en vragen stelde over posters die andere duo’s gemaakt hadden. Per poster had de vragensteller vijf minuten de tijd om de poster te bestuderen en er vragen over te stellen. Daarna ging hij door naar de volgende poster. Wanneer hij weer terugkwam bij zijn startpunt (bij zijn partner) werden de rollen omgekeerd en ging degene die net vragen beantwoord had rondlopen om vragen te stellen en de vragensteller ging vragen beantwoorden. Voor beide rollen gold dat hij/zij zes verschillende gesprekspartners kreeg. Men ging immers na vijf minuten door naar de volgende poster en er waren in totaal zeven posters. Lynch en Mclean laten in hun onderzoek zien dat de proefpersonen steeds accurater en vloeiender gaan spreken naarmate ze meer taken uitvoeren. Dit valt te verklaren, omdat de proefpersonen bij de eerste poster waarschijnlijk nog moeite hadden met formuleren en vragen bedenken, terwijl dit bij een aantal posters later al veel geautomatiseerder ging. Het herhaald uitvoeren van een bepaald type taak draagt dus bij aan een verbetering van de spreekvaardigheid.

Plough en Gass (1993) rapporteren daarentegen geen effect van herhaling van taken op de spreekvaardigheid. In het onderzoek werden 36 proefpersonen in paren ingedeeld en verdeeld over twee condities (groepen). Aan alle paren werd gevraagd om twee typen taken uit te voeren, namelijk: ‘zoek het verschil’ en ‘voer een discussie’. Groep één was bekend met dergelijke taken en groep twee niet. De proefpersonen in groep één hadden namelijk al twee keer geoefend in het

klaslokaal met dergelijke type taken. Er werd verwacht dat groep één beter zou presteren, aangezien zij bekend was met dergelijke taken. De spreekvaardigheid werd gemeten door scores te geven voor de onderdelen 'overlaps', 'echoes', 'confirmation checks', 'clarification requests', 'interruptions' en 'sentence completion'. De totaalscore van deze onderdelen bepaalde het uitvoeringscijfer. Er bleek geen verschil te zijn in prestatie op de taken tussen de twee groepen. Het was wel zo dat de proefpersonen in groep één meer om bevestiging (hogere score op 'confirmation checks') en verheldering (hogere score op 'clarification requests') vroegen dan de proefpersonen in groep twee, maar er was geen verschil in uitvoering (dus op de totaalscore). Dit onderzoek toont aan dat proefpersonen niet automatisch een taak beter uitvoeren naarmate ze meer ervaring hebben met het uitvoeren van een dergelijke taak.

Er bestaan dus vooral positieve resultaten met betrekking tot het effect van herhaling van taken, maar er zijn ook onderzoeken die geen effect van herhaling laten zien. In sommige onderzoeken zien we dat een (exacte) herhaling van (een bepaald type) taken zorgt voor een verbetering van de spreekvaardigheid, maar in andere onderzoeken (zoals bij Plough en Gass, 1993) wordt dit effect niet gevonden. Een verklaring hiervoor is dat in het onderzoek van Plough en Gass verveling een rol gespeeld kan hebben. De groep die al bekend is met het type taak, toont wellicht minder enthousiasme en inzet dan de groep die nog niet bekend is met het type taak. Hierdoor maken zij geen gebruik van hun voorsprong en kan het gebeuren dat de tweede groep gelijke resultaten boekt. Doordat de resultaten uit de verschillende studies elkaar deels tegen spreken is het belangrijk om binnen dit scriptieonderzoek te kijken of er een effect is van herhaling. De resultaten zijn niet eenduidig genoeg om te kunnen zeggen dat herhaling per definitie zorgt voor verbetering van spreekvaardigheid. Er is meer onderzoek nodig om dergelijke conclusies te kunnen trekken.

Tot slot wordt in deze scriptie verslag gedaan van onderzoek naar een mogelijk effect van het gebruik van hulpmiddelen op de spreekvaardigheid bij het uitvoeren van gesprekstaken. Denk hierbij aan het gebruik van een woordenweb (mindmap), foto's, liedjes, geschreven teksten, etc. Deze hulpmiddelen kunnen de loop van het gesprek beïnvloeden en wellicht ook het niveau van spreekvaardigheid van de sprekers. Het zou goed mogelijk zijn dat het gesprek anders verloopt door het gebruik van hulpmiddelen, maar dit is, voor zover is kunnen nagaan, niet eerder onderzocht. Het is mogelijk dat hulpmiddelen sprekers op weg kunnen helpen. Dan kunnen ze wellicht sneller associaties maken met andere woorden en gebeurtenissen, waardoor ze meer te vertellen hebben. Maar het is ook mogelijk dat hulpmiddelen juist een negatief effect hebben op de spreekvaardigheid. Ze kunnen spontaan taalgebruik in de weg staan. Denk hierbij bijvoorbeeld aan een woordenweb met kernwoorden. De spreker kan zich verplicht voelen om de kernwoorden te gebruiken, waardoor hij geremd wordt in zijn taaluitingen.

Hulpmiddelen worden wel gebruikt in verschillende onderzoeken (denk bijvoorbeeld aan de posters in het onderzoek van Lynch en McLean, 2000), maar hun mogelijke effect op de spreekvaardigheid wordt niet onderzocht. Binnen dit scriptieonderzoek wordt gekeken of er een effect is van het gebruik van hulpmiddelen en wat dat effect inhoudt. Als er een effect blijkt te zijn, dan zal dit gevolgen hebben voor de ontwikkeling van taken in de toekomst. Naast een effect van onderwerp en herhaling wordt er dus ook onderzocht of er een effect van hulpmiddelen bestaat.

In het volgende hoofdstuk zal besproken worden hoe dit onderzoek opgezet is om tot een antwoord te komen op de hoofdvraag (reeds genoemd in de inleiding): *wat is het effect van het herhaaldelijk*

uitvoeren van taken op de spreekvaardigheid van tweedetaalverwervers en welke rol spelen het onderwerp van de taak en mogelijke hulpmiddelen hierbij?

3. Operationalisatie

In dit hoofdstuk wordt beschreven op welke manier de hoofdvraag onderzocht is. Allereerst worden de proefpersonen besproken die hebben meegewerkt aan het onderzoek. Daarna wordt besproken wat voor taken er gebruikt zijn in het onderzoek. Vervolgens wordt de beoordeling van de uitgevoerde taken toegelicht. Hierbij is er vooral aandacht voor de criteria waarmee het spreekvaardigheidsniveau gemeten kan worden en het beoordelingsmodel waarmee de uitvoering van de gesprekstaken vervolgens beoordeeld is. Tot slot wordt nog stilgestaan bij de problemen die voor kunnen komen bij de beoordeling van spreekvaardigheid en wat eraan is gedaan om deze problemen zoveel mogelijk te ondervangen.

3.1 Proefpersonen

De proefpersonen zijn voormalige deelnemers aan het NIFLAR-project. Zij hebben aan dit project meegewerkt om hun spreekvaardigheid te verbeteren. Het betreft 40 eerste- en tweedejaarsstudenten, die aan de Universiteit van Olomouc in Tsjechië Nederlands studeren. Deze studenten leren Nederlands als tweede taal. In deze scriptie zullen zij aangeduid worden als T2'ers (tweedetaalverwervers). De Tsjechische studenten hebben weinig ervaring met spreken in het Nederlands. Zij spreken af en toe in het Nederlands tegen hun docent, maar niet tegen hun medeleerlingen. Zij zijn door NIFLAR in contact gebracht met Nederlandse studenten, zodat zij de mogelijkheid krijgen om hun spreekvaardigheid te verbeteren. Of de spreekvaardigheid van de T2'ers daadwerkelijk verbeterd is door het uitvoeren van gesprekstaken wordt onderzocht in deze scriptie. De Nederlandse studenten, afkomstig van de Universiteit Utrecht (29 in totaal), volgen een educatieve master om een eerstegraads bevoegdheid voor het docentschap te verkrijgen. Het zijn, op twee sprekers na, allemaal moedertaalsprekers van het Nederlands. Deze studenten zullen voortaan T1'ers genoemd worden. De twee sprekers die een andere taal dan het Nederlands als moedertaal hebben, zijn near-native speakers van het Nederlands en daarom wel meegenomen in dit scriptieonderzoek.

Bij de uitvoering van de taken worden twee vliegen in één klap geslagen, want de Tsjechische studenten kunnen oefenen in het spreken van Nederlands en de Nederlandse studenten kunnen oefenen in het lesgeven aan NT2'ers (mensen die Nederlands als tweede taal verwerven). Beide groepen studenten die aan het onderzoek deelnemen, hebben dus belang bij het succesvol uitvoeren van de gesprekstaken. Zij zullen dan ook gemotiveerd zijn om actief deel te nemen aan de taken, omdat het in hun eigen belang is.

3.2 Taken

De taken zijn uitgevoerd door middel van contact tussen T1'ers en T2'ers via een webcam en een headset. De studenten zijn hierbij ingedeeld in paren, waarbij elk paar bestaat uit een T1'er en een T2'er. Deze paren hebben elk drie taken uitgevoerd met steeds een week tijd ertussen. De taken houden in dat elk paar een gesprek voert over een door de T1'er uitgewerkt thema. We hebben hier dus te maken met gestuurde taken. De T1'er bepaalt het thema en het verloop van het gesprek. De T2'er geeft slechts inhoudelijk invulling, maar geeft geen sturing aan het gesprek. Er is hierbij sprake van grotendeels unplanned speech, omdat de precieze inhoud van het gesprek tijdens de taakuitvoering gevormd wordt. De sprekers vormen het gesprek door op elkaar te reageren. De T1'er geeft wel sturing (door bepaalde onderwerpen aan te snijden), maar hij weet niet hoe de T2'er op vragen zal reageren en zal dus ook moeten improviseren. Voor de T2'er is het geheel unplanned.

De studenten hebben drie keer met dezelfde gesprekspartner gesproken over drie verschillende onderwerpen. De hoofdonderwerpen zijn:

- Relaties (familie, liefde, homoseksualiteit)
- Feestdagen (verjaardagen, kerst, cultuurspecifieke feestdagen)
- Studentenleven (studie, studentenhuisvesting, vrijetijdsbesteding)
- Cultuuruitwisseling (cultuurspecifieke gerechten, Nederland, Tsjechië, verschillen tussen Nederland en Tsjechië)

De onderwerpen van de taken lopen erg uiteen, dus is het verstandig om te kijken of een onderwerp van een taak effect heeft op de spreekvaardigheid. Als dit namelijk het geval is, dan is de keuze van het onderwerp (mede) bepalend voor de prestaties die sprekers leveren. Dit zou kunnen betekenen dat verschillende sprekers over hetzelfde onderwerp zouden moeten spreken om hun kansen op goede spreekprestaties zo gelijk mogelijk te houden. Het is dus goed om een mogelijk effect van onderwerp te onderzoeken. Een voorbeeld van een ontwikkelde taak over studentenleven (wonen) is te vinden in bijlage 1.

Naast het onderwerp verschilt ook het soort en de hoeveelheid hulpmiddelen die een T2'er tot zijn beschikking heeft. Bij sommige taken beschikt de T2'er over een woordenlijst (woordenweb) die de T1'er in elkaar heeft gezet ter ondersteuning van het uitvoeren van de taak. In andere gevallen vindt het gesprek plaats aan de hand van een afbeelding (foto) of artikel. Ook is het mogelijk dat er combinaties van hulpmiddelen (denk bijvoorbeeld aan een videofilm en foto's) of helemaal geen hulpmiddelen gebruikt zijn. Hieronder staat een overzicht met de gebruikte hulpmiddelen:

- Woordenweb
- Afbeeldingen/foto's
- Agenda
- Artikel
- Videofilm (bijvoorbeeld van YouTube)
- Onderwerpenlijst
- Liedje met gatentekst (invultekst)
- *(geen hulpmiddel)*

De namen van de hulpmiddelen spreken waarschijnlijk wel voor zich, maar een kleine toelichting bij het hulpmiddel 'onderwerpenlijst' kan mijns inziens geen kwaad. Hiermee wordt een lijst bedoeld met onderwerpen waar de sprekers het tijdens het gesprek over gaan hebben. Deze lijst wordt verstrekt door de T1'er aan de T2'er aan het begin van het gesprek. De T2'er weet dus direct wat er aan de orde zal komen binnen het gesprek.

Bij de analyse van de spreekvaardigheid wordt dus gekeken naar mogelijke effecten van onderwerp, herhaling (ervaring in het doen van taken) en het gebruik van hulpmiddelen op de spreekvaardigheid van T2'ers.

3.3 Analyse van uitgevoerde taken

In totaal zijn 116 video's beschikbaar van de webcam-opnames die gemaakt zijn bij het uitvoeren van de taken. Van deze video's zijn 105 video's geschikt bevonden voor analyse. Dit lagere aantal komt door het ontbreken van één van de taken van enkele sprekers, waardoor er geen drie taken beschikbaar zijn van die spreker. Ook de slechte kwaliteit van enkele opnames of de onmogelijkheid om een opname te openen en bekijken hebben gezorgd voor het verwijderen van enkele video's uit dit onderzoek. Van de overgebleven video's zijn fragmenten bekeken. Er is gekozen voor fragmenten, omdat de video's elk minstens 25 minuten lang zijn. Het was niet haalbaar om alle video's in hun geheel te bekijken. Daardoor zijn er per video drie fragmenten geselecteerd die door verschillende beoordelaars bekeken zijn. Deze fragmenten duren elk ongeveer drie minuten en zijn afkomstig uit het begin, midden en einde van de video's, zodat er een beeld geschetst kan worden van het gemiddelde spreekvaardigheidsniveau. Aan het begin van het gesprek kunnen zenuwen en opstartproblemen immers een rol spelen en aan het eind van het gesprek zijn de sprekers wellicht moe en minder geconcentreerd. Ook is het mogelijk dat de sprekers gedurende het gesprek steeds beter gaan spreken. In dat geval zullen in het eindfragment de hoogste scores op spreekvaardigheid behaald worden. Doordat er dus een verschil verwacht wordt in spreekprestaties gedurende één gesprek (één video) worden verschillende momenten uit het gesprek bekeken. Deze fragmenten zijn beoordeeld op basis van criteria. Deze worden beschreven in de volgende subparagraaf.

3.3.1 Beoordelingscriteria

Om het effect van taakuitvoering op de spreekvaardigheid te kunnen onderzoeken, is het belangrijk om criteria op te stellen waarmee het niveau van de spreekvaardigheid gemeten kan worden. Op die manier kan het spreekvaardigheidsniveau bij de uitvoering van taak één vergeleken worden met het spreekvaardigheidsniveau bij de uitvoering van taak twee en drie.

Binnen het NIFLAR-project wordt het rapport van de Council of Europe (2001) als uitgangspunt genomen. Bij het selecteren van beoordelingscriteria voor dit scriptieonderzoek is dit rapport eveneens als uitgangspunt gekozen, omdat dit scriptieonderzoek aansluit bij het NIFLAR-project (zie inleiding). Het rapport van de Council of Europe (2001) draagt de naam '*Common European Framework of Reference for Languages: Learning, Teaching, Assessment*' (hierna te noemen: CEFR) en is een resultaat van een Europees project. Binnen dit project zijn dertig

verschillende beoordelingsmodellen verzameld uit een aantal landen. Deze modellen zijn vervolgens met elkaar vergeleken en de beste en meest gebruikte criteria zijn gekozen om te dienen als basis voor de beoordeling van spreekvaardigheid in het Europese taalonderwijs. Het uiteindelijke doel van het project is het mogelijk maken van een uitwisseling van ideeën over talenonderwijs en de verbetering van het talenonderwijs (De Jong, 2003). Dit project startte al in 1971 en geeft een basis voor onderwijsprogramma's en methodes op het gebied van taalvaardigheid. Hiermee wordt het vreemde talenonderwijs in Europa enigszins op elkaar afgestemd en hoeven docenten niet meer zelf aan de slag met beoordelingsmodellen voor taalvaardigheidsoefeningen. De modellen die het CEFR biedt, pleiten dus voor een uniforme beoordeling van spreekvaardigheid en meer overeenstemming over de beoordeling van spreekvaardigheid in Europa.

Binnen dit scriptieonderzoek zijn de volgende aspecten van spreekvaardigheid als criteria gekozen: 'complexiteit', 'nauwkeurigheid', 'vloeiendheid', 'uitspraak' en 'algemene indruk'. Deze criteria hiervan worden ook genoemd in het rapport van het CEFR. Het CEFR hanteert daarnaast ook nog een aantal andere criteria die niet zijn overgenomen in dit scriptieonderzoek. Per criterium zal worden besproken waarom het wel of juist niet gebruikt is binnen dit onderzoek.

Een criterium als *complexiteit* is van belang, omdat deze categorie laat zien in hoeverre een persoon in staat is om variatie aan te brengen in zinsbouw en woordkeus. Hoe flexibeler iemand is, hoe beter zijn spreekvaardigheid is. Ook *nauwkeurigheid* is van belang, omdat deze categorie laat zien hoeveel fouten een spreker maakt. Als iemand veel fouten maakt, dan laat dat zien dat diegene het taalverwervingsproces nog niet volledig doorlopen heeft. Naast complexiteit en nauwkeurigheid is ook *vloeiendheid* van belang, omdat deze categorie iets zegt over het spreektempo en het aantal aarzelingen van een spreker. Deze categorie wordt vaak als belangrijk beschouwd, omdat je een goede indruk krijgt van de spreekvaardigheid van een taalverwerver door de moeite die de spreker heeft om zich te uiten. Daarnaast wordt ook *uitspraak* van belang geacht. Bij deze categorie is ook goed te merken of een taalverwerver zich de klankpatronen, klemtonen en intonaties van een taal eigen gemaakt heeft. Tot slot wordt het criterium *algemene indruk* gehanteerd. Dit criterium komt ook voor bij het CEFR, al wordt het in dat rapport *adequaatheid* genoemd. Bij *algemene indruk* kan men een beoordeling geven voor de spreekprestatie in zijn geheel. Dit is een belangrijk criterium, want het is een natuurlijke manier van beoordelen. De beoordelaar geeft in één score weer wat hij van de spreekprestatie vond. Dit doet hij op gevoel. Je hoeft geen ervaren beoordelaar te zijn om te bepalen wat voor indruk iemand op jou maakt (Rijlaarsdam en Wesdorp, 1984). Maar omdat het moeilijk is om deze score te interpreteren (want wat houdt een algemene indruk in? Waar let je op?), is het slechts een extra criterium waarbij de beoordelaar een oordeel kan geven over de algehele prestatie van een persoon.

