

# ANTAT-CU4

Een onderzoek naar de betrouwbaarheid van de ANTAT-CU4  
bij gezonde sprekers



Masterscriptie Taal-Spraak Pathologie

Auteur: Judith Zwartjens  
Studentnummer: 4598776  
Begeleider: Dr. M.B. Ruiters  
Tweede lezer: Prof. Dr. R.W.N.M. van Hout,  
Datum: augustus 2017



## Voorwoord

In de periode van januari 2017 tot augustus 2017 heb ik, in het kader van mijn afstuderen aan de Opleiding Taal-en Spraakpathologie aan de Radboud Universiteit Nijmegen, aan de scriptie gewerkt, die nu voor u ligt. Bij nader inzien is dat wellicht niet meer van deze tijd, dus zal ik het anders verwoorden. Ik heb hard gewerkt aan de scriptie, die u nu voor u op het scherm ziet. Dit brengt me meteen op de aanleiding van het onderzoek. In het huidige onderzoek wordt namelijk een verbeterde versie van de “ouderwetse” Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (ANTAT; Blomert, Kean, Koster, & Schokker, 1995) onderzocht. Op basis van responsen van gezonde sprekers wordt gekeken naar de interne consistentie, test-hertestbetrouwbaarheid en de inter- en intrabeoordelaarsbetrouwbaarheid van deze nieuwe versie, de ANTAT-CU4.

In de scriptie wordt eerst aandacht besteed aan het onderwerp afasie en de testinstrumenten die beschikbaar zijn voor het onderzoeken van de communicatieve vaardigheden van sprekers met afasie. Daaropvolgend komen verschillende theoretische termen aan bod, die van belang zijn bij het onderzoek. Dan worden de gebruikte onderzoeksmethoden beschreven en worden de resultaten besproken die uit het onderzoek naar voren kwamen. Als laatste is er een kritische terugblik op het onderzoek en worden aanbevelingen gedaan voor toekomstig onderzoek omtrent dit onderwerp.

Graag wil ik hierbij van de gelegenheid gebruik maken om iedereen te bedanken die mij heeft geholpen tijdens het onderzoek, het schrijfproces en de afronding van de scriptie. Op de eerste plaats wil ik graag mijn dank uitspreken aan mijn begeleider dr. M.B. Ruiten. Zij voorzag me bij ieder overleg van voldoende feedback om het uiterste uit mezelf en het onderzoek te halen. Ook wil ik graag Prof. dr. R.W.N.M. van Hout bedanken voor zijn geduld en advies, wanneer ik vragen had betreffende statistische onderdelen van mijn onderzoek. Tevens mijn dank aan alle proefpersonen die aan het onderzoek hebben deelgenomen. Dankzij hun creatieve verhalen en vertellingen konden de verbale vaardigheden van gezonde sprekers worden onderzocht.

Daarnaast wil ik graag Marja Halters bedanken, die altijd een luisterend oor had. Ook ben ik grote dank verschuldigd aan Iris van den Berg, die als derde beoordelaar wilde optreden tijdens het onderzoek en altijd voor me klaarstond. Als laatste wil ik mijn familie bedanken voor hun steun, bemoedigende woorden en dat ze altijd op de achtergrond voor mij aanwezig waren.

Judith Zwartjens, Best, augustus 2017



## Inhoudsopgave

Voorwoord.....	iii
Samenvatting .....	vii
Summary.....	ix
1. Inleiding.....	1
1.1 Indeling van de scriptie .....	7
2. Theoretische achtergrond.....	8
2.1 Waarom is een nieuwe ANTAT nodig? .....	8
2.2 Ontwikkeling van ANTAT naar ANTAT-CU4 .....	8
2.3 Een kwantitatief scoringssysteem .....	10
2.4 Functionele communicatie .....	11
2.5 Propositionen en Content Units.....	12
2.6 Verbale effectiviteit (% Content Units).....	15
2.7 Verbale efficiëntie (aantal Content Units per minuut) .....	15
2.8 Componenten van betrouwbaarheid .....	16
3. Methode .....	23
3.1 Onderzoeksopzet.....	23
3.2 Proefpersonen .....	24
3.3 Materiaal .....	27
3.4 Procedure .....	27
3.5 Uitwerking audio-opnamen, meten en scores .....	29
3.6 Analyse.....	33
4. Resultaten .....	36
4.1 Betrouwbaarheid van de ANTAT-CU4; 20 scenario's.....	36
4.2 Betrouwbaarheid van de ANTAT-CU4; 23 scenario's.....	37
4.3 Test-hertestbetrouwbaarheid ANTAT-CU4; 20 scenario's.....	38
4.4 Test-hertestbetrouwbaarheid van de ANTAT-CU4; 23 scenario's .....	39
4.5 Begrip van testinstructie.....	40
4.6 Interbeoordelaarsbetrouwbaarheid ANTAT-CU4; 20 scenario's .....	41
4.7 Interbeoordelaarsbetrouwbaarheid ANTAT-CU4; 23 scenario's .....	42
4.8 Intrabeoordelaarsbetrouwbaarheid; 20 scenario's.....	43
4.9 Intrabeoordelaarsbetrouwbaarheid; 23 scenario's.....	45
5. Discussie.....	47