Het rapport van het CEFR geeft in totaal vijf criteria voor de beoordeling van kwalitatieve aspecten van gesproken taal, te weten: reikwijdte, nauwkeurigheid, vloeiendheid, interactie en coherentie. Hiervan zijn de criteria *reikwijdte, nauwkeurigheid* en *vloeiendheid* overgenomen. Met *reikwijdte* wordt de mate bedoeld waarin men zich kan uitdrukken, dus de woordenschat en zinsconstructies die iemand beheerst. Dit komt overeen met de term *complexiteit* in dit scriptieonderzoek. De andere twee criteria worden kort toegelicht met een uitleg waarom ze niet zijn meegenomen binnen dit onderzoek. Het begrip *interactie* laat zien hoe goed iemand in staat is om deel te nemen aan een gesprek. Dit criterium is binnen dit onderzoek niet meegenomen, omdat het bij dit onderzoek niet gaat om de bijdrage die de T2'er levert om het gesprek gaande te houden, maar om de uitingen die hij geeft als reactie op vragen van de T1'er. De T1'er neemt in de gesprekken het initiatief, dus het heeft geen zin om de interactievaardigheden van de T2'er te

meten. De T1'er en de T2'er zijn gezamenlijk verantwoordelijk voor de interactie. De interactie is geen individuele prestatie van de T2'er. Het zou dan ook vreemd zijn om dit wel mee te laten wegen bij de meting van het spreekvaardigheidsniveau. Het begrip *coherentie* laat zien hoe goed iemand zinnen aan elkaar kan verbinden tot een geheel (bijvoorbeeld door het gebruik van verbindingswoorden). Dit criterium is niet meegenomen, omdat bij de vooranalyse van de uitvoering van de gesprekstaken opviel dat de T2'ers veelal korte antwoorden geven, waarbij dus laag gescoord zou worden op 'coherentie' wegens het ontbreken van elementen als verbindingswoorden. Dit zou geen goed beeld geven van de T2'ers, want zij laten in de gesprekken wel degelijk zien dat ze begrijpen waar het gesprek over gaat en dat ze in staat zijn om adequate reacties te geven. Aangezien de sprekers beginnende taalverwervers zijn, wordt er weinig belang gehecht aan *coherentie*.

Binnen het onderzoeksproject NIFLAR worden, in navolging van het CEFR, de volgende criteria gehanteerd om spreekvaardigheid te beoordelen: *vloeiendheid*, *complexiteit, accuraatheid* en *adequaatheid.* De criteria *vloeiendheid, complexiteit* en *adequaatheid* zijn, zoals gezegd, meegenomen in het onderzoek. *Adequaatheid* wordt het in dit onderzoek meegerekend bij *algemene indruk*. Bij *adequaatheid* wordt gekeken of de gestelde doelen behaald worden en of het gesprek goed verloopt. In dit scriptieonderzoek is niet specifiek naar het behalen van de doelen gekeken, maar wel naar het verloop van het gesprek en de indruk die het gesprek geeft. Dit is gemeten in het criterium *algemene indruk*. Het criterium *accuraatheid* is binnen dit scriptieonderzoek opgesplitst in aparte categorieën (aspecten van spreekvaardigheid). Dit criterium is namelijk, binnen NIFLAR, een alomvattend begrip voor elementen als uitspraak, zinsbouw en vocabulaire. Binnen dit scriptieonderzoek is ervoor gekozen om de verschillende elementen van dit criterium los van elkaar te bekijken (in de categorieën *complexiteit* en *uitspraak*).

De criteria zijn verwerkt tot een beoordelingsmodel waarmee een aantal beoordelaars de video's hebben beoordeeld. Hierbij zijn per criterium verschillende vragen gesteld waarmee een score op het betreffende aspect van spreekvaardigheid gemeten kan worden. Dit beoordelingsmodel en de beoordelaars die met het model gewerkt hebben, worden in de volgende subparagrafen besproken.

3.3.2 Beoordelingsmodel

Een voorwaarde voor een realistisch beoordelingsmodel voor de onderwijspraktijk is dat het model een efficiënte en tijdsbesparende manier is om de spreekvaardigheid van een spreker te meten. Het moet voor verschillende sprekers en spreeksituaties te gebruiken zijn. Ook moet het model een hoge betrouwbaarheid geven (Van Gelderen, 1992). Het CEFR lijkt een dergelijk model te bieden, maar dit model kan helaas niet volstaan voor de onderwijspraktijk. Per criterium worden verschillende factoren genoemd die van invloed zijn op het criterium, waardoor het moeilijk is om een genuanceerd beeld te krijgen. Onder *nauwkeurigheid* op B2-niveau (een term uit het CEFR voor een 'gevorderde taalgebruiker') in het genoemde CEFR-model ('beoordeling van kwalitatieve aspecten van gesproken taal') staan bijvoorbeeld de volgende voorwaarden waar een spreker aan moet voldoen wil hij op *nauwkeurigheid* ingeschaald worden op B2-niveau: 'Toont een betrekkelijk grote beheersing van de grammatica. Maakt geen fouten die tot onbegrip leiden en kan de meeste van zijn of haar vergissingen corrigeren' (Council of Europe, 2000: 29). De spreker moet laten zien dat hij én een grote beheersing van de grammatica heeft én geen fouten maakt die tot onbegrip leiden én de

meeste van zijn vergissingen kan corrigeren. Hier wordt een spreker dus tegelijkertijd op drie verschillende aspecten beoordeeld. Als hij niet voldoet aan één van de voorwaarden (hij heeft bijvoorbeeld geen grote beheersing van de grammatica), dan kan hij niet ingeschaald worden op B2-niveau, terwijl hij er op basis van de andere twee voorwaarden wellicht wel hoort. Ook kan er getwist worden over de betekenis van 'groot' (in 'grote beheersing'). In dit onderzoek is dan ook niet gewerkt met dit model. Hierbij wordt dus ook afgeweken van het NIFLAR-project die hier wel gebruik van maakt. Binnen dit scriptieonderzoek is ervoor gekozen om de criteria die in de vorige paragraaf besproken zijn, te verwerken in korte stellingen en hier een eigen beoordelingsmodel mee te creëren.

Het model dat uiteindelijk ontwikkeld is, is een scoreformulier met verschillende criteria die bevraagd worden. Per criterium zijn verschillende vragen gesteld (bijvoorbeeld over spreektempo, aantal aarzelingen en aantal pauzes bij *vloeiendheid*) om een genuanceerd beeld te krijgen van verschillende aspecten van spreekvaardigheid. Het beoordelingsmodel is zo gemaakt, dat de beoordelaars snel en gemakkelijk hun oordeel kunnen geven. Dit is belangrijk, aangezien de beoordelaars geen tot weinig ervaring hebben met het beoordelen van spreekvaardigheid. Per vraag zijn er zeven antwoordmogelijkheden. Deze antwoorden zijn elk voorzien van een label. Dat wil zeggen: elk antwoord heeft een waarde toegekend gekregen, zodat er geen misverstand kan ontstaan over de inhoud van de antwoordmogelijkheden. Het beoordelingsmodel is te vinden in bijlage 2.

Bij dit beoordelingsmodel is een instructieformulier gemaakt. Verschillende auteurs, onder wie Rijlaarsdam (1982), Wijgh (1996) en Luoma (2004), stellen dat het handig is om een dergelijke instructie te maken wanneer meerdere beoordelaars met hetzelfde model moeten werken. Deze instructie moet in ieder geval de volgende elementen bevatten:

- definitie van de criteria (voorzien van een voorbeeld)
- uitleg van het scoremodel

Door middel van de instructie kan verwarring en misinterpretatie bij de beoordelaars voorkomen worden en is er een grotere kans op een hoge interbeoordelaarsbetrouwbaarheid. Deze instructie is ook te vinden in bijlage 2.

Doordat het model zeven antwoordmogelijkheden heeft, is het een model met een 7-puntsschaal. Voor deze vorm is gekozen, omdat in de praktijk blijkt dat de uitersten van een schaal niet vaak gebruikt worden en er dan dus uiteindelijk een 5-puntsschaal overblijft. Door voor een 7-puntsschaal te kiezen, blijft er op deze manier genoeg nuance. Dit is wederom een punt waarop dit onderzoek verschilt van het onderzoek binnen het NIFLAR-project. Dit project maakt immers gebruik van 5-puntsschalen.

3.3.3 Beoordeling

Om een betrouwbaar oordeel te krijgen over de spreekvaardigheid van sprekers is het belangrijk om meerdere oordelen met elkaar te vergelijken. Om dit te kunnen bewerkstelligen zijn deelnemers aan een tweedejaarscursus 'Methoden in communicatiekundig onderzoek' (MCO) van de Universiteit Utrecht gevraagd om een aantal fragmenten (begin, midden of eindfragmenten) uit de video's te bekijken en een oordeel te geven over het spreekvaardigheidsniveau van de T2'ers. Om de motivatie

en het enthousiasme van deze studenten te bevorderen zijn zij gevraagd mee te denken over beoordelingscriteria die iets zeggen over het spreekvaardigheidsniveau van een T2'er. Daarnaast zijn zij actief betrokken bij de totstandkoming van het definitieve beoordelingsmodel. Hiervoor is gekozen om mogelijke problemen bij de beoordeling van spreekvaardigheid (zogenaamde 'effecten') (Wesdorp, 1978. Rijlaarsdam, 1982. Van Gelderen, 1992. Koet, 2007) tegen te gaan. Dit zijn belangrijke punten waarop gelet moet worden bij de beoordeling van spreekprestaties. Er zijn zes verschillende effecten:

- signifisch effect
- halo-effect
- sequentie-effect
- normverschuiving
- persoonlijke vergelijking
- contaminatie-effect

Bij het uitvoeren van dit onderzoek is ernaar gestreefd om deze problemen zoveel mogelijk te ondervangen. Het eerste probleem, het *signifisch effect,* houdt in dat beoordelaars een verschillende interpretatie kunnen hebben van het beoordelingsmodel en/of de criteria waarop zij moeten beoordelen. Dit effect zorgt voor een lage interbeoordelaarsbetrouwbaarheid. Dit effect is tegengegaan door de beoordelaars actief te betrekken bij het maken van een beoordelingsschema. Er is een voorlopige versie aan de beoordelaars voorgelegd waarop zij konden reageren wat onduidelijk of juist duidelijk was en wat de verbeterpunten waren, zodat het beoordelingsmodel geoptimaliseerd kon worden. In het oude beoordelingsmodel stond bijvoorbeeld de vraag:

De gebruikte woordenschat van de T2'er is

Heel beperkt	*Redelijk Beperkt*	*Enigszins Beperkt*	*Neutraal*	*Enigszins uitgebreid*	*Redelijk uitgebreid*	*Heel uitgebreid*
○	○	○	○	○	○	○

Deze vraag is na overleg met de MCO-studenten veranderd in de stelling:

2. De L2'er varieert in woordkeus

Helemaal niet	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

Door het overleg met de MCO-studenten is er overeenstemming bereikt over de te hanteren criteria (welke wel en niet belangrijk zijn bij de beoordeling van spreekvaardigheid). Daarnaast is er tijdens een college geoefend met het gebruiken van het beoordelingsmodel door gezamenlijk een voorbeeldfragment te bekijken en te beoordelen, waarbij nog enkele onduidelijkheden met betrekking tot het beoordelingsmodel naar boven zijn gekomen. Er is niet alleen gelet op de vraagstelling, maar ook op de interpretatie van de antwoordmogelijkheden. Elke antwoordmogelijkheid heeft een label gekregen, zodat je zeker weten dat elke beoordelaar hetzelfde verstaat onder een bepaalde waarde. Hiermee wordt voorkomen dat beoordelaars een verschillende interpretatie van een criterium of antwoord hebben, waardoor de interbeoordelaarsbetrouwbaarheid naar alle waarschijnlijkheid hoger zal worden.

Het tweede effect, het *halo-effect,* betekent dat de score op een dominant criterium uitstraalt (vandaar de term 'halo') naar een ander criterium. De score op criterium a heeft dus invloed op de beoordeling van criterium b. Dit effect is tegengegaan door verschillende vormen te gebruiken voor de antwoordschalen op het beoordelingsmodel. Bij sommige criteria loopt de schaal van negatief naar positief, maar bij andere criteria loopt hij juist van positief naar negatief. Ook is er gekozen voor een variërende vraagstelling en antwoordmogelijkheid. De vraagstelling bestaat uit vragen en stellingen, zodat de lezer goed moet lezen wat er bedoeld wordt. En de antwoordmogelijkheid varieert van 'eens tot oneens' bij sommige vragen en 'altijd tot nooit' bij andere vragen.

Het *sequentie-effect* kan van toepassing zijn bij beoordeling van meerdere spreekvaardigheidstaken achter elkaar. Dit effect betekent namelijk dat het niveau van de voorgaande prestaties van andere sprekers invloed heeft op de beoordeling van de huidige spreekprestatie. Als er een aantal slechte sprekers geweest zijn, dan kun je de neiging hebben om een redelijke spreker een acht als rapportcijfer te geven, omdat hij eruit springt in vergelijking tot de voorgaande sprekers. Maar als dezelfde redelijke spreker zijn opdracht uitvoert na een aantal uitzonderlijk goede sprekers, dan is de beoordelaar sneller geneigd om de diezelfde prestatie een zes te geven. Dit effect is moeilijker te ondervangen, omdat hierbij de persoonlijke manier van beoordelen van belang is. Toch is er wel een poging toe gedaan door elke beoordelaar de fragmenten in een andere volgorde te laten bekijken. Zo hebben de verschillende beoordelaars niet dezelfde videofragmenten als eerst/laatst.

De invloed van een persoonlijke beoordeling geldt ook voor *normverschuiving*. Dit effect houdt in dat beoordelaars zich aanpassen aan het gemiddelde niveau van de sprekers. Als een klas over het algemeen heel goed presteert op spreekvaardigheid, dan kan het toch gebeuren dat alle sprekers gemiddeld een zeven krijgen. Als de klas minder goed presteert op spreekvaardigheid, maar iedereen zit ongeveer op hetzelfde niveau, dan zal de beoordelaar niet alleen vijven uitdelen, maar ook achten geven aan diegenen die in vergelijking tot de rest van de klas eruit springen. Dit effect zal vrijwel geen rol spelen in dit onderzoek, aangezien de beoordelaars dezelfde groep sprekers beoordelen. Het gemiddelde groepsniveau is voor elke beoordelaar gelijk.

Ook het effect *persoonlijke vergelijking* wordt door beoordelaars zelf veroorzaakt. De ene beoordelaar is nu eenmaal soepeler en minder kritisch dan de andere beoordelaar en zal dus ook standaard hogere beoordelingen geven dan de strengere beoordelaar. Dit effect is dan ook bijna niet tegen te gaan, omdat de beoordelaars zich vaak niet bewust zijn van hun (niet-)kritische houding en er geen standaardnorm is waarmee je de spreekprestatie kunt vergelijken. Om dit effect tegen te gaan, zijn er labels geplaatst bij alle antwoordmogelijkheden op het beoordelingsmodel, zodat er meer consensus ontstaat over de waarden van de schaal. Toch zal de *persoonlijke vergelijking* hoogstwaarschijnlijk een rol blijven spelen.

Een laatste effect dat wordt genoemd in de literatuur is het *contaminatie-effect.* Hierbij speelt de beoordelaar wederom een belangrijke rol. De beoordelaar kan namelijk belang hebben bij een goede beoordeling. Dit komt voor als een student optreedt als beoordelaar en zijn medestudent moet beoordelen en daarbij zijn medestudent niet af wil vallen. Maar niet alleen studentbeoordelaars handelen uit eigen belang. Ook een docent kan wat milder zijn bij de beoordeling van specifieke leerlingen, omdat hij de leerling in kwestie bijvoorbeeld graag mag of omdat de leerling er (in de ogen van de docent onterecht) slecht voor staat. Ook kan de reputatie van een school op het spel staan. Wanneer er veel onvoldoendes gehaald worden op een school, dan wek je de indruk dat er slecht onderwijs gegeven wordt, omdat iedereen onder de maat presteert.

De docent kan ervoor zorgen dat zijn school goed overkomt door zijn leerlingen goede cijfers te geven. Aangezien de beoordelaars (MCO-studenten) vrijwel niets te maken hebben met de studenten die deel hebben genomen aan de gesprekstaken is het zeer onwaarschijnlijk dat dit effect een rol gespeeld heeft.

Samenvattend lijkt het erop dat alleen de *persoonlijke vergelijking* en het *sequentie-effect* in dit onderzoek een rol kunnen spelen. Het *sequentie-effect* is deels ondervangen door de randomisering en opsplitsing van de videofragmenten, maar toch is het niet ondenkbaar dat de uitvoering van voorgaande sprekers invloed heeft op de beoordeling van sprekers die daarop volgen. Ook de *persoonlijke vergelijking* zal hoogstwaarschijnlijk een rol blijven spelen, aangezien dit moeilijk tegen te gaan is. De andere effecten zijn zoveel mogelijk ondervangen.

Een ander probleem dat specifiek een rol speelt bij dit scriptieonderzoek is het gebrek aan ervaring op het gebied van beoordelen. De beoordelaars bestaan uit studenten die niet vaak te maken hebben met het beoordelen van spreekvaardigheid en voor wie het dus ook onwennig is. De beoordelaars zijn onervaren in het gebruiken van een dergelijk model in vergelijking tot docenten die een dergelijk model in de onderwijspraktijk gebruiken. Het is goed mogelijk dat het gebrek aan ervaring een rol heeft gespeeld bij de beoordeling van de fragmenten. Om dit zoveel mogelijk tegen te gaan, is er een instructie toegevoegd aan het model en is er een mondelinge toelichting gegeven aan de studenten over het gebruik van het model. Op deze manier zijn de studenten enigszins vertrouwd gemaakt met het gebruiken van een dergelijk model.

In dit hoofdstuk is beschreven hoe het onderzoek opgezet is. In het volgende hoofdstuk wordt een overzicht gegeven van de resultaten van de beoordelingen van de taken.

4. Resultaten

Er zijn binnen dit onderzoek 105 video-opnames verzameld. Deze opnames zijn opgesplitst in drie delen: een begin-, midden- en eindfragment. Er zijn dus 315 fragmenten die beoordeeld konden worden. Wegens technische mankementen en tijdsdruk zijn niet alle beoordelaars erin geslaagd de aan hun toebedeelde fragmenten te analyseren. Uiteindelijk zijn er slechts 60 fragmenten die door drie beoordelaars gescoord zijn. Van 141 fragmenten zijn twee beoordelingen beschikbaar en van 255 fragmenten is één beoordeling beschikbaar.

Allereerst wordt een interbeoordelaarsbetrouwbaarheid gegeven. Uit deze gegevens kan worden afgeleid in hoeverre de oordelen van beoordelaar één, twee en drie overeen komen. Vervolgens wordt beschreven in hoeverre de verschillende vragen van het beoordelingsmodel met elkaar samenhangen (per criterium en overall). Daarna wordt een antwoord op de hoofdvraag gegeven. Er wordt gekeken in hoeverre de gesprekstaken effect gehad hebben op de spreekvaardigheid. Daarbij wordt onderscheid gemaakt tussen een effect van taak, onderwerp en hulpmiddel. Hieruit kunnen we afleiden of het uitvoeren van gesprekstaken nut heeft voor de ontwikkeling van de spreekvaardigheid.

4.1 Interbeoordelaarsbetrouwbaarheid

De beoordelaars zijn ingedeeld per meetmoment. Dit houdt in dat elke groep beoordelaars naar een type fragment uit de video-opname kijkt. Een beoordelaar bekijkt dus alleen het begin, midden of eind van het fragment. Het is belangrijk dat de beoordelaars op één lijn zitten bij het beoordelen van de fragmenten. Dit geeft namelijk aan dat het beoordelingsmodel duidelijk is en dat de beoordelaars de fragmenten op een gelijke manier beoordelen. Om dit in kaart te brengen wordt er een interbeoordelaarsbetrouwbaarheid uitgerekend. De interbeoordelaarsbetrouwbaarheid per vraag staat in tabel 1. Hierbij is alleen de groep beoordelaars van de eindfragmenten meegenomen. Dit komt door het lage aantal beoordelingen van de begin- en middenfragmenten. Bij de beginfragmenten zijn er slechts 4 fragmenten die door drie beoordelaars beoordeeld zijn. Bij de middenfragmenten zijn er 15 fragmenten met drie beoordelaars. Dit is allebei weinig in vergelijking tot de eindfragmenten (N = 41). De interbeoordelaarsbetrouwbaarheid is weinig veelzeggend wanneer er slechts 4 items zijn meegenomen in de analyse (beginfragmenten). Daarom is ervoor gekozen om alleen een Cronbach's Alpha te berekenen van de beoordelaars op de eindfragmenten.

Vraag	*Alpha B1+B2 (N=72)*	*Alpha B1+B2+B3 (N=41)*
1	,47	,70 (,74 zonder B1)
2	,51	,69 (,72 zonder B1)
3	,15	,40 (,65 zonder B1)
4	,23	,30
5	,22	,38 (,46 zonder B3)
6	,35	,71 (,79 zonder B2)
7	,53	,68 (,69 zonder B3)
8	,46	,65
9	,23	,30 (,38 zonder B2)
10	0	,30
11	0	,14 (,56 zonder B1)
12	,62	,64

Tabel 1. Interbeoordelaarsbetrouwbaarheid bij eindfragmenten

Als we naar de tweede kolom van de tabel kijken (Alpha B1+B2), dan zien we dat de beoordelaars 1 en 2 bij geen enkele vraag op één lijn zitten. De Cronbach's Alpha is bij geen enkele vraag hoog genoeg om acceptabel te zijn. Hiervoor wordt namelijk de benedengrens van ,70 aangehouden. Alleen de Alpha bij vraag 12 is redelijk hoog, maar nog niet hoog genoeg om betrouwbaar te zijn.

Wanneer de scores van beoordelaar 3 ook meegenomen worden (derde kolom), dan is de Cronbach's Alpha op verschillende vragen wel hoog genoeg om betrouwbaar genoemd te worden. Dit geldt voor vraag 1 en 6. De scores op de vragen 2 en 7 komen dicht in de buurt van de benedengrens van ,70.

De beoordelaars zitten dus niet voor elke vraag op één lijn. Bij verschillende vragen wordt de Cronbach's Alpha (veel) hoger wanneer één van de beoordelaars niet meegenomen wordt. Dit is te zien in de tabel in kolom drie tussen haakjes. Het gebrek aan overeenstemming tussen de

beoordelaars kan komen door persoonlijke beweegredenen van de beoordelaars (denk bijvoorbeeld aan de *persoonlijke vergelijking* uit hoofdstuk drie), maar het kan ook betekenen dat het beoordelingsmodel niet duidelijk genoeg was. Het houdt in elk geval in dat de gevonden resultaten in dit onderzoek niet zomaar aangenomen mogen worden, aangezien er onvoldoende overeenstemming is tussen de beoordelaars. Bij de verdere analyse van de resultaten zal gewerkt worden met de data van beoordelaar 1. Dit is niet voor de hand liggend, omdat de interbeoordelaarsbetrouwbaarheid bij sommige vragen (een stuk) hoger wordt wanneer de scores van beoordelaar 1 niet meegenomen worden. Hier staat tegenover dat er ook vragen zijn waarbij de interbeoordelaarsbetrouwbaarheid hoger wordt wanneer de scores van beoordelaar 2 of 3 niet meegenomen worden. Er is uiteindelijk gekozen om alleen met beoordelaar 1 verder te werken, omdat dan een zo groot mogelijk aantal items wordt meegewogen in de analyse (N = 255). Dit zijn de scores van zes verschillende beoordelaars.

4.2 Inter-item correlatie

Het is niet alleen van belang dat de scores van de beoordelaars betrouwbaar met elkaar samenhangen. Het geldt ook voor de vragen uit het beoordelingsmodel. Per criterium/aspect van spreekvaardigheid zijn een aantal vragen gesteld (alleen bij 'algemene indruk' is er maar één vraag gesteld). Het is belangrijk om na te gaan of deze verschillende vragen wel hetzelfde aspect (bijvoorbeeld 'uitspraak') meten. De verwachting is dat wanneer een spreker hoog scoort op vraag 1 (een deelvraag van 'complexiteit') hij dan ook hoog scoort op vraag 2 (de andere deelvraag van 'complexiteit'). Ook de inter-item correlatie van alle vragen bij elkaar moet betrouwbaar zijn. Er wordt namelijk verwacht dat alle vragen samen iets zeggen over de spreekvaardigheid. In tabel 2 staan de inter-item correlaties per aspect van spreekvaardigheid en voor alle vragen tezamen ('overall'). Het aspect 'algemene indruk' is hierbij niet meegenomen, omdat hij maar één vraag over gesteld werd. Deze vraag (vraag 12) is wel meegenomen bij de berekening van de correlatie van alle vragen bij elkaar.

Aspect	*Vragen*	*Cronbach's Alpha*
Complexiteit	1 - 2	,81
Nauwkeurigheid	3 - 5	,71
Vloeiendheid	6 - 8	,86
Uitspraak	9 - 11	,80
Overall	1 - 12	,91

Tabel 2. Inter-item correlaties

In tabel 2 is te zien dat de Cronbach's Alpha voor alle aspecten hoger is dan de betrouwbaarheidsgrens van ,70. Dit betekent dat de vragen per aspect betrouwbaar samenhangen (dat vraag 1 inderdaad samenhangt met vraag 2 als onderdeel van 'complexiteit'). De Cronbach's Alpha van alle vragen tezamen is ,91. Deze hoge correlatie geeft aan dat de vragen betrouwbaar samenhangen en dat alle vragen dus een deel van de spreekvaardigheid meten. Doordat de vragen betrouwbaar samenhangen mag er per aspect van spreekvaardigheid een gemiddelde somscore berekend worden.

Het is goed om te vermelden dat er voor de analyse van de scores meerweg ANOVA's (Univariate) zijn gebruikt, terwijl toetsing met behulp van 'repeated measures' meer op zijn plaats is. Het gaat immers om een herhaalde uitvoering. Een analyse op basis van een repeated measures is binnen dit onderzoek echter niet haalbaar, aangezien er te weinig data beschikbaar zijn. Van de meeste proefpersonen is namelijk geen volledige beoordeling beschikbaar van alle taken (er ontbreekt bijvoorbeeld een score op taak 1 – begin, taak 1 – midden, taak 1 – eind, taak 2 – begin, etc.). Om zoveel mogelijk gegevens te analyseren (om toch een goed beeld te krijgen van mogelijke effecten van taakuitvoering), is gekozen om meerweg ANOVA's te gebruiken. Hiermee kunnen ook eventuele interactie-effecten berekend worden. In de volgende drie paragrafen worden de resultaten per effect besproken (herhaling, onderwerp, hulpmiddel).

4.3 Effect van herhaling

In het theoretisch kader zijn verschillende onderzoeken naar voren gekomen die een effect van taakherhaling lieten zien. In deze paragraaf wordt nagegaan of eenzelfde effect gevonden kan worden binnen het huidige onderzoek. De resultaten zullen per aspect van spreekvaardigheid besproken worden. Het aantal beoordelingen dat is meegenomen bij de analyses (N) staat in tabel 3.

	Taak 1	*Taak 2*	*Taak 3*	Totaal
Begin	38	37	30	105
Midden	9	36	28	73
Eind	23	29	25	77

Tabel 3. Meegenomen beoordelingen (N) per taak per opnamemoment

4.3.1 Complexiteit

Om erachter te komen of het uitvoeren van meerdere taken een rol speelt op de score voor complexiteit is de gemiddelde somscore op complexiteit van taak 1, 2 en 3 met elkaar vergeleken. In tabel 4 is te zien wat de gemiddelde somscores (met de standaarddeviatie tussen haakjes) op deze taken zijn. Deze scores zijn uitgesplitst per opnamemoment (begin, midden of eind van het gesprek).

	Taak 1	*Taak 2*	*Taak 3*
Begin	7,92 (1,72)	7,27 (2,27)	8,57 (1,52)
Midden	7,44 (1,13)	6,33 (2,43)	6,64 (2,45)
Eind	7,70 (2.08)	8,38 (2,29)	8,84 (1,82)

Tabel 4. Gemiddelde somscores op complexiteit

Met behulp van een meerweg ANOVA is berekend of de verschillende somscores van elkaar verschillen. Hieruit blijkt dat er een hoofdeffect van opnamemoment is ($F = 8,32$; $df = 2, 246$; $p < 0,001$). Er is ook een hoofdeffect van taak ($F = 2,51$; $df = 2, 246$; $p = 0,08$). Deze p-waarde is significant, aangezien hij door twee gedeeld mag worden, omdat er eenzijdig getoetst wordt. Er is geen interactie-effect van opnamemoment en taak. Er is dus een verschil in scores afhankelijk van

het moment in het gesprek en een verschil afhankelijk van het taaknummer, maar deze factoren beïnvloeden elkaar niet. Een Post Hoc (Scheffe) toets van de scores op opnamemoment laat zien wat deze verschillen inhouden. Uit deze toets blijkt dat de scores op het beginmoment verschillen van de scores op het middenmoment (MD = 1,29; SE = 0,32; p < 0,001), waarbij geldt 'MD' = 'Mean Difference' en 'SE' = 'Standard Error'. Dit betekent dat de scores op complexiteit aan het begin van het gesprek hoger zijn dan de scores op complexiteit in het midden van het gesprek. Ook verschillen de scores op het middenmoment van de scores op het eindmoment (MD = -174; SE = 0,34; p < 0,001). De scores op het middenmoment zijn lager dan de scores op het eindmoment. De scores op het eindmoment verschillen niet significant van de scores op het beginmoment (MD = 0,45; SE = 0,31; p = 0,36). Een tweede Post Hoc toets (van de scores op taak) laat zien dat er geen verschil is tussen de scores op taak 1 en taak 2 (MD = 0,53; SE = 0,32; p = 0,26) en geen verschil op taak 1 en 3 (MD = -0,21; 0,34; p = 0,82). Er is wel een verschil tussen de scores op taak 2 en 3 (MD = -0,75; SE = 0,31; p = 0,05). Er wordt hoger gescoord op taak 3 dan op taak 2.

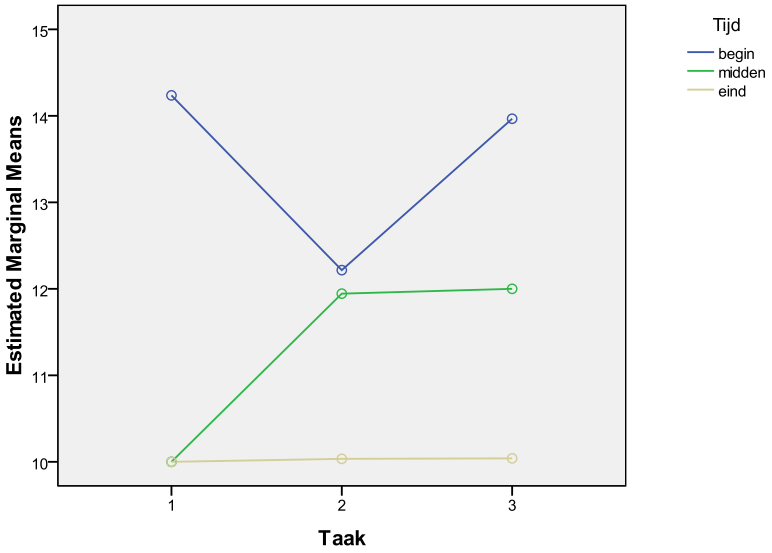
4.3.2 Nauwkeurigheid

Om erachter te komen of het uitvoeren van meerdere taken ook een rol speelt op de score voor nauwkeurigheid is de gemiddelde somscore op nauwkeurigheid van taak 1, 2 en 3 met elkaar vergeleken. In tabel 5 is te zien wat de gemiddelde somscores (met de standaarddeviatie tussen haakjes) op deze taken zijn. Deze scores zijn uitgesplitst per opnamemoment (begin, midden of eind van het gesprek).

	Taak 1	*Taak 2*	*Taak 3*
Begin	14,24 (1,97)	12,22 (3,09)	13,97 (2,06)
Midden	10,00 (1,66)	11,94 (2,56)	12,00 (3,06)
Eind	10,00 (2.07)	10,03 (2,47)	10,04 (1,46)

Tabel 5. Gemiddelde somscores op nauwkeurigheid

Een meerweg ANOVA laat zien dat er geen hoofdeffect van taak is (F = 1,65; df = 2, 246; p = 0,20). Er is wel een hoofdeffect van opnamemoment (F = 46,47; df = 2, 246; p < 0,001), maar dit effect valt weg, omdat er ook een interactie-effect van opnamemoment en taak is (F = 4,09; df = 4, 246; p = 0,003). Er is dus een verschil in scores afhankelijk van het taaknummer én het moment in het gesprek. Uit een Post Hoc toets blijkt dat de scores op nauwkeurigheid aan het begin van het gesprek verschillen van de scores op het middenmoment (MD = 1,72; SE = 0,37; p < 0,001) en van de scores op het eindmoment (MD = 3,42; SE = 0,36; p < 0,001). De proefpersonen scoren hoger op nauwkeurigheid aan het begin dan in het midden of aan het eind van het gesprek. De scores op het midden verschillen ook significant van de scores op het eind (MD = 1,70; SE = 0,39; p < 0,001). De proefpersonen scoren hoger op nauwkeurigheid op het midden dan op het eind van het gesprek. Grafiek 1 laat zien dat de verschillende scores op de fragmenttypen niet overal even ver uit elkaar liggen. Dit komt door het interactie-effect.



Grafiek 1. Interactie-effect opnamemoment (tijd) en taak op nauwkeurigheid

In de grafiek is te zien dat de lijnen (gemiddelde score per opnamemoment) niet parallel lopen. Bij taak 2 liggen de scores op begin en midden erg dicht op elkaar, terwijl dat bij taak 1 en 3 niet het geval is (bij taak 1 nog minder dan bij taak 3). De scores op het begin en het midden worden beïnvloed door het taaknummer. Dit geldt niet voor de scores op 'eind', aangezien die lijn vlak loopt. De grafiek laat zien dat het verschil tussen de scores op begin en midden/eind bij taak 1 veel groter is dan bij taak 2 en 3. Bij taak 2 liggen de scores van begin en midden zelfs heel dicht bij elkaar. Maar de lijn met de scores op 'begin' blijft hoger dan de andere twee lijnen. Dit houdt in dat de proefpersonen, naarmate ze meerdere taken uitvoeren, beter blijven presteren aan het begin van het gesprek dan in het midden of aan het eind van het gesprek.

4.3.3 Vloeiendheid

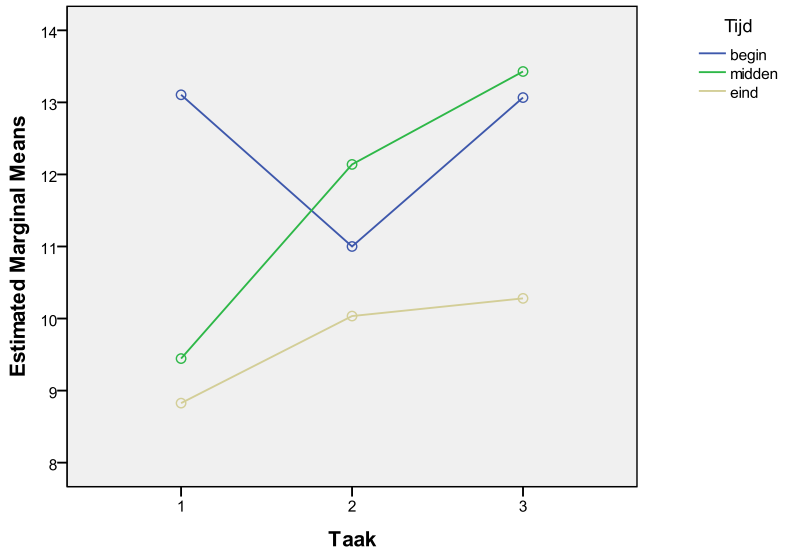
Het derde aspect van spreekvaardigheid dat geanalyseerd wordt, is vloeiendheid. In tabel 6 staan de gemiddelde somscores per taak en opnamemoment voor vloeiendheid.