5.1 Wat is de mate van de interne consistentie van de ANTAT-CU4?.....	47
5.2 Wat is de mate van de test-hertestbetrouwbaarheid van de ANTAT-CU4 bij gezonde sprekers? .....	48
5.3 Wat is de mate van interbeoordelaarsbetrouwbaarheid van de ANTAT-CU4? .....	52
5.4 Intrabeoordelaarsbetrouwbaarheid .....	55
5.5 Onderzoeksopzet.....	55
5.6 Handleiding.....	56
5.7 Aanbevelingen ten aanzien van het ANTAT-CU4-schema .....	56
5.8 Onderzoeksgroep .....	58
5.9 Conclusie.....	59
Referenties.....	60
Bijlagen .....	65
Bijlage A: Afbeelding ‘Cookie theft picture’ .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage B: Schematisch overzicht proefpersonen (n=31).....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage C: Overzicht proefpersonen interbeoordelaarsbetrouwbaarheid (n=6) ....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage D: Overzicht proefpersonen intrabeoordelaarsbetrouwbaarheid (n=18) ..	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage E: Testinstructie afname scenario’s ANTAT-CU4.....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage F: Gebruikte afbeeldingen bij afname ANTAT-CU4..	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage G: Informatiebrief deelnemers (Dassek, 2016).....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage H: Informatiebrief zeven overige participanten .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage I: Toestemmingsverklaring deelnemers (Dassek, 2016).....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage J: Toestemmingsverklaring zeven overige participanten .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage K: Vragenlijst persoonsgegevens (Dassek, 2016) ....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage L: Vragenlijst persoonsgegevens zeven overige participanten .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage M: ANTAT-CU4-schema, opgesteld door Dassek (2016).....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage N: Effectiviteit 20 scenario’s .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage O: Efficiëntie 20 scenario’s.....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage P: Effectiviteit 23 scenario’s .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Bijlage Q: Efficiëntie 23 scenario’s.....	<b>Fout! Bladwijzer niet gedefinieerd.</b>

Bijlage R: Ruwe scores intrabeoordelaarsbetrouwbaarheid. **Fout! Bladwijzer niet gedefinieerd.**

Bijlage S: Aangepast ANTAT-CU4-schema..... **Fout! Bladwijzer niet gedefinieerd.**

## Samenvatting

*Inleiding:* Afasie is een verworven taalstoornis ten gevolge van hersenletsel (Bastiaanse, 2011; Dharmaperwira-Prins & Maas, 1998). Met de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (ANTAT; Blomert, Kean, Koster, & Schokker, 1995) kunnen de verbale communicatievaardigheden van sprekers met afasie, met behulp van scenario's, worden onderzocht. Echter, kan dan alleen een uitspraak worden gedaan over een deelaspect van de verbale communicatie, namelijk de verbale effectiviteit. Met de ANTAT wordt de verbale efficiëntie niet onderzocht. Ook wordt bij de ANTAT een kwalitatief scoringsmodel gehanteerd. Ruiters, Kolk, Rietveld, Dijkstra en Lotgering (2011) hebben onderzocht of het construct van de ANTAT kon worden verbeterd door het kwalitatieve scoringsmodel te vervangen door een kwantitatief scoringsmodel (ANTAT-CU4-schema), waarmee wordt gepoogd om zo objectief mogelijk zowel de verbale effectiviteit als de efficiëntie te meten. Niet alleen werd het scoringsmodel aangepast, ook de scenario's van de ANTAT zijn bijna allemaal vervangen of aangepast. De reden hiervoor was dat de scenario's niet meer als actueel, duidelijk en inleefbaar werden beschouwd (Filipinski, 2014). Het onderzoeksinstrument met de nieuwe en verbeterde scenario's en het aangepaste scoringsmodel wordt de ANTAT-CU4 genoemd. In het huidige onderzoek worden verschillende betrouwbaarheidsmaten van de ANTAT-CU4 onderzocht.

*Methode:* De ANTAT-CU4 werd tweemaal, met een tussentijd van acht weken, bij een groep van 31 gezonde sprekers afgenomen. Alle testafnamen werden getranscribeerd en gescoord op basis van het ANTAT-CU4-schema. Deze scores werden gebruikt om onder andere de interne consistentie en de test-hertestbetrouwbaarheid van de ANTAT-CU4 te onderzoeken. Dit laatste werd gedaan door de behaalde scores op het gebied van verbale effectiviteit en efficiëntie met elkaar te vergelijken. Tevens werd de inter- en intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4 onderzocht.

*Resultaten:* De interne consistentie van de ANTAT-CU4 bleek matig (tweede testafnamemoment) tot adequaat (eerste testafnamemoment) te zijn. Tussen beide testafnamen werd geen significant verschil gevonden wat betreft verbale effectiviteit. Wel was de verbale efficiëntie op het tweede testafnamemoment significant toegenomen.

In totaal hebben drie beoordelaars de prestaties van zes proefpersonen beoordeeld. De gegeven scores van een van de beoordelaars had een zeer zwakke correlatie met de andere twee beoordelaars. Deze twee beoordelaars lieten een hoge onderlinge correlatie zien, maar desondanks waren de gegeven scores significant verschillend.

Wat betreft de intrabeoordelaarsbetrouwbaarheid werd door de beoordelaar op het tweede meetmoment een significant hogere gemiddelde score gegeven dan op het eerste meetmoment.

*Discussie en conclusie:* Op beide testafnamemomenten was de interne consistentie van de ANTAT-CU4 te laag om van een goede betrouwbaarheid te spreken. In vervolgonderzoek zullen de scenario's, instructies en de CU-schema's kritisch bekeken moeten worden om na te gaan of met het testinstrument daadwerkelijk alleen de verbale effectiviteit en efficiëntie worden onderzocht.

Op beide testafnamen produceerden gezonde sprekers ongeveer evenveel essentiële informatie-eenheden (Content Units; Yorkston & Beukelman, 1980), maar op de tweede testafname produceerden ze een hoger aantal essentiële informatie-eenheden in een kortere tijd. De verbale effectiviteit was dus nagenoeg gelijk, terwijl de verbale efficiëntie toenam. Dit was in lijn met onderzoek van Linebarger, McGall, Virata en Berndt (2007).

De verschillen in de gegeven scores van twee van de drie beoordelaars waren klein, maar desondanks significant. De mate van achtergrondkennis en ervaring in het gebruik van het ANTAT-CU4-schema kunnen hierbij mogelijk een rol hebben gespeeld.

De verschillen in de gegeven scores voor de intrabeoordelaarsbetrouwbaarheid waren minimaal. Een mogelijk verklaring is dat er sprake was van een leereffect bij de beoordelaar of dat er sprake was van een statistisch artefact.

Het is van belang dat er een goede handleiding komt met duidelijke instructies voor het gebruik van het ANTAT-CU4-schema. Een ander belangrijk punt is dat het huidige onderzoek werd uitgevoerd bij gezonde sprekers. Uit onderzoek van Brookshire en Nicholas (1994b) bleek dat gezonde sprekers over de tijd heen minder stabiel scoren dan personen met afasie. Mogelijk kan een betrouwbaarheidsonderzoek van de ANTAT-CU4 bij sprekers met afasie dus een andere uitkomst hebben.



## Summary

*Introduction:* Aphasia is an acquired language disorder due to brain injury (Bastiaanse, 2011; Dharmaperwira-Prins & Maas, 1998). With the Amsterdam-Nijmegen Everyday Language Test (ANELT; Blomert, Kean, Koster, & Schokker, 1995), the verbal communication skills of speakers with aphasia can be investigated using scenarios. However, with the ANELT one can only judge one aspect of verbal communication, namely verbal effectiveness. The degree of verbal efficiency is not investigated with the ANELT. The verbal effectiveness is being judged with the use of a qualitative scoring model. Ruiters, Kolk, Rietveld, Dijkstra and Lotgering (2011) investigated whether the construct of the ANELT could be improved by replacing the qualitative scoring model with a quantitative scoring model (ANELT-CU4 schedule). With the ANELT-CU4 schedule they aimed to measure verbal effectiveness as well as verbal efficiency as objective as possible. Apart from the adjusted scoring model, almost all the scenarios of the ANELT were replaced or modified. This was because the scenarios were no longer considered to be current, clear and imaginable (Filipinski, 2014). The research tool with the new and improved scenarios and the adjusted scoring model is called the ANELT-CU4. The current research examines various measurements of reliability of the ANELT-CU4.