	Taak 1	*Taak 2*	*Taak 3*
Begin	13,11 (3,46)	11,00 (3,57)	13,07 (3,23)
Midden	9,44 (2,13)	12,14 (3,91)	13,43 (3,90)
Eind	8,83 (2,71)	10,03 (3,45)	10,28 (2,42)

Tabel 6. De gemiddelde somscores op vloeiendheid

Uit een meerweg ANOVA blijkt dat er een hoofdeffect van opnamemoment is ($F = 14{,}05$; $df = 2, 246$; $p < 0{,}001$) en een hoofdeffect van taaknummer ($F = 5{,}07$; $df = 2, 246$; $p = 0{,}01$). Deze hoofdeffecten vallen weg, omdat er tevens een interactie-effect is ($F = 3{,}56$; $df = 2, 246$; $p = 0{,}01$). Dit houdt in dat de scores op vloeiendheid zowel afhankelijk zijn van het taaknummer als van het moment van opname (begin, midden, eind). Uit de Post Hoc toets blijkt dat de scores op het beginmoment niet verschillen van de scores op het middenmoment (MD = 0,05; SE = 0,52; $p > 0{,}99$), maar wel van de

scores op het eindmoment (MD = 2,60; SE = 0,51; p < 0,001). De scores op het middenmoment verschillen ook van het eindmoment (MD = 2,55; SE = 0,55; p < 0,001). De proefpersonen scoren dus hoger op vloeiendheid aan het begin en in het midden van het gesprek dan aan het eind. Uit een tweede Post Hoc toets (voor taak) blijkt dat de scores op taak 1 niet verschillen van de scores op taak 2 (MD = 0,10; SE = 0,53; p = 0,98) en ook niet van de scores op taak 3 (MD = -1,12; SE = 0,55; p = 0,13). De scores op taak 2 verschillen wel van taak 3 (MD = -1,22; SE = 0,50; p = 0,05). Op taak 3 wordt hoger gescoord dan op taak 2. Het interactie-effect is te zien in grafiek 2.



Grafiek 2. Interactie-effect opnamemoment (tijd) en taak op vloeiendheid

In de grafiek is te zien dat de scores bepaald worden door het taaknummer én door het opnamemoment. Op de verschillende taken (taaknummers) worden uiteenlopende scores behaald en op de verschillende opnamemomenten worden uiteenlopende resultaten behaald. De lijnen lopen niet parallel aan elkaar. De score op 'begin' bij taak 1 en 3 is veel hoger dan bij taak 2. Deze score wordt dus beïnvloed door het taaknummer. Dit is het gevolg van het interactie-effect.

4.3.4 Uitspraak

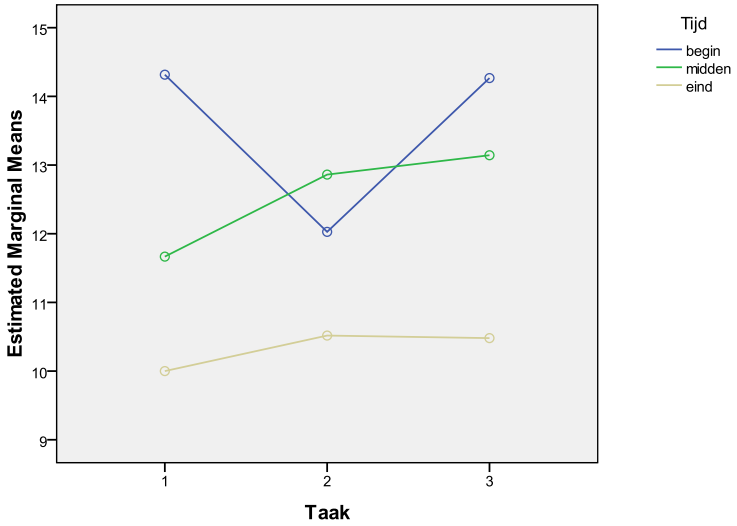
Het vierde aspect van spreekvaardigheid is uitspraak. In tabel 7 staan de gemiddelde somscores op uitspraak.

	Taak 1	*Taak 2*	*Taak 3*
Begin	14,32 (2,21)	12,03 (3,30)	14,27 (2,03)
Midden	11,67 (1,87)	12,86 (1,78)	13,14 (2,64)
Eind	10,00 (2,47)	10,52 (2,69)	10,48 (2,02)

Tabel 7. De gemiddelde somscores op uitspraak

Een meerweg ANOVA laat zien dat er zowel een hoofdeffect van opnamemoment (F = 38,67; df = 2, 246; p < 0,001) als een hoofdeffect van taak is (F = 2,74; df = 2, 246; p = 0,07). Deze p-waarde is significant, omdat hij gehalveerd mag worden wegens eenzijdige toetsing. Deze effecten vallen weg,

omdat er ook een interactie-effect van opnamemoment en taak is (F = 4,31; df = 4, 246; p = 0,002). Er is dus een verschil in scores afhankelijk van het taaknummer én het moment in het gesprek. Uit een Post Hoc toets blijkt dat de scores op uitspraak op het beginmoment niet verschillen van de scores op het middenmoment (MD = 0,67; SE = 0,37; p = 0,19), maar wel verschillen van de scores op het eindmoment (MD = 3,14; SE = 0,37; p < 0,001). De scores op het middenmoment verschillen ook van de scores op het eindmoment (MD = 2,47; SE = 0,40; p < 0,001). Aan het begin en in het midden van de gesprekken scoren de proefpersonen dus hoger op uitspraak dan aan het eind van het gesprek. Een tweede Post Hoc toets laat zien dat de scores op taak 1 en 2 niet verschillen (MD = 0,66; SE = 0,38; p = 0,21), net zoals de scores op taak 1 en 3 (MD = -0,19; SE = 0,40; p = 0,89). De scores op taak 2 en 3 verschillen wel (MD = -0,66; SE = 0,38; p = 0,06). Deze p-waarde mag wederom gehalveerd worden. De proefpersonen scoren dus hoger op taak 3 dan op taak 2.
Grafiek 3 laat zien dat de verschillende scores per opnamemoment niet parallel lopen.



Grafiek 3. Interactie-effect opnamemoment (tijd) en taak op uitspraak

De scores op het beginmoment en het middenmoment lopen (net als bij nauwkeurigheid) niet parallel aan elkaar. Dit heeft te maken met de invloed van het taaknummer (interactie-effect). Bij taak 2 is er een dieptepunt in de score op het beginmoment. Dit is opvallend, want dit is ook te zien bij de score op het beginmoment op nauwkeurigheid (grafiek 1). Het lijkt erop dat de proefpersonen bij de tweede taak aanvankelijk moeite hadden met spreken.

4.3.5 Algemene indruk

Het laatste aspect van spreekvaardigheid waarop de proefpersonen beoordeeld zijn, is de algemene indruk. In tabel 8 staan de gemiddelde scores op algemene indruk.

	Taak 1	*Taak 2*	*Taak 3*
Begin	5,96 (0,93)	5,14 (1,53)	5,88 (1,06)
Midden	5,11 (0,93)	5,67 (1,60)	6,04 (1,53)
Eind	4,41 (1,66)	4,69 (1,42)	4,96 (1,44)

Tabel 8. De gemiddelde scores op algemene indruk

Een meerweg ANOVA laat zien dat er bij algemene indruk twee hoofdeffecten zijn, namelijk van opnamemoment ($F = 12,42$; $df = 2, 246$; $p < 0,001$) en van taak ($F = 3,05$; df 2, 246; $p = 0,05$). Er zijn dus verschillen te vinden in de scores op verschillende momenten in de opnames en in de scores op de verschillende taken. Er is geen interactie-effect. Uit de Post Hoc toets blijkt wat deze verschillen inhouden. Voor opnamemoment geldt dat de scores op het beginmoment niet verschillen van de scores op het middenmoment (MD = -0,09; SE = 0,21; $p = 0,91$), maar wel van het eindmoment (MD = 0,95; SE = 0,21; $p < 0,001$). De proefpersonen scoren beter op het beginmoment dan op het eindmoment. Dit verschil is ook te vinden tussen de scores op het middenmoment en het eindmoment. Er wordt hoger gescoord op het middenmoment dan op het eindmoment (MD = 1,05; SE = 0,22; $p < 0,001$).
Voor taaknummer geldt dat alleen de scores op taak 2 verschillen van de scores op taak 3 (MD = -0,46; SE = 0,20; $p = 0,08$). Deze p-waarde mag door twee gedeeld worden omdat er eenzijdig getoetst wordt. Op taak 3 wordt hoger gescoord dan op taak 2. Er is geen verschil tussen de scores van taak 1 en 3 (MD = 0,15; SE = 0,21; $p = 0,79$) en de scores van taak 1 en 2 (MD = -0,31; SE = 0,22; $p = 0,37$).

4.3.6 Samenvatting effect van herhaling

Bij alle aspecten van spreekvaardigheid is te zien dat het opnamemoment zorgt voor een verschil in scores. Het is dus van belang om goed na te denken over de selectie van fragmenten voor de beoordeling van spreekvaardigheid. Op vier van de vijf aspecten van spreekvaardigheid zijn de scores aan het begin en in het midden van het gesprek hoger dan aan het eind van het gesprek. Dit geldt alleen niet voor complexiteit. Daar zijn de begin- en eindscores hoger dan de middenscores. De scores worden dus (deels) bepaald door het moment in het gesprek.

Als gekeken wordt naar het effect van taakherhaling, dan valt op dat de proefpersonen niet altijd hoger scoren naarmate ze meer taken uitvoeren. Bij vier van de vijf aspecten van spreekvaardigheid wordt er alleen een verschil gevonden tussen de scores op taak 2 en taak 3. Dit verschil is niet terug te vinden bij nauwkeurigheid. Hier geldt een interactie-effect. Het taaknummer hangt samen met het opnamemoment. Zij beïnvloeden gezamenlijk de score op nauwkeurigheid. In de resultaten is geen verschil te vinden tussen de scores op taak 1 en 2. Dit houdt in dat een effect van herhaling pas merkbaar is wanneer er al twee taken gedaan zijn. De scores op taak 2 en 3 verschillen immers wel. Bij de uitvoering van de tweede taak is er nog geen effect van herhaling, maar bij de uitvoering van de derde taak wel. Om dus een effect van herhaling te vinden, moeten er minstens drie taken uitgevoerd worden.

4.4 Effect van onderwerp

Naast een effect van taakuitvoering is er in het theoretisch kader ook stilgestaan bij een effect van onderwerp. Binnen dit onderzoek worden vier verschillende onderwerpen behandeld. Dit betekent niet dat in elke taak alle vier de onderwerpen naar voren komen of dat elke proefpersonen alle vier de onderwerpen behandeld (in één van de drie taken). Het komt voor dat een spreker drie keer een taak uitvoert over hetzelfde onderwerp (bijvoorbeeld feestdagen of relaties). In tabel 9 staat de

verdeling van de onderwerpen over de taken. Opgemerkt moet worden dat de onderwerpen 'studentenleven' en 'relaties' bij de analyse samengevoegd zijn tot 'persoonlijk leven'. Op deze manier zijn de onderwerpen qua totaal aantal gelijkmatig verdeeld over de taken en is er een duidelijker contrast tussen de verschillende onderwerpen. Persoonlijk leven is heel concreet, feestdagen is al iets abstracter (over feestdagen in het algemeen en de persoonlijke viering van feestdagen) en cultuuruitwisseling is het meest abstract.

	Taak 1	*Taak 2*	*Taak 3*	Totaal
Cultuuruitwisseling	27	30	24	81
Feestdagen	15	54	21	90
Persoonlijk leven	28	18	38	84
Totaal	70	102	83	255

Tabel 9. De verdeling van de onderwerpen over de taken.

De resultaten zullen, net als bij het effect van taakherhaling, besproken worden per aspect van spreekvaardigheid.

4.4.1 Complexiteit

Het eerste aspect van spreekvaardigheid dat geanalyseerd wordt is complexiteit. In tabel 10 staan de gemiddelde somscores op complexiteit per onderwerp, per taak en per opnamemoment.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Cultuur-uitwisseling	7,83 (1,19)	7,33 (1,53)	8,42 (2,11)	7,60 (2,72)	6,80 (2,97)	8,80 (1,81)	8,75 (1,91)	7,00 (3,02)	8,63 (1,19)
Feestdagen	6,86 (1,77)	7,00 (-)	6,86 (1,77)	7,33 (2,20)	6,61 (2,30)	8,44 (2,26)	8,86 (0,69)	5,86 (1,95)	8,71 (2,43)
Persoonlijk leven	8,37 (1,86)	7,60 (1,14)	7,00 (2,16)	6,78 (2,05)	5,13 (1,73)	3,00 (-)	8,33 (1,63)	6,85 (2,41)	9,10 (1,91)

Tabel 10. De gemiddelde somscores op complexiteit

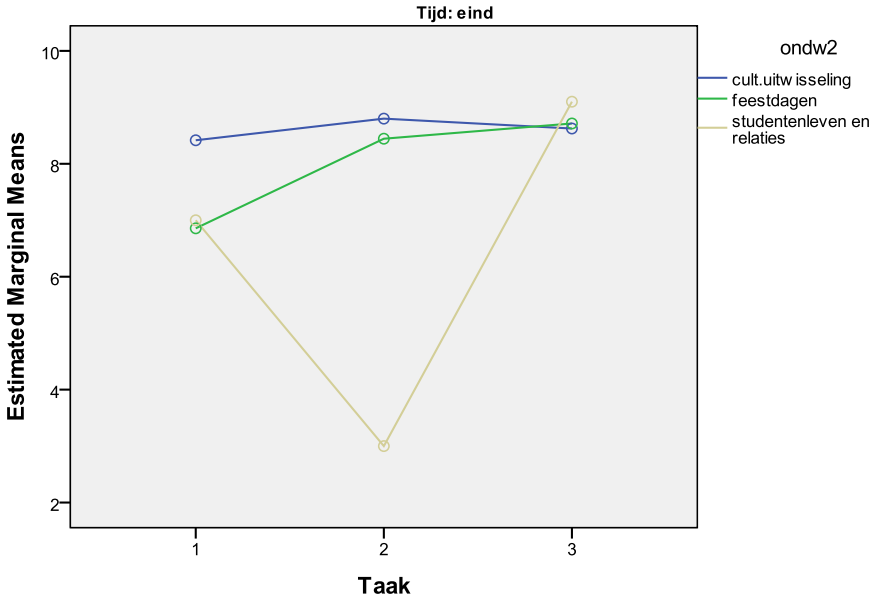
Met behulp van een aantal meerweg ANOVA's is berekend of de gemiddelde somscores van elkaar verschillen. Dit is berekend per opnamemoment, dus voor het begin, midden en einde van het gesprek.

Voor het *begin* geldt dat uit de ANOVA blijkt dat er een hoofdeffect is van taak ($F = 4,24$; $df = 2, 96$; $p = 0,02$), maar geen hoofdeffect van onderwerp ($F = 0,29$; $df = 2, 96$; $p = 0,75$). Er is ook geen interactie-effect ($F = 1,10$; $df = 4, 96$; $p = 0,36$). Het taaknummer zorgt dus wel voor een verschil in scores op complexiteit, maar het onderwerp van de taak niet. Een Post Hoc (Scheffe) toets laat zien wat dit verschil inhoudt. De scores op taak 2 zijn lager dan op taak 3 (MD = -1,30; SE = 0,47; $p = 0,02$).

De proefpersonen scoren beter op complexiteit bij taak 3 dan bij taak 2. Tussen de scores op taak 1 en 2 (MD = 0,65; SE = 0,44; p = 0,34) en taak 1 en 3 (MD = -0,65; SE = 0,46; p = 0,38) is geen verschil.

Bij de *middenfragmenten* is er wederom geen hoofdeffect van onderwerp (F = 0,26; df = 2, 64; p = 0,77), maar ook geen hoofdeffect van taak (F = 0,63; df = 2, 64; p = 0,53) en ook geen interactie-effect (F = 0,78; df = 4, 64; p = 0,54). Voor de middenfragmenten geldt dus dat de scores niet afhankelijk zijn van het taaknummer of het onderwerp van de taak.

Voor de *eindfragmenten* zijn er wel twee hoofdeffecten te vinden. Er is een effect van taak (F = 4,27; df = 2, 68; p = 0,02) en een effect van onderwerp (F = 3,44; df = 2, 68; p = 0,04). Deze effecten worden verder niet geïnterpreteerd, omdat er ook een interactie-effect is van taak en onderwerp (F = 2,31; df = 4, 68; p = 0,07). De p-waarde mag (net als bij 'complexiteit' en 'algemene indruk' in paragraaf 4.3) door twee gedeeld worden omdat er eenzijdig getoetst wordt. Bij de eindfragmenten is de score op een onderwerp is dus afhankelijk van het taaknummer. In grafiek 4 is te zien wat dit interactie-effect inhoudt.



Grafiek 4. Interactie-effect taak en onderwerp op complexiteit.

In de grafiek is te zien dat het niet zo is dat proefpersonen op 'persoonlijk leven' (= studentenleven en relaties) altijd lager scoren dan op 'feestdagen'. De score op 'persoonlijk leven' is bij taak 1 en taak 3 immers hoger dan de score op 'feestdagen'. Hetzelfde geldt voor de score op 'culturele uitwisseling'. Die is niet altijd hoger dan de score op de andere twee onderwerpen. Ook wordt er niet hoger gescoord naarmate het aantal taakuitvoeringen toeneemt (de score op taak 2 is niet altijd hoger dan op taak 1 en de score op taak 3 is niet altijd hoger dan op taak 1 of 2). De score op een onderwerp is dus afhankelijk van het taaknummer. Bij taak 2 is het bijvoorbeeld verstandig om niet te kiezen voor het onderwerp 'persoonlijk leven', terwijl dat voor taak 1 of 3 niet (veel) uitmaakt.