*Method:* A group of 31 healthy speakers was tested twice with the ANELT-CU4, with an eight-week interval. All test samples were transcribed and scored based on the ANELT-CU4 schedule. These scores were used to examine, inter alia, the internal consistency and test-retest reliability of the ANELT-CU4. The latter was done by comparing the achieved scores of verbal effectiveness and efficiency. The inter- and intra-rater reliability of the ANELT-CU4 was also investigated.

*Results:* The internal consistency of the ANELT-CU4 ranged from moderate (second test session) to adequate (first test session). There was no significant difference in verbal effectiveness between the two tests. Verbal efficiency had significantly increased in the second test session.

In total, three reviewers rated the performance of six speakers. The scores given by one of the reviewers had a very weak correlation with the other two reviewers. These two reviewers showed a high mutual correlation, but nevertheless the given scores were significantly different.

Regarding the inter-rater reliability, the investigator gave a significantly higher average score at the second measuring moment compared to the first measuring moment.

*Discussion and conclusion:* In both test moments, the internal consistency of the ANELT-CU4 was too small to speak of a good reliability. In follow-up research, the scenarios, instructions and the CU-schedules will have to be critically reviewed to investigate whether the ANELT-CU4 truly only examines verbal effectiveness and efficiency.

On both test sessions, healthy speakers produced approximately the same amount of essential Content Units (CUs; Yorkston & Beukelman, 1980), but in the second test run they produced a higher number of essential Content Units in a shorter period of time. The verbal effectiveness was thus almost equal, while verbal efficiency increased. This was similar with the research of Linebarger, McGall, Virata and Berndt (2007).

The differences in the scores given by two out of three reviewers were small, but nevertheless significant. Possible causes may have been the extent of background knowledge and experience in using the ANELT-CU4 schedule.

The differences in the given scores for the intra-rater reliability were minimal. One possible explanation is that there was a learning effect or that there was a statistical artefact.

It is important that a manual will be available with clear instructions for using the ANELT-CU4 schedule. Another important issue is that current research was conducted with healthy speakers. Research from Brookshire and Nicholas (1994b) showed that healthy speakers are more instable over time than persons with aphasia. Research of the reliability of the ANELT-CU4 with speakers with aphasia may therefore have a different outcome.

## 1. Inleiding

In Nederland zijn naar schatting 30.000 mensen die een vorm van afasie hebben (Afasie Vereniging Nederland, 2017; SHV Hersenletsel, 2017). Afasie is een verworven taalstoornis ten gevolge van hersenletsel (Bastiaanse, 2011; Dharmaperwira-Prins & Maas, 1998). Met verworven wordt bedoeld dat de stoornis ontstaat na de fase van taalverwerving (Dharmaperwira-Prins & Maas, 1998). Afasie kan verschillende oorzaken hebben, zoals een beroerte of traumatisch hersenletsel (Kemmerer, 2015; Bastiaanse, 2011). Hallowell en Chapey (2008) beschrijven dat afasie vaak beschouwd wordt als een symbolische verwerkingsstoornis waarbij men problemen heeft met de formulering en interpretatie van taalkundige symbolen. Ze benoemen dat het van belang is om te erkennen dat een of meerdere modaliteiten van de communicatie aangedaan kunnen zijn. Hierdoor kan een afatische stoornis grote invloed hebben op het sociaal functioneren (Berns, et al., 2015). De communicatie met de omgeving wordt bemoeilijkt doordat mensen met afasie onder andere problemen kunnen hebben met het begrijpen van gesproken taal, spreken, lezen en schrijven (Bastiaanse, 2011). Dharmaperwira-Prins en Maas (1998) benoemen dat de meeste personen met afasie zich bewust zijn van de taalproblemen en de symptomen die erbij horen. Naast taalstoornissen kunnen mensen met afasie volgens Patterson en Chapey (2008) ook cognitieve stoornissen hebben, die van invloed zijn op de taal, en kunnen mensen met afasie moeilijkheden hebben met het begrijpen en gebruiken van gebaren. Door de verschillende stoornissen bestaat er grote variatie in de problemen die mensen met afasie kunnen ervaren. Het is belangrijk om te onderzoeken op welk vlak van de communicatie de personen met afasie de meeste belemmering ervaren, omdat persoonlijk contact gezien wordt als de meest fundamentele vorm van sociale betrokkenheid in de maatschappij (Sacks, 1972).