4.4.2 Nauwkeurigheid

Het tweede aspect van spreekvaardigheid is nauwkeurigheid. In tabel 11 staan de gemiddelde somscores op dit aspect.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Cultuur-uitwisseling	14,17 (2,21)	10,00 (1,00)	10,08 (1,93)	12,80 (2,20)	13,50 (1,72)	10,20 (2,35)	14,63 (2,26)	12,63 (3,82)	9,88 (1,73)
Feestdagen	13,14 (2,19)	8,00 (-)	8,86 (1,68)	11,83 (3,42)	11,44 (2,23)	9,89 (2,65)	14,29 (1,98)	10,57 (2,30)	9,81 (2,28)
Persoonlijk leven	14,68 (1,64)	10,40 (1,95)	11,75 (2,22)	12,33 (3,46)	11,13 (3,48)	11,00 (-)	13,47 (2,00)	12,38 (2,87)	10,40 (1,68)

Tabel 11. De gemiddelde somscores op nauwkeurigheid

Uit de meerweg ANOVA blijkt dat er op het *beginmoment* alleen een hoofdeffect van taak is (F = 5,33; df = 2, 96; p = 0,01). Er is geen hoofdeffect van onderwerp (F = 0,69; df = 2, 96; p = 0,51) en geen interactie-effect van taak en onderwerp (F = 0,72; df = 4, 96; p = 0,58). Dit betekent dat er in de scores voor nauwkeurigheid alleen verschillen bestaan tussen taken. Dit is niet afhankelijk van het onderwerp. Een Post Hoc toets laat zien waar deze verschillen voorkomen. Hieruit blijkt namelijk dat de scores op taak 1 verschillen van taak 2 (MD = 2,02; SE = 0,57; p = 0,01). Op taak 1 wordt hoger gescoord dan op taak 2 (onafhankelijk van het onderwerp van de taak). Er is geen verschil tussen de scores op taak 1 en 3 (MD = 0,27; SE = 0,60; p = 0,90). Daarnaast blijkt dat er een verschil is tussen de scores op taak 2 enerzijds en de scores op taak 3 anderzijds (MD = -1,75; SE = 0,61; p = 0,02). Er wordt op taak 3 hoger gescoord dan op taak 2.

Op het *middenmoment* zijn geen effecten te vinden. Er is geen hoofdeffect van taak (F = 2,39; df = 2, 64; p = 0,11), geen hoofdeffect van onderwerp F = 1,53; df = 2, 64; p = 0,22) en geen interactie-effect (F = 0,70; df = 4, 64; p = 0,59). Op het middenmoment is er dus geen verschil te vinden in de scores afhankelijk van het taaknummer en/of het onderwerp.

Dit geldt ook voor de scores op het *eindmoment*. Er is geen hoofdeffect van taak (F = 0,06; df = 2, 68; p = 0,94), geen hoofdeffect van onderwerp (F = 0,73; df = 2, 68; p = 0,49) en geen interactie-effect (F = 1,26; df = 4, 68; p = 0,29). Ook hier is er dus geen verschil te vinden in de scores. De scores op nauwkeurigheid zijn aan het eind van het gesprek dus niet afhankelijk van het taaknummer en/of het onderwerp.

4.4.3 Vloeiendheid

Naast complexiteit en nauwkeurigheid zijn de proefpersonen gescoord op vloeiendheid. Ook hier wordt gekeken of het onderwerp effect heeft op de scores op vloeiendheid. In tabel 12 staan de gemiddelde somscores op vloeiendheid.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Cultuur-uitwisseling	13,17 (2,98)	8,33 (1,15)	8,83 (2,25)	10,90 (3,57)	11,80 (3,65)	11,10 (2,81)	12,75 (4,40)	13,75 (4,13)	9,00 (2,51)
Feestdagen	10,00 (4,12)	11,00 (-)	9,00 (3,37)	11,11 (3,53)	13,00 (3,53)	9,67 (3,69)	13,43 (2,82)	12,71 (4,54)	10,14 (1,57)
Persoonlijk leven	14,21 (2,92)	9,80 (2,59)	8,50 (3,51)	10,89 (4,04)	10,63 (4,93)	6,00 (-)	13,07 (2,89)	13,62 (3,69)	11,40 (2,50)

Tabel 12. De gemiddelde somscores op vloeiendheid

Uit de meerweg ANOVA blijkt dat er op het *beginmoment* alleen een hoofdeffect van taak is (F = 3,17; df = 2, 96; p = 0,05). Er is geen hoofdeffect van onderwerp (F = 1,00; df = 2, 96; p = 0,37) en geen interactie-effect (F = 1,56; df = 4, 96; p = 0,19). Er is alleen een verschil in scores op vloeiendheid afhankelijk van het taaknummer. Wat dit verschil inhoudt, blijkt uit de Post Hoc toets. Die toets laat zien dat de scores op taak 1 verschillen van de scores op taak 2 (MD = 2,11; SE = 0,79; p = 0,03), maar niet van de scores op taak 3 (MD = 0,04; SE = 0,83; p > 0,99). Op taak 1 wordt dus hoger gescoord dan op taak 2, maar met de scores op taak 3 is er geen verschil. De scores op taak 2 verschillen wel van taak 3 (MD = -2,07; SE = 0,84; p = 0,05). Er wordt hoger gescoord op taak 3 dan op taak 2.

Voor het *middenmoment* is er ook alleen een hoofdeffect van taak (F = 2,60; df = 2, 64; p = 0,08). Deze p-waarde mag door twee gedeeld worden (i.v.m. eenzijdige toetsing). Er is geen hoofdeffect van onderwerp (F = 0,18; df = 2, 64; p = 0,84) en geen interactie-effect (F = 0,59; df = 4, 64; p = 0,67). Uit een Post Hoc toets blijkt er alleen een verschil is tussen de scores op taak 1 en 3 (MD = -3,98; SE = 1,47; p = 0,03). De proefpersonen scoren hoger op taak 3 dan op taak 1. Er is geen verschil tussen de scores op taak 1 en taak 2 (MD = -2,69; SE = 1,43; p = 0,18) en ook geen verschil tussen de scores op taak 2 en 3 (MD = -1,29; SE = 0,96; p = 0,41).

Voor het *eindmoment* geldt dat er geen effect gevonden kan worden. Er is geen hoofdeffect van taak (F = 1,40; df = 2, 68; p = 0,26), geen hoofdeffect van onderwerp (F = 0,35; df = 2, 68; p =0,71) en geen interactie-effect (F = 1,53; df = 4, 68; p = 0,20). Op het eindmoment wordt de score op vloeiendheid dus niet bepaald door het taaknummer en/of het onderwerp.

4.4.4 Uitspraak

Een vierde aspect van spreekvaardigheid is uitspraak. In tabel 13 staan de gemiddelde somscores die behaald zijn op uitspraak.

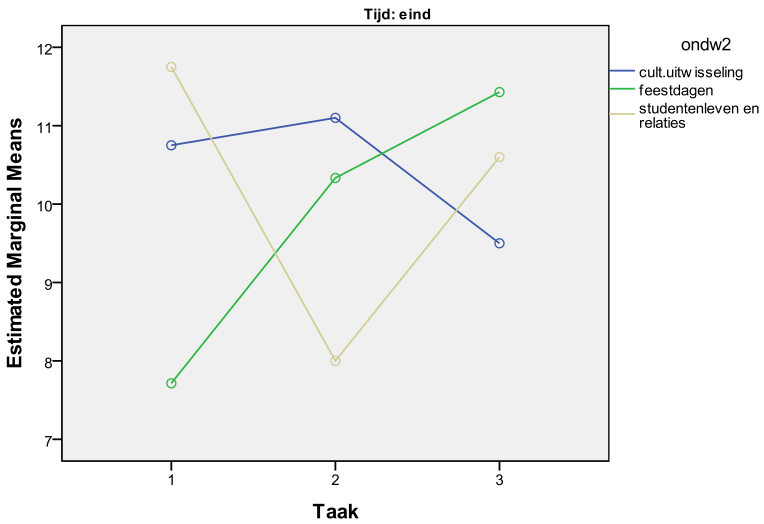
	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Cultuur-uitwisseling	14,33 (2,50)	12,33 (0,58)	10,75 (2,14)	12,00 (3,59)	13,10 (1,52)	11,10 (3,07)	14,88 (1,73)	13,00 (1,85)	9,50 (1,31)
Feestdagen	12,29 (2,06)	9,00 (-)	7,71 (0,76)	12,28 (3,14)	13,28 (1,64)	10,33 (2,52)	14,71 (1,98)	13,00 (3,27)	11,43 (1,72)
Persoonlijk leven	15,05 (1,62)	11,80 (2,17)	11,75 (2,87)	11,56 (3,64)	11,63 (2,00)	8,00 (-)	13,73 (2,19)	13,31 (2,87)	10,60 (2,46)

Tabel 13. De gemiddelde somscores op uitspraak

Een meerweg ANOVA laat zien dat er op het *beginmoment* een hoofdeffect van taak is (F = 8,08; df = 2, 96; p = 0,001). Er is geen hoofdeffect van onderwerp (F = 0,43; df = 2, 96; p = 0,65) en geen interactie-effect (F = 1,82; df = 4, 96; p = 0,13). Uit de Post Hoc toets blijkt dat de scores op taak 1 hoger zijn dan op taak 2 (MD = 2,29; SE = 0,60; p = 0,001). De scores van taak 1 en 3 verschillen niet (MD = 0,05; SE = 0,63; p > 0,99). Er is wel een verschil tussen de scores op taak 2 en 3. Op taak 3 wordt hoger gescoord dan op taak 2 (MD = -2,24; SE = 0,64; p = 0,003).

Op het *middenmoment* laat de ANOVA andere resultaten zien. Hier is namelijk geen hoofdeffect van taak (F = 2,17; df = 2, 64; p = 0,12), geen hoofdeffect van onderwerp (F = 0,67; df = 2, 64; p = 0,52) en geen interactie-effect (F = 1,16; df = 4, 64; = 0,34) te vinden. De scores op het middenmoment zijn dus niet afhankelijk van taaknummer en onderwerp.

Op het *eindmoment* is er een interactie-effect van taak en onderwerp (F = 3,34; df = 4, 68; p = 0,02) te vinden. De score op uitspraak is afhankelijk van zowel het taaknummer als het onderwerp. Er is geen hoofdeffect van taak (F = 0,36; df = 2, 68; p = 0,70) en geen hoofdeffect van onderwerp (F = 0,52; df = 2, 68; p = 0,60). Grafiek 5 laat het interactie-effect zien.



Grafiek 5. Interactie-effect taak en onderwerp op uitspraak

In de grafiek is te zien dat de lijn door elkaar lopen. Bij 'cultuuruitwisseling' is te zien dat de scores op taak 1 en 2 (veel) hoger zijn dan op taak 3. Dit geldt voor 'feestdagen' juist andersom. De scores op 'persoonlijk leven' (studentenleven en relaties) wisselen nog sterker. Op taak 1 en 3 wordt hoger gescoord dan op taak 2. Er is geen duidelijke lijn in de resultaten te vinden. Het is niet zo dat de scores op een bepaald onderwerp altijd hoger zijn dan de scores op andere onderwerpen (er is ook geen hoofdeffect). De grafiek laat duidelijk zien dat het taaknummer een duidelijke invloed heeft op de score, maar dat die invloed per onderwerp verschillend is (bijvoorbeeld bij het verschil in verloop van de scorelijnen van 'cultuuruitwisseling' en 'feestdagen'). Bij taak 1 kan bijvoorbeeld het best gekozen worden voor het onderwerp 'persoonlijk leven', maar bij taak 3 voor 'feestdagen'. Dit wisselt per taaknummer.

4.4.5 Algemene indruk

Het laatste aspect dat geanalyseerd wordt bij het zoeken naar een effect van onderwerp op spreekvaardigheid is de algemene indruk. De gemiddelde somscores op deze categorie staan in tabel 14.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Cultuur-uitwisseling	6,13 (0,68)	5,33 (1,53)	4,58 (1,62)	4.90 (1,73)	5,80 (1,32)	4,90 (1,29)	5,94 (1,45)	5,88 (1,55)	4,88 (0,99)
Feestdagen	5,14 (0,99)	5,00 (-)	3,29 (1,38)	5,42 (1,39)	6,06 (1,16)	4,56 (1,54)	6,07 (0,98)	5,57 (1,62)	5,00 (1,41)
Persoonlijk leven	6,16 (0,93)	5,00 (0,71)	5,88 (0,85)	4,93 (1,64)	4,63 (2,39)	5,00 (-)	5,77 (0,92)	6,39 (1,50)	5,00 (1,41)

Tabel 14. De gemiddelde somscores op algemene indruk

Op het *beginmoment* is er een hoofdeffect van taak (F = 4,93; df = 2, 96; p =0,01). Er is dus een verschil in score op algemene indruk afhankelijk van het taaknummer. Er is geen hoofdeffect van onderwerp (F = 0,06; df = 2, 96; p = 0,94) en geen interactie-effect (F = 1,54; df = 4, 96; p = 0,20). Een Post Hoc toets laat zien dat de scores op taak 1 verschillen van taak 2 (MD = 0,83; SE = 0,28; p = 0,02). Er wordt hoger gescoord op taak 1 dan op taak 2. Op taak 3 wordt ook hoger gescoord dan op taak 2 (MD = 0,75; SE = 0,30; p = 0,05). Er is geen verschil tussen de scores op taak 1 en taak 3 (MD = 0,08; SE = 0,29; p = 0,97).

Op het *middenmoment* laat de meerweg ANOVA geen effecten zien. Er is geen hoofdeffect van taak (F = 1,05; df = 2, 64; p = 0,36), geen hoofdeffect van onderwerp (F = 0,24; df = 2, 64; p = 0,79) en geen interactie-effect (F = 1,51; df = 4, 64; p = 0,21). Op het middenmoment zijn de scores op algemene indruk dus niet afhankelijk van het taaknummer en/of het onderwerp.

Deze resultaten gelden ook voor het *eindmoment*. Er is geen hoofdeffect van taak (F = 0,38; df = 2, 68; p = 0,68), geen hoofdeffect van onderwerp (F = 1,73; df = 2, 68; p = 0,19) en geen interactie-effect (F = 1,41; df = 4, 68; p = 0,24). Ook op het eindmoment zijn de scores dus niet afhankelijk van het taaknummer en/of het onderwerp.

4.4.6 Samenvatting effect van onderwerp

Op het *beginmoment* is bij elk aspect van spreekvaardigheid een effect van taak te vinden. Overal geldt dat de scores op taak 3 hoger zijn dan op taak 2. Bij vier van de vijf aspecten geldt dat de scores op taak 1 hoger zijn dan op taak 2. Dit geldt alleen niet voor complexiteit. Dit resultaat is vreemd, aangezien dit tegen een effect van herhaling pleit. De proefpersonen gaan slechter scoren op spreekvaardigheid bij het uitvoeren van meerdere taken. Maar dit geldt alleen wanneer er slechts twee taken gedaan worden. Bij de derde taak wordt de score op spreekvaardigheid weer hoger. Bij geen enkel aspect van spreekvaardigheid is een effect van onderwerp te vinden. Het onderwerp lijkt geen invloed uit te oefenen op de scores.

Op het *middenmoment* zijn bijna geen effecten te vinden. Alleen op vloeiendheid is er een effect van taak. Dit effect houdt in dat proefpersonen in het midden van het gesprek op taak 3 hoger scoren dan op taak 1. Op taak 2 wordt daarentegen niet hoger gescoord dan op taak 1, dus de proefpersonen gaan het niet steeds beter doen. Maar ze boeken wel een vooruitgang ten opzichte van het startpunt (score op taak 1 vergeleken met de score op taak 3). Maar zoals gezegd geldt dit alleen voor vloeiendheid.

Op het *eindmoment* zijn twee interactie-effecten te vinden, namelijk op complexiteit en uitspraak. Daar wordt aangetoond dat het taaknummer invloed uitoefent op de score, maar dat de mate van invloed per onderwerp verschillend is. De score op een onderwerp is dus afhankelijk van het taaknummer en de score op een taak is dus afhankelijk van het onderwerp. Dit resultaat laten de andere drie aspecten van spreekvaardigheid niet zien.

Alleen op het eindmoment lijkt het onderwerp dus een rol te spelen en dan slechts op twee aspecten van spreekvaardigheid (complexiteit en uitspraak). Geen enkel resultaat laat zien dat scores op een

bepaald onderwerp altijd hoger of lager zijn dan op een ander onderwerp. Wel opvallend is dat de scores op 'persoonlijk leven' een dieptepunt kennen bij taak 2 (zie grafiek 4 en 5). Het lijkt dus niet verstandig om 'persoonlijk leven' als onderwerp te kiezen voor de tweede taak. Daarnaast is het af te raden om 'feestdagen' als onderwerp te kiezen voor taak 1, maar is het aan te raden om dit onderwerp te kiezen voor taak 3. Bij taak 1 worden namelijk de laagste scores behaald op 'feestdagen', maar bij taak 3 worden juist de hoogste scores behaald op dit onderwerp. Het taaknummer hangt dus samen met het onderwerp voor twee aspecten van spreekvaardigheid. Zoals gezegd geldt dit alleen voor het eindmoment.

4.5 Effect van hulpmiddel

Tot slot wordt er, naast een effect van taakherhaling en onderwerp, nog gekeken of het gebruik van hulpmiddelen van invloed is op de spreekvaardigheid. In tabel 15 staat de verdeling van de hulpmiddelen over de taken. Opgemerkt moet worden dat verschillende hulpmiddelen bij elkaar gevoegd zijn. Dit is gedaan, omdat sommige hulpmiddelen weinig voorkwamen (bijvoorbeeld 'onderwerpenlijst' – N = 5) en doordat er op deze manier een gelijke verdeling ontstaat per soort hulpmiddel. Uit de analyses kunnen dan conclusies getrokken worden per soort hulpmiddel. Hierdoor worden de conclusies generaliseerbaar. Er zijn nu vier categorieën: geen hulpmiddel, woordenweb, afbeeldingen (afbeeldingen/foto's, videofilms) en overig (liedje met gatentekst, onderwerpenlijst, agenda, artikel).

	Taak 1	*Taak 2*	*Taak 3*	Totaal
Geen hulpmiddel	22	15	39	76
Woordenweb	28	25	15	68
Afbeeldingen	-	62	9	71
Overig	20	-	20	40
Totaal	70	102	83	255

Tabel 15. De verdeling van de hulpmiddelen over de taken.

De resultaten zullen, net als bij het effect van herhaling en het effect van onderwerp, besproken worden per aspect van spreekvaardigheid.

4.5.1 Complexiteit

Als eerst zal gekeken worden of het gebruik van hulpmiddelen effect heeft op de complexiteit. In tabel 16 staan de gemiddelde somscores op complexiteit per taak, per opnamemoment en per hulpmiddel.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Geen hulpmiddel	8,64 (1,12)	8,00 (1,00)	7,25 (0,89)	8,80 (0,84)	5,80 (2,59)	7,80 (1,10)	9,23 (0,83)	7,54 (2,44)	9,00 (1,73)
Woordenweb	7,12 (1,96)	7,00 (1,00)	6,78 (1,99)	7,50 (1,51)	5,40 (2,07)	7,80 (2,87)	8,80 (0,84)	5,60 (2,30)	9,00 (2,55)
Afbeeldingen	-	-	-	6,82 (2,63)	6,90 (2,49)	8,68 (2,41)	7,33 (2,52)	6,00 (2,00)	9,33 (1,16)
Overig	8,36 (1,43)	7,33 (1,53)	9,67 (2,25)	-	-	-	7,89 (1,52)	6,00 (2,65)	7,75 (1,71)

Tabel 16. De gemiddelde somscores op complexiteit

Een meerweg ANOVA laat zien dat op het *beginmoment* er een hoofdeffect van hulpmiddel is (F = 3,09; df = 3, 95; p = 0,03). De scores op complexiteit zijn dus afhankelijk van de gebruikte hulpmiddelen. Er is geen hoofdeffect van taak (F = 0,93; df = 2, 95; p = 0,40) en geen interactie-effect (F = 0,76; df = 4, 95; p = 0,55). Een Post Hoc (Scheffe) toets laat zien wat het hoofdeffect van hulpmiddel inhoudt. De scores op complexiteit zijn hoger wanneer er geen hulpmiddelen worden gebruikt ten opzichte van een woordenweb (MD = 1,41; SE = 0,47; p = 0,03) en ten opzichte van afbeeldingen (MD = 2,05; SE = 0,50; p = 0,001). Er zijn geen verschillen tussen de scores op 'geen hulpmiddelen' en 'overige hulpmiddelen' (MD = 0,78; SE = 0,53; p = 0,54), 'woordenweb' en 'afbeeldingen' (MD = 0,64; SE = 0,49; p = 0,64), 'woordenweb' en 'overige hulpmiddelen' (MD = -0,63; SE = 0,52; p = 0,69) en 'afbeeldingen' en 'overige hulpmiddelen' (MD = -1,27; SE = 0,54; p = 0,15). Met uitzondering van de categorie 'overige hulpmiddelen' is het verstandig om aan het begin van het gesprek geen hulpmiddelen te gebruiken. Het weglaten van hulpmiddelen heeft immers een positief effect op de score op complexiteit.

Op het *middenmoment* laat de ANOVA geen effecten zien. Er is geen hoofdeffect van hulpmiddel (F = 0,65; df = 3, 63; p = 0,58), geen hoofdeffect van taak (F = 1,34; df = 2, 63; p = 0,27) en geen interactie-effect (F = 0,53; df = 4, 63; p = 0,72). In het midden van het gesprek zijn de scores op complexiteit dus niet afhankelijk van het gebruik van hulpmiddelen.

Ditzelfde geldt voor de resultaten op het *eindmoment.* Ook daar is geen hoofdeffect van hulpmiddel (F = 0,78; df = 3, 67; p = 0,51), geen hoofdeffect van taak (F = 0,71; df = 2, 67; p = 0,50) en geen interactie-effect (F = 1,72; df = 4, 67; p = 0,16) te meten.

4.5.2 Nauwkeurigheid

Een tweede aspect van spreekvaardigheid is nauwkeurigheid. In tabel 17 staan de gemiddelde somscores voor deze categorie.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Geen hulpmiddel	15,73 (1,01)	10,67 (2,52)	10,62 (2,07)	14,60 (1,14)	13,60 (2,30)	10,20 (1,64)	14,46 (1,51)	13,62 (3,04)	10,38 (1,26)
Woordenweb	13,62 (1,96)	9,33 (1,53)	9,44 (2,24)	13,80 (2,20)	11,20 (2,35)	9,20 (1,30)	14,60 (2,30)	10,20 (2,05)	10,80 (1,64)
Afbeeldingen	-	-	-	10,95 (3,14)	11,90 (2,64)	10,21 (2,88)	13,00 (2,65)	11,00 (3,00)	8,67 (1,53)
Overig	13,64 (2,01)	10,00 (1,00)	10,00 (1,90)	-	-	-	13,22 (2,44)	10,71 (2,63)	9,00 (0,82)

Tabel 17. De gemiddelde somscores op nauwkeurigheid

De ANOVA toont aan dat er op het *beginmoment* wederom een hoofdeffect van hulpmiddel is (F = 5,04; df = 3, 95; p = 0,003). Er is geen hoofdeffect van taak (F = 0,92; df = 2, 95; p = 0,40) en geen interactie-effect (F = 1,05; df = 4, 95; p = 0,38). De Post Hoc toets laat zien dat er een verschil is tussen 'geen hulpmiddel' en 'afbeeldingen' (MD = 3,77; SE = 0,61; p < 0,001), 'woordenweb' en 'afbeeldingen' (MD = 2,64; SE = 0,60; p = 0,001) en 'overige hulpmiddelen' en 'afbeeldingen' (MD = 2,25; SE = 0,67; p = 0,01). In alle gevallen wordt hoger gescoord op nauwkeurigheid als er geen afbeeldingen gebruikt worden, maar gekozen wordt voor geen of andere hulpmiddelen. Er is geen verschil gevonden tussen de scores bij 'geen hulpmiddel' en 'woordenweb' (MD = 1,13; SE = 0,58; p = 0,29), 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 1,52; SE = 0,65; p = 0,15) en 'woordenweb' en 'overige hulpmiddelen' (MD = 0,39; SE = 0,64; p = 0,95).

Ook op het *middenmoment* wordt er alleen een hoofdeffect van hulpmiddel (F = 2,33; df = 3, 63; p = 0,08) gemeten. De p-waarde mag door twee gedeeld worden i.v.m. eenzijdige toetsing. Er is geen hoofdeffect van taak (F = 2,02; df = 2, 63; p = 0,14) en geen interactie-effect (F = 0,30; df = 4, 63; p = 0,88). Het hoofdeffect van hulpmiddel houdt in dat er een verschil is op de scores bij 'geen hulpmiddel' en 'woordenweb' (MD = 2,58; SE = 0,83; p = 0,03) en bij 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 2,69; SE = 0,99; p = 0,07). Deze p-waarde mag door twee gedeeld worden. Er wordt hoger gescoord wanneer er geen hulpmiddel gebruikt wordt ten opzichte van een woordenweb en overige hulpmiddelen. Er wordt geen verschil gevonden tussen de scores op 'geen hulpmiddel' en 'afbeeldingen' (MD = 1,40; SE = 0,77; p = 0,36), 'woordenweb' en 'afbeeldingen' (MD = -1,18; SE = 0,80; p = 0,54), 'woordenweb' en 'overige hulpmiddelen' (MD = 0,11; SE = 1,02; p > 0,99) en 'afbeeldingen' en 'overige hulpmiddelen' (MD = 1,29; SE = 0,97; p = 0,62).

Op het *eindmoment* is er geen hoofdeffect van hulpmiddel (F = 0,75; df = 3, 67; p = 0,53), geen hoofdeffect van taak (F = 0,05; df = 2, 67; p = 0,95) en geen interactie-effect (F = 0,97; df = 4, 67; p = 0,43) te meten. Aan het eind van het gesprek beïnvloedt het gebruik van eventuele hulpmiddelen de prestaties op nauwkeurigheid dus niet.

4.5.3 Vloeiendheid

Het derde aspect van spreekvaardigheid is vloeiendheid. In onderstaande tabel (tabel 18) staan de gemiddelde somscores op vloeiendheid.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Geen hulpmiddel	14,00 (3,16)	9,33 (3,51)	8,13 (1,96)	12,80 (1,30)	13,20 (3,27)	9,00 (1,00)	15,00 (1,87)	15,39 (2,81)	11,31 (2,59)
Woordenweb	11,50 (3,69)	10,67 (0,58)	8,89 (3,41)	11,90 (3,35)	11,70 (4,47)	8,20 (5,02)	13,00 (3,32)	12,40 (5,22)	9,80 (1,79)
Afbeeldingen	-	-	-	10,18 (3,85)	12,10 (3,90)	10,79 (3,29)	8,33 (3,52)	11,33 (4,04)	8,00 (1,00)
Overig	14,55 (2,58)	8,33 (1,15)	9,67 (2,58)	-	-	-	11,89 (2,80)	11,43 (3,60)	9,25 (1,89)

Tabel 18. De gemiddelde somscores op vloeiendheid

Een meerweg ANOVA wijst uit dat er op het *beginmoment* geen hoofdeffect van taak (F = 0,16; df = 2, 95; p = 0,85) en geen interactie-effect van taak en hulpmiddel is (F = 1,73; df = 4, 95; p = 0,15). Er is wel een hoofdeffect van hulpmiddel (F = 4,64; df = 3, 95; p = 0,01). Uit een Post Hoc toets blijkt dat de scores op vloeiendheid verschillen bij 'geen hulpmiddel' en 'woordenweb' (MD = 2,37; SE = 0,82; p = 0,05), 'geen hulpmiddel' en 'afbeeldingen' (MD = 4,28; SE = 0,87; p < 0,001) en 'afbeeldingen' en 'overige hulpmiddelen' (MD = -3,39; SE = 0,96; p = 0,01). Er wordt hoger gescoord op vloeiendheid wanneer er aan het begin van het gesprek geen hulpmiddelen gebruikt worden in vergelijking met het gebruik van een woordenweb of afbeeldingen. Daarnaast wordt er hoger gescoord wanneer 'overige hulpmiddelen' gebruikt worden in plaats van 'afbeeldingen'. Er is geen verschil tussen de scores op 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 0,89; SE = 0,93; p = 0,82), 'woordenweb' en 'afbeeldingen' (MD = 1,91; SE = 0,86; p = 0,18) en 'woordweb' en 'overige hulpmiddelen' (MD = -1,48; SE = 0,91; p = 0,46).

Op het *middenmoment* is er geen hoofdeffect van hulpmiddel (F < 1; df = 3, 63; p = 0,40), maar wel een hoofdeffect van taak (F = 2,71; df = 2, 63; p = 0,07). Deze p-waarde mag door twee gedeeld worden (eenzijdige toetsing). Er is geen interactie-effect (F = 0,59; df = 4, 63; p = 0,67). Een Post Hoc toets laat zien dat de score op taak 3 hoger is dan op taak 1 (MD = 1,29; SE = 0,94; p = 0,03). De scores op taak 1 en taak 2 verschillen niet (MD = -2,69; SE = 1,38; p = 0,16), net zoals de scores op taak 2 en 3 (MD = -1,29; SE = 0,94; p = 0,39).

Op het *eindmoment* is er geen hoofdeffect van hulpmiddel (F = 0,13; df = 3, 67; p = 0,94), geen effect van taak (F = 0,33; df = 2, 67, p = 0,72) en geen interactie-effect (F = 1,93; df = 4, 67; p = 0,12). Het gebruik van hulpmiddel heeft aan het eind van het gesprek dus geen invloed op de vloeiendheid.

4.5.4 Uitspraak

Het vierde aspect van spreekvaardigheid waarbij gezocht wordt naar een effect van hulpmiddelen is uitspraak. De gemiddelde somscores op uitspraak staan in tabel 19.

	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Geen hulpmiddel	15,55 (1,29)	13,00 (2,00)	10,50 (2,51)	14,60 (1,67)	13,60 (1,67)	9,20 (1,30)	15,15 (1,41)	14,31 (1,97)	10,92 (2,33)
Woordenweb	13,69 (2,12)	9,67 (0,58)	8,33 (1,41)	12,60 (2,55)	12,30 (1,95)	9,40 (2,70)	14,40 (2,30)	12,80 (3,96)	11,40 (1,14)
Afbeeldingen	-	-	-	11,18 (3,60)	12,95 (1,72)	11,16 (2,83)	13,67 (1,53)	12,67 (0,58)	9,67 (1,16)
Overig	14,00 (2,68)	12,33 (0,58)	11,83 (2,32)	-	-	-	13,11 (2,42)	11,43 (2,51)	8,50 (0,58)

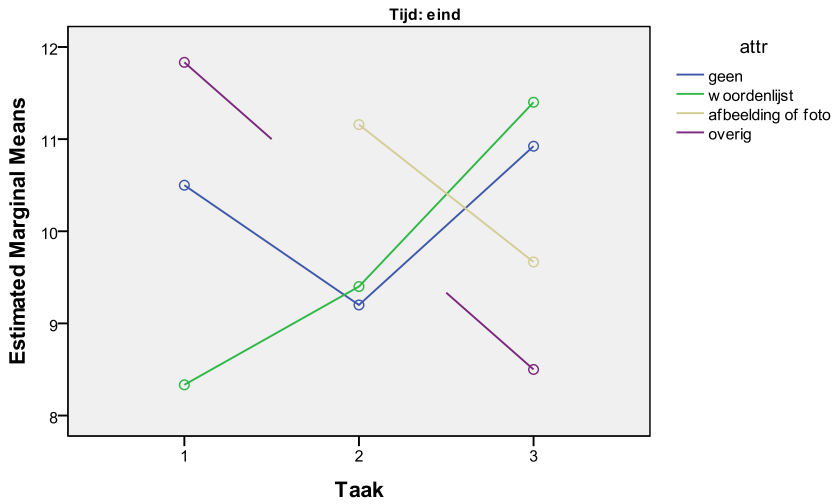
Tabel 19. De gemiddelde somscores op uitspraak

De meerweg ANOVA laat op het *beginmoment* een hoofdeffect van hulpmiddel zien (F = 3,91; df = 3, 95; p = 0,01). Er is geen hoofdeffect van taak (F = 2,18; df = 2, 95; p = 0,12) en geen interactie-effect van taak en hulpmiddel (F = 0,46; df = 4, 95; p = 0,77). Er zijn verschillende hulpmiddelen die invloed uitoefenen op de score voor vloeiendheid. Dit laat de Post Hoc toets zien. Er wordt hoger gescoord bij het achterwege laten van hulpmiddel ('geen hulpmiddel') ten opzichte van 'afbeeldingen' (MD = 3,73; SE = 0,68; p < 0,001) en 'woordenweb' (MD = 1,76; SE = 0,65; p = 0,07). Hierbij mag de p-waarde door twee gedeeld worden i.v.m. eenzijdige toetsing. Er wordt ook hoger gescoord bij het gebruik van een 'woordenweb' ten opzichte van 'afbeeldingen' (MD = 1,97; SE = 0,67; p = 0,04) en bij 'overige hulpmiddelen' ten opzichte van 'afbeeldingen' (MD = 2,12; SE = 0,75; p = 0,05). Er is geen verschil tussen 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 1,61; SE = 0,73; p = 0,19) en tussen 'woordenweb' en 'overige hulpmiddelen' (MD = -0,15; SE = 0,72; p > 0,99). Het gebruik van afbeeldingen aan het begin van het gesprek wordt dus eigenlijk afgeraden, omdat er hogere scores gehaald worden op 'geen hulpmiddel', 'woordenweb' en 'overige hulpmiddelen' ten opzichte van 'afbeeldingen'.

Op het *middenmoment* is er wederom een hoofdeffect van hulpmiddel (F = 2,64; df = 3, 63; p = 0,06). Deze p-waarde mag weer door twee gedeeld worden (eenzijdig toetsen). Er is geen hoofdeffect van taak (F = 1,11; df = 2, 63; p = 0,34) en geen interactie-effect (F = 1,11; df = 4, 63; p = 0,36). Het effect van hulpmiddel komt alleen terug bij een verschil in scores tussen 'geen hulpmiddel' en 'woordenweb (MD = 1,95; SE = 0,65; p = 0,04) en 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 2,25; SE = 0,78; p = 0,05). Bij het ontbreken van hulpmiddelen is de score op uitspraak hoger. Het lijkt dus een positieve uitwerking op de uitspraak te hebben om in het midden van het gesprek geen hulpmiddelen te gebruiken. Dit geldt alleen niet ten opzichte van 'afbeeldingen'. Er is namelijk geen verschil op de scores op uitspraak tussen 'geen hulpmiddel' enerzijds en 'afbeeldingen' anderzijds (MD = 1,04; SE = 0,61; p = 0,42). Er is ook geen verschil tussen de scores op 'woordenweb' en 'afbeeldingen' (MD = -0,92; SE = 0,64; p = 0,56), 'woordenweb' en 'overige hulpmiddelen' (MD =

0,30; SE = 0,80; p = 0,99) en ‘afbeeldingen’ en ‘overige hulpmiddelen’ (MD = 1,22; SE = 0,64; p = 0,48).

Aan het *eind* van het gesprek is er geen hoofdeffect van hulpmiddelen (F = 0,29; df = 3, 67; p = 0,83), geen hoofdeffect van taak (F = 0,16; df = 2, 67; p = 0,85), maar wel een interactie-effect van hulpmiddel en taak op de uitspraak (F = 3,55; df = 4, 67; p = 0,01). Dit is te zien in grafiek 6.