Volgens Albin Davis (2014) hebben bijna alle mensen met afasie op een bepaald niveau moeite met het begrijpen van taal. Zo kunnen personen met afasie volgens Kemmerer (2015) bijvoorbeeld moeite hebben met foneemperceptie en woordherkenning, maar volgens Albin Davis hebben maar weinig van hen moeite met het begrijpen van losse woorden. Uit onderzoek van Hagoort (1992) kwam naar voren dat de taalbegripstoornissen, die mensen met afasie ervaren, te wijten zijn aan het onvoldoende kunnen samenbrengen van lexicale eenheden. Tevens benoemt Hagoort dat dit verminderde lexicale selectie- en integratieproces bij personen met Wernicke-afasie (een van de afasiesyndromen) mogelijk zo erg vertraagd is, dat het lexicale proces niet meer volledig voltooid kan worden. Hierdoor hebben mensen met afasie moeite met het herkennen en begrijpen van woorden. Daarnaast benoemt Albin Davis dat bijna alle personen met afasie in enige mate moeite hebben met het begrijpen van zinnen.

Bij de taalproductie kunnen personen met afasie onder andere problemen hebben met het vinden van woorden voor objecten of handelingen (Kemmerer, 2015; Berndt, 2007). Papathanasiou, Coppens en Potagas (2013) beschrijven dat woordvindingsproblemen regelmatig voorkomen bij mensen met afasie, maar dat het soort woordvindingsproblemen differentieert voor de verschillende afatische syndromen. Verder kunnen personen met afasie moeite hebben met morfologische structuren van woorden en grammaticale structuren van zinnen (Kemmerer, 2015). Brookshire en Nicholas (1994a) beschrijven dat de meeste mensen met afasie moeite hebben met het effectief en efficiënt produceren van aaneengesloten taal om de luisteraar te informeren. Er bestaat dus veel variatie in de afatische problemen die voor kunnen komen. Om deze reden is het wenselijk om per persoon het type afasie en de ernst van de afatische stoornis in kaart te

brengen en de taalkundige en sociale communicatieve vaardigheden van iemand met afasie te onderzoeken.

Er zijn verschillende testen waarmee afasie onderzocht kan worden. Testen waarmee bekeken kan worden of bij iemand al dan geen sprake is van afasie, zijn bijvoorbeeld de Token Test (De Renzi & Vignolo, 1962) en de Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972). De Token Test wordt vaak als onderdeel gebruikt van de Akense Afasie Test (AAT; Graetz, De Bleser, & Willmes, 1992). Met de AAT wordt niet alleen onderzocht of de cliënt al dan geen afasie heeft, maar wordt tevens de ernst van de afasie vastgesteld en de afasie geclassificeerd in een van de vier hoofdsyndromen (Broca, Wernicke, Globale en Amnestische afasie). Afasietesten waarmee de soort afasie en de ernst van de afatische stoornis onderzocht kunnen worden, kunnen gericht zijn op het taalbegrip of de taalproductie. Zo richt de Token Test zich op het auditief taalbegrip, terwijl de Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1978) de woordproductie (woordvinding) onderzoekt.

Hoewel uit bovenstaande en soortgelijke testen veel informatie omtrent de stoornis gehaald kan worden, is het volgens Bastiaanse (2011) niet voldoende om alleen een afasietest af te nemen om de gevolgen van het hersenletsel in kaart te brengen. Ze beschrijft dat een spontane taalanalyse een cruciale aanvulling is op de beoordeling van de spontane taal door middel van afasieonderzoek. Met behulp van een spontane taalanalyse kan volgens Bastiaanse (2011) worden nagegaan wat de onderliggende oorzaak is voor de problemen in de alledaagse communicatie. Hoewel Bastiaanse overtuigd is van het belang van een spontane taalanalyse om erachter te komen welke problemen sprekers met afasie tegenkomen in de alledaagse communicatie, is niet iedereen het met haar eens dat met een spontane taalanalyse de onderliggende problemen van personen met afasie doorgrond kunnen worden. Zo beschrijven Hesketh en Bishop (1996) dat de cliënt in de spontane taal zelf kan bepalen wat hij<sup>1</sup> zegt. Op deze manier kan de cliënt aanpassingsgedrag laten zien. Zo kan er bijvoorbeeld voor gekozen worden om minder te spreken of om bepaalde woordsoorten te vermijden. Hierdoor kunnen taalproblemen mogelijk minder opvallen. Compensatiestrategieën bij personen met afasie worden gezien als lucratieve methoden om communicatieve hindernissen te overwinnen (Luria, 1970; Simmons-Mackie & Damico, 1997). Volgens Kolk en Heeschen (1990) is aanpassing echter optioneel, niet verplicht. Zo zouden aanpassingsstrategieën afwezig kunnen zijn als de cliënt zich bijvoorbeeld minder bewust is van zijn taalstoornis (Goodglass & Kaplan, 1983). Een conversatie lijkt een situatie te zijn waarin voldoende gelegenheid is voor de cliënt om aanpassingsstrategieën toe te passen, hoewel dit volgens Hesketh en Bishop (1996) waarschijnlijk afhankelijk is van de conversatiestijl van de persoon. Ze geven aan dat door het toepassen van compensatie- en aanpassingstechnieken de spontane taal niet de volledige capaciteiten en beperkingen van een spreker met afasie weerspiegelt.

Hoewel mensen met afasie wellicht hun taalgedrag kunnen aanpassen met compensatie- en aanpassingsstrategieën, zijn Prins en Bastiaanse (2004) van mening dat spontane taalanalyse bij mensen met afasie wel degelijk nuttig kan zijn. Ze benoemen dat de taalproblemen voornamelijk tijdens het spontane spreken naar voren komen. De alledaagse communicatie van cliënten met afasie kan belemmerd worden op verschillende linguïstische gebieden (zoals fonologie,

---

<sup>1</sup> Omwille van de leesbaarheid wordt 'hij' gebruikt in de tekst wanneer verwezen wordt naar iemand met afasie. Waar 'hij' staat kan uiteraard ook 'zij' gelezen worden.

woordvinding en morfosyntaxis). Volgens Prins en Bastiaanse kan met een geraffineerde analyse van de spontane taal, de invloed van deze linguïstische processen geïdentificeerd worden. Dit kan door het taalgebruik op verschillende variabelen kwalitatief te beoordelen (bijvoorbeeld gebruikte grammaticale vorm en parafasieën) en/of door de linguïstische variabelen te kwantificeren (bijvoorbeeld gemiddelde uitinglengte en de ratio inhouds- en functiewoorden). Een voorbeeld van een spontane taalanalyse is de Analyse voor Spontane Taal bij Afasie (ASTA; Boxum, Van der Scheer, & Zwaga, 2013). Volgens Boxum et al. (2013) kan de ASTA een uitspraak doen over de onderliggende stoornissen en de samenhang tussen deze stoornissen, die bij personen met afasie een belemmerende rol kunnen spelen tijdens het spontaan spreken.

Behalve het afnemen van een spontane taalanalyse benoemt Bastiaanse (2011) ook dat het van belang is om na te gaan hoe goed een cliënt met afasie informatie kan overbrengen met de communicatieve functies die nog intact zijn. Om dit te onderzoeken zijn verschillende testen beschikbaar. Zo is er bijvoorbeeld de Communicative Abilities in Daily Living (CADL; Holland, 1980), waarmee volgens Dharmaperwira-Prins en Maas (1998) veel informatie verkregen kan worden over de communicatieve vaardigheden van de cliënt in verschillende situaties. Hoewel de CADL niet vertaald is naar het Nederlands, wordt hij hier toch benoemd om te dienen als een voorbeeld van afasieonderzoek. Naast afasieonderzoeken als de CADL is er de Scenario Test (Van der Meulen, Van Gelder-Houthuizen, Wiegers, Wielaert, & Van de Sandt-Koenderman, 2008), die de verbale en non-verbale communicatieve vaardigheden van mensen met afasie meet met behulp van scenario's. Doordat er gebruik wordt gemaakt van scenario's, worden meer specifieke reacties uitgelokt. Door bepaalde reacties uit te lokken, kunnen volgens Hesketh en Bishop (1996) taalcapaciteiten tot uiting komen, die niet in de spontane taal zichtbaar zijn. Volgens hen kan op deze manier het taalvermogen van mensen met afasie beter in beeld gebracht worden dan wanneer alleen gekeken wordt naar de spontane taal, omdat de spontane taal niet volledig de onderliggende capaciteiten en beperkingen weerspiegelt.

Een ander onderzoeksinstrument dat gebruik maakt van scenario's is de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (ANTAT; Blomert, Kean, Koster, & Schokker, 1995). Met deze test worden de verbale communicatievaardigheden van sprekers met afasie onderzocht met behulp van alledaagse scenario's waarbij sociale interactie vereist is (Blomert, Kean, Koster, & Schokker, 1994). Er wordt op talig gebied verwacht dat de persoon met afasie de testinstructie kan begrijpen en ook dat de cliënt enigszins een boodschap onder woorden kan brengen. Om deze reden is de ANTAT bedoeld voor personen met een milde afasie. De ANTAT bestaat uit twee parallelle versies van tien alledaagse scenario's. De cliënt wordt gevraagd wat hij in deze bepaalde situaties zou zeggen. De scenario's hebben het karakter van een script en zorgen er volgens Blomert et al. (1994) voor dat de cliënt zo min mogelijk stress ervaart, waardoor de optimale prestaties van de cliënt bevorderd worden. Het onderzoek bestaat uit korte communicatieve situaties, die lijken op situaties die in het dagelijks leven daadwerkelijk voor zouden kunnen komen. Op deze manier wordt een soort monoloog uitgelokt. Er wordt onderzocht welke verbale communicatievaardigheden de persoon met afasie inzet om in het dagelijks leven te communiceren met zijn omgeving. Tevens wordt gekeken naar de communicatieve effectiviteit van de spreker.