Grafiek 6. Interactie-effect taak en hulpmiddel op uitspraak

In de grafiek lopen de scorelijnen van de verschillende hulpmiddelen door elkaar. Het zijn geen rechte lijnen. Het ene hulpmiddel scoort niet standaard hoger dan het andere hulpmiddel (er is ook geen hoofdeffect op hulpmiddel). Het is ook niet zo dat er op taak 3 standaard hoger gescoord wordt dan op taak 1 of 2 (er is geen hoofdeffect op taak). Bij ‘overige hulpmiddelen’ is te zien dat de score bij taak 1 heel hoog is, maar bij taak 3 heel laag. Bij ‘woordenlijst’ (woordenweb) is dit precies andersom. Het verschilt dus per hulpmiddel welke invloed deze uitoefent op de score voor uitspraak. Dat verschilt ook per taaknummer. Er is dus een duidelijke interactie tussen taak en hulpmiddel. Bij taak 1 is het verstandig om hulpmiddelen uit de categorie ‘overig’ te gebruiken, aangezien hier de hoogste score op gehaald wordt. Bij taak 2 is het juist verstandig om afbeeldingen te gebruiken en bij taak 3 een woordenweb of geen hulpmiddel. Het verschilt dus per taak welk hulpmiddel de hoogste scores op uitspraak oplevert.

4.5.5 Algemene indruk

Het laatste aspect van spreekvaardigheid dat besproken wordt is algemene indruk. In tabel 20 staan de gemiddelde somscores voor algemene indruk.

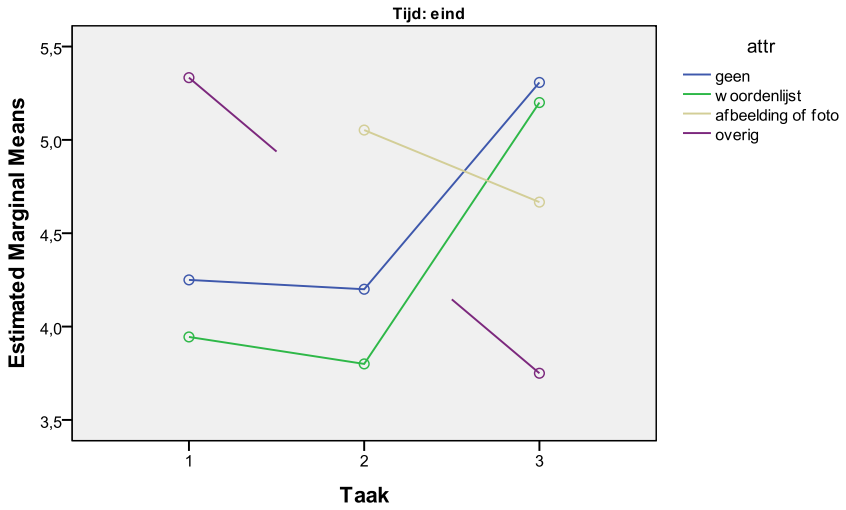
	Taak 1			*Taak 2*			*Taak 3*		
	Begin	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*	*Begin*	*Midden*	*Eind*
Geen hulpmiddel	6,36 (0,45)	4,67 (0,58)	4,25 (1,17)	6,20 (0,45)	6,00 (1,22)	4,20 (0,45)	6,23 (0,60)	6,77 (1,30)	5,31 (1,11)
Woordenweb	5,47 (1,10)	5,33 (0,58)	3,94 (1,81)	5,60 (0,81)	5,10 (2,08)	3,80 (1,48)	6,00 (1,17)	5,40 (1,95)	5,20 (1,64)
Afbeeldingen	-	-	-	4,68 (1,76)	5,86 (1,42)	5,05 (1,47)	5,00 (2,00)	5,33 (1,15)	4,67 (0,58)
Overig	6,27 (0,72)	5,33 (1,53)	5,33 (1,86)	-	-	-	5,61 (1,14)	5,43 (1,40)	3,75 (0,96)

Tabel 20. De gemiddelde somscores op algemene indruk

Op het *beginmoment* laat de ANOVA een hoofdeffect van hulpmiddel zien (F = 3,38; df = 3, 95; p = 0,02). Er is geen hoofdeffect van taak (F = 0,17; df = 2, 95; p = 0,84) en geen interactie-effect van taak en hulpmiddel (F = 0,64; df = 4, 95; p = 0,63). De Post Hoc toets laat een aantal verschillen zien in de scores op algemene indruk. Allereerst wordt er hoger gescoord op algemene indruk wanneer er 'geen hulpmiddelen' gebruikt zijn ten opzichte van 'afbeeldingen' (MD = 1,56; SE = 0,31; p < 0,001). Er wordt ook hoger gescoord wanneer een 'woordenweb' gebruikt wordt ten opzichte van 'afbeeldingen' (MD = 0,88; SE = 0,31; p = 0,05). En tot slot wordt er ook hoger gescoord op algemene indruk wanneer 'overige hulpmiddelen' gebruikt worden ten opzichte van 'afbeeldingen' (MD = 1,26; SE = 0,34; p = 0,01). Wanneer afbeeldingen gebruikt worden, dan wordt er lager gescoord op algemene indruk. Het gebruik van afbeeldingen aan het begin van het gesprek kun je dus beter vermijden. De andere scores verschillen niet van elkaar: 'geen hulpmiddel' en 'woordenweb' (MD = 0,68; SE = 0,30; p = 0,16), 'geen hulpmiddel' en 'overige hulpmiddelen' (MD = 0,30; SE = 0,33; p = 0,85) en 'woordenweb' en 'overige hulpmiddelen' (MD = -0,38; SE = 0,33; p = 0,73).

In het *midden* van het gesprek oefenen hulpmiddelen geen invloed uit op de scores op algemene indruk. Er is namelijk geen hoofdeffect van hulpmiddel (F = 0,34; df = 3, 63; p = 0,80), geen hoofdeffect van taak (F = 0,62; df = 2, 63; p = 0,54) en geen interactie-effect (F = 0,95; df = 4, 63; p = 0,44). In het midden van het gesprek maakt het dus niet uit welke hulpmiddelen je gebruikt.

Aan het *eind* van het gesprek meet de ANOVA geen hoofdeffect van hulpmiddel (F = 0,35; df = 3, 67; p = 0,79), geen hoofdeffect van taak (F = 0,41; df = 2, 67; p = 0,66), maar wel een interactie-effect van hulpmiddel en taak (F = 2,17; df = 4, 67; p = 0,08). Deze p-waarde mag weer gehalveerd worden doordat de resultaten eenzijdig getoetst worden. Dit interactie-effect wordt geïllustreerd met behulp van grafiek 7.



Grafiek 7. Interactie-effect taak en hulpmiddel op algemene indruk

In de grafiek staan zowel dalende als stijgende lijnen. De scores waarbij het hulpmiddel 'woordenlijst' (woordenweb) gebruikt is zijn bij taak 1 en 2 nog vrij laag, maar stijgen bij taak 3 enorm. De scores bij 'overige hulpmiddelen' daarentegen dalen van hoog (bij taak 1) naar laag (bij taak 3). Het verschilt dus per hulpmiddel welke invloed deze uitoefent op de score voor uitspraak. Dat verschilt ook per taaknummer. Dit was ook te zien in paragraaf 4.5.4 bij 'uitspraak'. Er is wederom een duidelijke interactie tussen taak en hulpmiddel. Bij taak 1 worden de hoogste scores behaald met behulp van 'overige hulpmiddelen'. Bij taak 2 worden de hoogste scores behaald wanneer afbeeldingen gebruikt worden en bij taak 3 worden de hoogste scores behaald wanneer er een woordenweb of juist geen hulpmiddel gebruikt wordt. Het verschilt dus per taak welk hulpmiddel de hoogste scores op uitspraak oplevert.

4.5.6 Samenvatting effect van hulpmiddel

Op het *beginmoment* is er bij elk aspect van spreekvaardigheid een effect van hulpmiddel te vinden. Bij vier van de vijf aspecten houdt dat effect in dat de scores hoger zijn wanneer het gebruik van afbeeldingen vermeden wordt. Dit geldt alleen niet voor complexiteit. Bij complexiteit is de score bij het ontbreken van hulpmiddelen hoger dan bij het gebruik van een woordenweb of afbeeldingen. Ook bij vloeiendheid en uitspraak is de score bij het ontbreken van hulpmiddelen hoger dan bij het gebruik van een woordenweb. Deze resultaten laten zien dat het verstandig kan zijn om aan het begin van een gesprek geen hulpmiddelen te gebruiken, omdat er op die manier hogere scores worden gehaald dan bij het gebruik van een woordweb of afbeeldingen. Wanneer er wel gekozen wordt voor een hulpmiddel, dan moet het gebruik van afbeeldingen vermeden worden, omdat er met behulp van de andere hulpmiddelen (woordenweb + overig) hogere scores op spreekvaardigheid behaald worden.

Voor het *middenmoment* geldt dat er op complexiteit en algemene indruk geen effecten behaald worden. Bij nauwkeurigheid en uitspraak is er een effect van hulpmiddel. Dit houdt in dat er hogere scores behaald worden wanneer er geen hulpmiddel gebruikt wordt ten opzichte van een woordenweb en overige hulpmiddelen. In het midden van het gesprek is er geen verschil met de

scores die behaald worden wanneer er gebruikt gemaakt wordt van afbeeldingen ten opzichte van andere hulpmiddelen. Er is alleen een positief effect van 'geen hulpmiddel' op nauwkeurigheid en uitspraak. Het lijkt verstandig om ook in het midden van het gesprek geen hulpmiddel te gebruiken, omdat er op twee van de vijf categorieën een hogere score behaald wordt wanneer er geen hulpmiddel bij de taakuitvoering gebruikt worden. Bij vloeiendheid is er een effect van taak. Op taak 3 wordt hoger gescoord op vloeiendheid dan op taak 1. De proefpersonen gaan op vloeiendheid dus vooruit, onafhankelijk van het gebruik van hulpmiddelen.

Voor het *eindmoment* geldt dat er geen hoofdeffecten te vinden zijn. Voor uitspraak en algemene indruk is er wel een interactie-effect van taak en hulpmiddel. Dit houdt in dat de combinatie van een taaknummer en een hulpmiddel samen de score op spreekvaardigheid beïnvloeden. Voor beide aspecten (uitspraak en algemene indruk) geldt dat op taak 1 het best gekozen kan worden voor hulpmiddelen uit de categorie 'overig'. Bij taak 2 daarentegen is het verstandiger om voor afbeeldingen te kiezen. Dit hulpmiddel geeft op taak 2 immers de hoogste score. Voor taak 3 ligt het gebruik van een woordenweb of geen hulpmiddel het meest voor de hand. Daarop wordt bij taak 3 immers de hoogste score behaald. Het gebruik van hulpmiddelen of juist het ontbreken ervan beïnvloedt de prestatie op complexiteit, nauwkeurigheid en vloeiendheid niet.

5. Conclusie

Binnen dit onderzoek is gezocht naar het antwoord op de vraag: *wat is het effect van het herhaaldelijk uitvoeren van taken op de spreekvaardigheid van tweedetaalverwervers en welke rol spelen het onderwerp van de taak en mogelijke hulpmiddelen hierbij?* In de resultatensectie is gebleken dat het herhaaldelijk uitvoeren van taken pas invloed heeft op de spreekvaardigheid van tweedetaalverwervers wanneer er meer dan twee taken uitgevoerd worden. De proefpersonen zijn niet beter gaan scoren op taak 2 ten opzichte van taak 1, maar wel op taak 3 ten opzichte van taak 2. Dit geldt niet voor nauwkeurigheid. Daar wordt geen verschil gevonden in de scores. Er geldt dus niet dat een spreker steeds beter gaat scoren op spreekvaardigheid naarmate hij meerdere taken uitvoert. Een effect van herhaling komt pas voor bij de uitvoering van drie taken en alleen op complexiteit, vloeiendheid, uitspraak en algemene indruk. Deze resultaten sluiten niet aan bij de uitkomsten van de onderzoeken van Yule, Powers en McDonald (1992), Gass, Mackey, Fernandez en Alvarez-Torres (1999) en Lynch en McLean (2000). Binnen die onderzoeken is namelijk gebleken dat de proefpersonen steeds beter gingen scoren op spreekvaardigheid. Hier geldt dat pas na twee taken.

Naast een effect van herhaling is gezocht naar een effect van het onderwerp van een taak op de spreekvaardigheid. Op het *beginmoment* zijn er alleen effecten van taak te vinden. Deze effecten houden in dat de proefpersonen hoger scoren op taak 3 vergeleken met taak 2. Dit sluit aan bij wat zojuist vermeld is. Daarnaast is er een verschil in scores op taak 1 en 2. Dit verschil is niet gevonden bij het effect van herhaling. Dit verschil houdt in dat de scores op taak 1 hoger zijn dan op taak 2. De proefpersonen gaan dus slechter presteren naarmate zij meer taken uitvoeren. Dit resultaat geldt alleen wanneer de taakherhaling tot twee taken beperkt wordt. Bij taak 3 wordt de score immers weer hoger dan op taak 2. Het onderwerp heeft aan het begin van het gesprek geen invloed op de spreekvaardigheid. Ditzelfde geldt voor de scores op het *middenmoment*. Alleen op vloeiendheid is

er een effect van taak, waarbij proefpersonen op taak 3 hoger scoren dan op taak 1. Op taak 2 wordt niet hoger gescoord dan op taak 1, dus de proefpersonen gaan het niet steeds beter doen, maar ze boeken wel een vooruitgang over de drie taken. De score op taak 3 voor vloeiendheid is immers hoger dan op taak 1. Alleen bij de scores op het *eindmoment* is er een interactie-effect van onderwerp en taak te vinden. Dit geldt voor complexiteit en uitspraak. Voor deze aspecten geldt dat het taaknummer en het onderwerp samen de score op spreekvaardigheid beïnvloeden. Per taak wordt er op een ander onderwerp de hoogste score behaald. Er is dus geen effect van onderwerp, wat inhoudt dat er niet standaard hoger gescoord wordt op bijvoorbeeld 'persoonlijk leven' ten opzichte van 'culturele uitwisseling'. In dit onderzoek is dus niet eenzelfde resultaat gevonden als in de onderzoeken van Gass en Varonis (1984) en Lange (2000). Aan het eind van het gesprek is het wel van belang welk onderwerp er gekozen wordt. Daarbij moet rekening gehouden worden met het taaknummer. Bij taak 1 kan het best gekozen worden voor 'persoonlijk leven', maar bij taak 2 is het verstandiger om te kiezen voor 'culturele uitwisseling'. Het is dus belangrijk om na te denken over een onderwerp bij het ontwerpen van taken. Maar het is niet zo dat de scores op spreekvaardigheid voor bepaalde onderwerpen hoger zijn dan voor andere onderwerpen. Dit is afhankelijk van het taaknummer. Het is dus belangrijk om bij de keuze van het onderwerp rekening te houden met het taaknummer om de spreekprestatie zo optimaal mogelijk te maken.

Naast een effect van herhaling en onderwerp is ook gezocht naar een effect van hulpmiddel. Voor de scores op het *beginmoment* geldt dat er op alle aspecten van spreekvaardigheid een effect van hulpmiddel gevonden is. Het hulpmiddel heeft dus invloed op de spreekvaardigheid. Er moet dus kritisch overwogen worden welk hulpmiddel gekozen wordt bij de uitvoering van gesprekstaken. Aan het begin van het gesprek kan het gebruik van afbeeldingen bijvoorbeeld beter vermeden worden. Het is zelfs beter om helemaal geen hulpmiddel te gebruiken, want bij deze categorie wordt op drie aspecten van spreekvaardigheid (complexiteit, vloeiendheid en uitspraak) een hogere score behaald dan bij het gebruik van een woordenweb of afbeeldingen. Bij de scores op het *middenmoment* is eenzelfde effect van hulpmiddel te zien. Dit geldt alleen voor nauwkeurigheid en uitspraak. Er worden hogere scores behaald wanneer er geen hulpmiddel gebruikt wordt ten opzichte van een woordenweb en overige hulpmiddelen. Dit toont aan dat het ook in het midden van het gesprek verstandig kan zijn om geen hulpmiddelen te gebruiken, omdat de scores op nauwkeurigheid en uitspraak dan hoger zijn dan bij het gebruik van andere hulpmiddelen (woordenweb en overig). Net als bij het effect van onderwerp is er bij het aspect vloeiendheid een effect van taak. De proefpersonen scores op taak 3 hoger dan op taak 1. Op taak 2 wordt niet hoger gescoord dan op taak 1, dus de proefpersonen gaan het niet steeds beter doen, maar ze boeken wel een vooruitgang over de drie taken. De score op taak 3 voor vloeiendheid is hoger dan op taak 1. De proefpersonen gaan er op vloeiendheid dus vooruit, onafhankelijk van het onderwerp en de hulpmiddelen. Bij de scores op het *eindmoment* zijn twee interactie-effecten te vinden, namelijk op uitspraak en algemene indruk. Op deze aspecten zorgt een combinatie van taaknummer en hulpmiddel voor een invloed op de score. Op taak 1 kan het best gekozen worden voor een hulpmiddel uit 'overige hulpmiddel'. Hier wordt namelijk de hoogste score op behaald. Voor taak 2 wordt de hoogste score behaald wanneer er afbeeldingen gebruikt worden en voor taak 3 wanneer er een woordenweb of helemaal geen hulpmiddel gebruikt wordt. Het is dus belangrijk om bij het gebruik van hulpmiddelen rekening te houden met het taaknummer om een zo optimaal mogelijke spreekprestatie te bereiken.

De resultaten zijn dus niet eenduidig. Het effect van herhaling geldt pas na de uitvoering van twee taken en alleen op complexiteit, vloeiendheid, uitspraak en algemene indruk. De proefpersonen

worden dus niet steeds beter naarmate zij meerdere taken uitvoeren. Zij hebben pas profijt van hun ervaring met het uitvoeren van taken wanneer zij al meerdere taken (minstens 2) uitgevoerd hebben. Het uitvoeren van gesprekstaken heeft dus alleen nut als er minstens drie taken uitgevoerd worden. Daarnaast zijn er geen hoofdeffecten van onderwerp gevonden, alleen twee interactie-effecten op het eindmoment. Het is dus niet het geval dat de scores op spreekvaardigheid bij bepaalde onderwerpen hoger zijn dan bij andere onderwerpen, maar dat dit samenhangt met het taaknummer. Elk taaknummer vraagt om een ander soort onderwerp om een zo hoog mogelijke score op spreekvaardigheid te halen. Dit is ook te zien bij het gebruik van hulpmiddelen, al zijn daar ook hoofdeffecten van hulpmiddelen te vinden. Op sommige aspecten van spreekvaardigheid wordt het hoogst gescoord wanneer er geen hulpmiddel gebruikt wordt en het laagst wanneer er afbeeldingen gebruikt worden. Het is dus belangrijk om een bewuste selectie te maken van te gebruiken hulpmiddelen.