Met communicatieve effectiviteit wordt bedoeld hoe begrijpelijk een spreker een boodschap over kan brengen naar de luisteraar (Ruiter, Kolk, Rietveld, Dijkstra, & Lotgering, 2011). Hoewel communicatieve effectiviteit zowel verbaal als non-verbaal kan zijn, wordt met de ANTAT alleen

gekeken naar de verbale effectiviteit. Een van de redenen hiervan is volgens Blomert et al. (1995) dat er weinig kennis is over de interactie van verbale en non-verbale aspecten van communicatie. Daarnaast zijn taalstoornissen bij mensen met afasie het meest opvallend (Blomert et al., 1994). De verbale communicatieve effectiviteit van een cliënt wordt kwalitatief beoordeeld op een 5-puntsschaal. Met behulp van de 5-puntsschaal beoordeelt de luisteraar hoe begrijpelijk de boodschap is die de spreker over wil brengen, namelijk: 1 “geheel niet”, 2 “enigszins”, 3 “middelmatig”, 4 “redelijk” of 5 “goed” (Blomert et al., 1995). Blomert et al. (1994) geven aan dat hiermee een valide en betrouwbare meting gedaan kan worden om de verbale communicatievaardigheden van de cliënt in kaart te brengen. Hier zit echter wel een kanttekening aan. Zo zouden de cijfers op de 5-puntsschaal suggereren dat de afstanden tussen de beoordelingen gelijk zijn, zoals bij een liniaal, terwijl dit niet het geval is (Eithward, 2012). Omdat de 5-puntsschaal een ordinale variabele is, zijn optellingen en gemiddelden van deze scores eigenlijk niet betrouwbaar (Eithward, 2012).

Hoewel met de 5-puntsschaal, volgens Blomert et al. (1994), een betrouwbare meting gedaan kan worden, is deze beoordeling wel subjectief. Een subjectieve beoordelingsmethode houdt in dat scores gegeven worden op basis van persoonlijke overtuigingen, meningen en verwachtingen van de testleider (Dollaghan, 2007). Zo beschrijft Niewold (2006) dat verschillende beoordelaars in de praktijk eenzelfde mening zullen hebben over wanneer een extreme waarde (1 “geheel niet” of 5 “goed”) toegekend zou moeten worden, maar dat de meningen over het toekennen van de tussenliggende waarden van elkaar verschillen.

Bij een objectieve beoordelingsmethode worden, in tegenstelling tot bij een subjectieve methode, de persoonlijke overtuigingen van de onderzoeker zoveel mogelijk buiten beschouwing gelaten. Maso en Smaling (1998) beschrijven objectief als recht doen aan het object van onderzoek. Daarmee bedoelen ze dat het onderwerp van de betreffende studie het belangrijkste is en dat je hier zoveel mogelijk over te weten moet komen, zonder er een vertekend beeld van te krijgen.

Met de ANTAT kan de verbale communicatieve effectiviteit op een betrouwbare manier in kaart gebracht worden. Helaas is dit niet voldoende als we de volledige functionele communicatievaardigheid van een spreker met afasie willen onderzoeken. Met functionele communicatie wordt volgens Filipinski (2014) door verschillende onderzoekers aangeduid hoe succesvol mensen (in dit geval personen met afasie) informatie kunnen overbrengen of ontvangen, waarbij ze gebruik maken van alle mogelijke taalmodaliteiten (verbaal, non-verbaal of schriftelijk) in alledaagse situaties. Volgens de American Speech-Language-Hearing Association (ASHA; 1990, zoals geciteerd door Frattali, 1992, p. 64) is functionele communicatie “*the ability to receive a message or to convey a message, regardless of the mode, to communicate effectively and independently in a given environment*”. Dat wil zeggen dat de functionele communicatievaardigheden de vaardigheden zijn, die nodig zijn om in het dagelijks leven een boodschap zowel effectief als efficiënt over te brengen (Ruiter, Rietveld, Hoskam, & Van Beers, 2016). Deze laatste definitie van functionele communicatie is ook de definitie die tijdens het huidige onderzoek aangehouden wordt, hoewel in het huidige onderzoek alleen onderzoek gedaan wordt naar de verbale vaardigheden en de non-verbale vaardigheden buiten beschouwing worden gelaten.