Uit de analyse van de scores blijkt dat bij alle aspecten van spreekvaardigheid te zien is dat de verschillende opnamemomenten zorgen voor verschillen in scores. Het is dus van belang om goed na te denken over de selectie van fragmenten voor de beoordeling van spreekvaardigheid. Bij kortere gesprekstaken is het mogelijk om een oordeel te vellen op basis van het gehele gesprek, maar bij deze gesprekken was dit niet mogelijk in verband met de duur (25 minuten). Uit de resultaten blijkt dat op vier van de vijf aspecten van spreekvaardigheid de scores aan het begin en in het midden van het gesprek hoger zijn dan aan het eind van het gesprek. Dit geldt alleen niet voor complexiteit. Daar zijn de begin- en eindscores hoger dan de middenscores. De scores worden dus (deels) bepaald door het moment in het gesprek. Bij de beoordeling van gesprekken moet dus een bewuste keuze gemaakt worden voor het moment van beoordelen.

Daarnaast is in hoofdstuk 4.1 gebleken dat de oordelen van de verschillende beoordelaars niet op alle vragen betrouwbaar genoeg samenhangen. Dit zorgt ervoor dat de resultaten uit dit onderzoek niet zomaar aangenomen kunnen worden, zeker omdat er in hoofdstuk 4 alleen verder gewerkt is met de oordelen van beoordelaar 1. Er ligt dus een vraag naar vervolgonderzoek. Maar dit onderzoek geeft wel een signaal af. Er is geen direct effect van herhaling bij het uitvoeren van gesprekstaken, maar op de lange termijn wel. Het herhaaldelijk uitvoeren van gesprekstaken (minstens 3) met moedertaalsprekers heeft een positief effect op de spreekvaardigheid (complexiteit, vloeiendheid, uitspraak, algemene indruk) van tweedetaalverwervers.

6. Discussie

De gevonden resultaten laten zien dat het geen effect heeft om een gesprekstaak twee keer uit te voeren. Bij de tweede uitvoering treedt er namelijk geen effect op met betrekking tot spreekvaardigheid. De spreekvaardigheid van proefpersonen verbetert pas bij de uitvoering van de derde taak. Voor het onderwijs is dit een belangrijk resultaat. De herhaling van gesprekstaken heeft alleen nut als de taak vaker dan één keer herhaald wordt. Met betrekking tot het onderwerp is het verstandig om vooraf een pre-test af te nemen om te kijken hoe de proefpersonen reageren op het onderwerp, aangezien de scores op de onderwerpen per taaknummer nogal wisselend zijn. Er zijn geen hoofdeffecten van onderwerp gevonden, maar wel interactie-effecten. Bij de analyse van het

effect van hulpmiddelen is wel duidelijk geworden dat het gebruik van hulpmiddelen niet direct betekent dat de middelen ook daadwerkelijk hulp bieden aan de proefpersonen. Bij verschillende aspecten van spreekvaardigheid ligt de score hoger wanneer er geen hulpmiddel gebruikt wordt ten opzichte van het gebruik van woordenwebben en afbeeldingen. Over de toevoeging van hulpmiddelen moet dus bewust nagedacht worden, omdat hulpmiddelen het taalgebruik niet altijd ten goede komen.

Door de lage interbeoordelaarsbetrouwbaarheid is het goed om een dergelijk onderzoek te herhalen om te zien of dezelfde resultaten gevonden kunnen worden. Ook is goed om een dergelijk onderzoek op te zetten met een controlegroep. Hierbij kan gedacht worden aan een experimentele groep die gesprekstaken uitvoert met moedertaalsprekers, een experimentele groep die gesprekstaken uitvoert met mede-tweedetaalverwervers en een controlegroep die geen gesprekstaken uitvoert. Voorafgaand en na afloop van de gesprekken krijgen alle sprekers dezelfde spreekvaardigheidstest, zodat gemeten kan worden of en hoeveel de sprekers vooruit zijn gegaan op spreekvaardigheid. In het huidige onderzoek was er geen controlegroep, dus het valt niet met zekerheid te zeggen dat externe factoren geheel uitgesloten waren. Het is ook niet bekend welke andere spreekoefeningen de sprekers naast de gesprekstaken hebben uitgevoerd, dus het valt niet met zekerheid te zeggen dat de gevonden resultaten te wijten zijn aan de drie uitgevoerde gesprekstaken. Dit was ook niet te ondervangen, omdat het materiaal (de video-opnames) er al lag voordat dit onderzoek ontving. Het onderzoek is als het ware rond het materiaal gevormd. Hierdoor was het niet mogelijk om rekening te houden met bijvoorbeeld mogelijke gender-effecten. Plough en Gass (1984) doen dit wel door vaste man-vrouw paren te creëren, maar dat was in dit geval niet mogelijk.

Een andere kanttekening bij dit onderzoek is dat de T1'ers niet allemaal moedertaalsprekers zijn. Er zijn twee studenten bij (Justyna en Kathy) die de educatieve master van de Universiteit Utrecht volgen en daarom deel nemen aan dit project, maar hun moedertaal is niet Nederlands. Ze zijn wel native-like en daarom zijn ze meegenomen in het onderzoek, maar het is best aannemelijk dat zij de spreekprestaties van de T2'ers (enigszins) beïnvloed hebben. Soms hebben Justyna en Kathy woordvindingsproblemen of problemen met uitspraak. Hier heeft vooral Kathy last van. Bij Justyna merk je eigenlijk bijna niet dat ze geen moedertaalspreker is. Maar door de haperingen van Kathy heeft haar gesprekspartner ('haar' T2'er) geen voorbeeld van de perfecte Nederlandse uitspraak. Of dit invloed heeft gehad, is niet onderzocht omdat Kathy's gesprekspartner alleen met Kathy gesproken heeft. Er is geen controlegroep. De prestaties van Kathy's gesprekspartner zijn wel meegenomen bij de analyse van de resultaten om zoveel mogelijk oordelen te kunnen analyseren. Bij een vervolgonderzoek kan ervoor gekozen worden om alle tweedetaalverwervers met dezelfde moedertaalspreker de gesprekstaken uit te laten voeren. Zo sluit je een invloed van verschillende moedertaalsprekers uit. Dit is wel lastiger te realiseren, omdat er dan veel meer gevraagd wordt van de T1'er. De T1'er voert immers niet slechts drie gesprekken uit, maar drie gesprekken per deelnemer. Ook zou het kunnen dat factoren als verveling en vermoeidheid een rol kunnen gaan spelen. De T1'er zal bij het twintigste gesprek waarschijnlijk minder enthousiast zijn dan bij het eerste gesprek.

Naast de kanttekening over de T1'ers is het goed om op te merken dat er veel technische problemen waren met de (instabiele) omgeving waar de beoordelaars de video's van de uitvoering van de gesprekstaken moesten bekijken. Hierdoor lukte het vaak niet om de video's te starten en zijn er veel missing value's ontstaan. Ook de geluidskwaliteit van de video-opnames laat te wensen over. Dit werkt niet motiverend voor de beoordelaars, aangezien zij sommige video's niet konden openen

en slecht konden verstaan. Op deze manier ging er veel meer tijd in het beoordelen zitten dan vooraf gepland was. Dit zorgde voor frustratie bij de beoordelaars.

Een probleem wat hier wellicht mee te maken heeft, is de interbeoordelaarsbetrouwbaarheid. De betrouwbaarheid bleek erg laag te zijn. De beoordelaars zitten dus niet met elkaar op één lijn. Dit komt wellicht door frustratie en ongenoegen tijdens het bekijken van de video's, maar het kan ook door het beoordelingsmodel komen. Bij een vervolgonderzoek zal dus kritisch naar het beoordelingsmodel gekeken moeten worden.

7. Bronvermelding

Bygate, M. 'Effects of task repetition: appraising the developing language of learners'. In: Willis en Willis (ed.), *Challenge and Change in Language Teaching.* Oxford, 1996. Pagina 136 – 146.

Bygate, M. 'Effects of task repetition on the structure and control of oral language'. In: Branden, K. van den, M. Bygate en J. Norris (ed.), *Task-based language teaching: a reader.* 2009. Pagina 249 – 274.

Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge, 2001.

Ellis, R., *Task-based Language Learning and Teaching.* Oxford, 2003.

Fulcher, G., 'Deluded by Artifices? The Common European Framework and Harmonization'. In: *Language Assessment Quarterly* 1 (4). 2004. Pagina 253 – 266.

Fulcher, G., *Testing second language speaking.* London, 2003.

Gelderen, A. van, *De evaluatie van spreekvaardigheid in communicatieve situaties. Globale beoordeling en gedetailleerde analyse van spreekprestaties van 11- en 12-jarigen.* Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. Academisch proefschrift. 1992.

Gass, S. en E. Varonis, 'The effect of familiarity on the comprehensibility of nonnative speech'. In: *Language Learning* 34. 1984. Pagina 65 – 89.

Jong, J. de, 'Het Europees Referentiekader'. In: *Levende Talen tijdschrift* 2 (3). 2003. Pagina 3 – 9.

Koet, T., *Polder English in Dutch ears. Empirical studies on the evaluation of the pronunciation of English as a foreign language.* Academisch proefschrift. 2007.

Luoma, S., *Assessing speaking.* Cambridge, 2004.

Lynch, T. en J. Maclean, 'Exploring the benefits of task repetition and recycling for classroom language learning'. In: *Language Teaching Research* 4 (3). 2000. Pagina 221 – 250.

Plough, I. en S. Gass, 'Interlocutor and task familiarity: effects on interactional structure'. In: Crookes, G. en S. Gass (ed.), *Tasks and language learning. Integrating theory and practice.* 1993. Pagina 35 – 56.

Rijlaarsdam, G., *Beoordelen van discussievaardigheid. Constructie van analytische schema's voor de*

beoordeling van discussieleider en discussiedeelnemers. Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. SCO-rapport. 1982.

Rijlaarsdam, G. en H. Wesdorp, *Het beoordelen van taalvaardigheid in het onderwijs: Nederlandse onderzoeksresultaten op het gebied van het meten van de schrijf-, spreek-, lees- en luistervaardigheid.* Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. SCO-rapport nr. 28. 1984.

Robinson, P., 'Task complexity, cognitive resources and syllabus design: a triadic framework for examining task influences on SLA'. In: Robinson, P. (ed.), *Cognition and Second Language Instruction.* Cambridge, 2001. Pagina 287 – 318.

Skehan, P., *A cognitive approach to Language Learning.* Oxford, 1998.

Wesdorp, H., *Evaluatie-technieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid.* Amsterdam: Stichting Research Instituut voor de toegepaste psychologie aan de Universiteit van Amsterdam. 1978.

Wijgh, I., *Gespreksvaardigheid in de vreemde taal. Begripsbepaling en toetsing.* Proefschrift Universiteit Utrecht, 1996.

Yule, G., M. Powers en D. McDonald, 'The variable effects of some task-based learning procedures on T2 communicative effectiveness'. In: *Language Learning* 42. 1992. Pagina 249 – 277.

Gastcollege Rick de Graaff over het NIFLAR project tijdens de cursus 'Language Education' 15-02-10 '*Effectieve taken voor Video-web communicatie'*

Website van onderzoeksproject NIFLAR:
www.niflar.eu

Website van de cursus 'Language Education' 2008/2009:
http://www.surfgroepen.nl/sites/languageeducation

Bijlagen

Bijlage 1: Een voorbeeld van een gesprekstaak
Bijlage 2: Beoordelingsmodel + toelichting

Bijlage 1

Een voorbeeld van een gesprekstaak

(uit de cursus 'Language Education' 2008/2009)

Taak 2: Wonen (Jesse, Lotte, Justyna)

Ontwikkeld voor niveau A2 (CEFR)

Doelen van het gesprek:

- Uitwisseling van informatie over de verschillen en overeenkomsten in twee culturen wat betreft het onderwerp "Wonen"
- Negotiation of meaning
- Aanmoediging tot spreken

Start van de taak:

Hoe gaat het? Weet je nog wat we vorige week hebben gedaan/besproken?

Introductiefilmpje glorix:

http://www.youtube.com/watch?v=GwOV-_8IyYk&feature=PlayList&p=9CF610A0A7C6E8DD&playnext=1&index=97

Wat zie je in het filmpje? Wat vind je hiervan? Is het ook zo bij jou thuis? Denk je dat alle studenten zo leven in Nederland?

Wij laten foto's van andere huizen zien (2 paren: flat en herenhuis, rijtjeshuis en spacebox)

Weet je hoe dit heet?
Welke verschillen zijn er?
Waar woon jij in? Dichtbij het centrum? Aan de rand van de stad? Een heel klein dorpje?
Waar zou je willen wonen?

Eigen foto student

Nu ben ik benieuwd naar jouw huis. Hoe ziet het er uit? Hoe groot is je kamer?
Wat staat er in je kamer? Beschrijven. (bed, bank, wastafel, gemeenschappelijke ruimte)

Waar ben je trots op? Wat kan beter?
Heb je een kamer- of een huisgenoot?
Wat doe je op je kamer (slapen/ eten/ feesten/ …)

Samenvatting:

Wat heb je geleerd? Wat was was nieuw?
Zijn er veel verschillen in het wonen in NL en Tsjechie? Welke?
Wat is beter in Tsjechie? Wat is beter in NL?

Opdracht voor de volgende keer

Woordenweb van taak 1 uitbreiden (over het onderwerp 'vrije tijd´)

Bijlage 2

Beoordelingsmodel videoanalyse

Naam beoordelaar:
Naam van tweedetaalverwerver (L2'er):
Naam van moedertaalspreker (L1'er):
Nummer van taak/sessie:

Complexiteit

1. De L2'er varieert in zinsbouw

Helemaal niet	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

2. De L2'er varieert in woordkeus

Helemaal niet	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

Nauwkeurigheid

3. De L2'er gebruikt inhoudelijk correcte woorden

Nooit	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

4. De L2'er maakt fouten bij de zinsbouw

Nooit	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

5. De L2'er maakt grammaticale fouten op woordniveau

Nooit	*heel soms*	*soms*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

Vloeiendheid

6. Het spreektempo van de L2'er is …

heel langzaam	*redelijk langzaam*	*een beetje langzaam*	*neutraal*	*een beetje snel*	*redelijk snel*	*heel snel*
○	○	○	○	○	○	○

7. Komen storende aarzelingen voor?

Helemaal niet	*bijna niet*	*soms*	*neutraal*	*geregeld*	*vaak*	*heel vaak*
○	○	○	○	○	○	○

8. Komen onnatuurlijke pauzes voor?

Helemaal niet	*bijna niet*	*soms*	*neutraal*	*geregeld*	*vaak*	*heel vaak*
○	○	○	○	○	○	○

Uitspraak

9. De L2'er spreekt met juiste intonatie

nooit	*meestal niet*	*af en toe*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

10. De L2'er spreekt woorden uit met een verkeerde klemtoon

nooit	*meestal niet*	*af en toe*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

11. De L2'er spreekt de klanken goed uit

nooit	*meestal niet*	*af en toe*	*neutraal*	*geregeld*	*vaak*	*altijd*
○	○	○	○	○	○	○

Algemene indruk

Wat is je algemene indruk van de taalvaardigheid van de L2'er? ….
Geef een cijfer op een schaal van 1 – 10, waarbij geldt: 1 = heel slecht, 10 = heel goed

Toelichting op het beoordelingsmodel

Complexiteit

1. Variatie in zinsbouw = worden er alleen korte zinnen gebruikt of gebruikt de L2'er soms ook langere, complexere zinnen? Gebruikt de L2'er een standaardvolgorde met het onderwerp voorop, gevolgd door het gezegde of wordt hierin afgewisseld (doordat bijvoorbeeld een bijwoord als 'gisteren' voorop geplaatst wordt)?
2. Variatie in woordkeus = gebruikt de L2'er constant dezelfde woorden of komen er veel verschillende woorden aan bod?

Nauwkeurigheid

3. Correcte woorden = is de L2'er in staat om de juiste woorden in de context te gebruiken? Bv. het gebruik van het woord 'lekker' bij een gesprek over eten. Dus: 'het eten is niet lekker' i.p.v. 'het eten is niet goed'.
4. Fouten bij zinsbouw = is er sprake van een verkeerde zinsvolgorde? (bijvoorbeeld: 'Ik denk dat hij weet dat niet' i.p.v. 'Ik denk dat hij dat niet weet')
5. Fouten op woordniveau = gebruikt de L2'er dingen als bijvoeglijke naamwoorden, voorzetsels en meervoudsvormen op de verkeerde manier? (denk aan: 'ik lopen' of 'een grote huis')

Vloeiendheid

6. Spreektempo = snelheid van spreken. Hier geldt niet: hoe sneller het tempo, hoe beter de L2'er spreekt. De L2'er kan namelijk ook te snel spreken. Wees je hiervan bewust. (Heel snel is dus niet automatisch heel goed)
7. Storende aarzelingen = dingen als "uhm"
8. Onnatuurlijke pauzes = geen pauzes om nadruk te geven aan wat gezegd is, maar pauzes die aangeven dat de spreker aan het nadenken of aan het herformuleren is.

Uitspraak

9. Intonatie = spreekt de L2'er onnatuurlijk monotoon of is er (goede) variatie in toonhoogte? Bijvoorbeeld: bij een vraag gaat je stem aan het eind van de zin omhoog en bij een 'gewone' zin omlaag.
10. Klemtoon = legt de L2'er de juiste klemtoon, dus 'opa' als /O-pa/ of /o-PA/? En 'familie' als FA-mi-lie, fa-MI-lie of fa-mi-LIE?
11. Goede uitspraak = spreekt de L2'er een woord als 'huis' uit als /huis/ of als /haus/? En een woord als 'chaos' als /gaos/ of /kaos/?

Algemene indruk

(geen toelichting)