Kort samengevat bestaat functionele communicatie dus uit twee aspecten, namelijk effectiviteit en efficiëntie (Yorkston & Beukelman, 1980; Nicholas & Brookshire, 1993, 1995; Ruiter et al., 2011; Ruiter et al., 2015). De (non-)verbale effectiviteit zegt iets over de begrijpelijkheid van de

boodschap en de (non-)verbale efficiëntie zegt iets over de snelheid waarmee een spreker essentiële informatie overbrengt (Ruiter et al., 2011). De effectiviteit en efficiëntie vormen samen een belangrijke indicatie van iemands functionele communicatievaardigheden (Ruiter et al., 2011). Volgens Blomert et al. (1994) is de ANTAT ontworpen om de verbale communicatievaardigheden te meten bij mensen met afasie. Met de ANTAT in zijn huidige vorm, wordt echter alleen de verbale effectiviteit gemeten en kunnen geen uitspraken gedaan worden over de verbale efficiëntie (Ruiter et al., 2011). Hierdoor kan de functionele communicatievaardigheid bij personen met afasie niet volledig in kaart gebracht worden. Daarnaast hanteert de ANTAT, zoals eerder benoemd, een subjectieve schaal voor het scoren van de communicatieve effectiviteit. Hierdoor wordt mogelijk minder betrouwbaar gescoord dan wanneer er sprake zou zijn van een objectieve scoringsmethode. Om deze redenen hebben Ruiter et al. (2011) onderzocht of het construct van de ANTAT verder verbeterd kon worden door het kwalitatieve scoringsmodel te vervangen door een kwantitatief scoringschema, waarbij rekening gehouden werd met het aantal essentiële informatie-eenheden. Zodoende kan een uitspraak gedaan worden over de hoeveelheid essentiële informatie, die een spreker met afasie over kan brengen. Met een kwantitatief scoringsmodel, het ANTAT-CU-schema, wordt gepoogd zo objectief mogelijk zowel de effectiviteit als de efficiëntie te meten.

In latere onderzoeken (Filipinski, 2014; Aan de Stegge, 2015; Giessen, 2015; Dassek, 2016) is de ANTAT-CU verder ontwikkeld tot de ANTAT-CU4. Zo werden onder andere 15 van de 20 scenario's van de originele ANTAT vervangen of aangepast. De reden voor deze verandering is dat bij het ontwikkelen van de originele ANTAT, Blomert et al. (1994) criteria hebben opgesteld waaraan de scenario's moeten voldoen. Toentertijd waren de scenario's actueel, duidelijk en inleefbaar, maar uit onderzoek van Filipinski (2014) blijkt dit nu niet meer het geval te zijn. De Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (Blomert et al., 1995) is verouderd. De ANTAT is aangepast op het gebied van de scenario's en het scoringsstelsel. Tevens is de 5-puntsschaal vervangen door het ANTAT-CU-schema.

### Doel van het onderzoek

In het huidige onderzoek zullen de nieuwe ontwikkelingen getoetst worden op het gebied van betrouwbaarheid, zodat de nieuwe ANTAT-CU4 verder ontwikkeld kan worden en zodoende in de toekomst gebruikt kan worden om de verbale functionele communicatievaardigheden van cliënten met afasie te toetsen.

### Onderzoeksvragen en hypothesen

Om de nieuwe ontwikkelingen van de ANTAT-CU4 te toetsen, wordt bekeken in hoeverre de nieuwe scenario's en het nieuwe scoringsmodel van de ANTAT-CU4 betrouwbaar zijn. Hiervoor wordt de data van 31 Nederlandstalige proefpersonen zonder afasie gebruikt. Om de betrouwbaarheid van de ANTAT-CU4 in kaart te brengen, worden verschillende aspecten van de ANTAT-CU4 onderzocht. Zo wordt er gekeken naar de interne consistentie van de ANTAT-CU4, door de samenhang tussen de scenario's te analyseren. Daarnaast wordt onder andere gekeken naar de mate van de test-hertestbetrouwbaarheid van de ANTAT-CU4 op het gebied van verbale effectiviteit en efficiëntie bij gezonde sprekers. Eveneens worden de inter- en intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4 bij gezonde sprekers onderzocht, gemeten op basis van het aantal correcte informatie-eenheden, oftewel Content Units (CUs;

Ruiter et al., 2011 gebaseerd op Yorkston & Beukelman, 1980). De onderzoeksvragen, die voor dit onderzoek zijn opgesteld, staan hieronder weergegeven. Er wordt per onderzoeksvraag uitgelegd op welke manier een antwoord verkregen wordt. Tevens worden de hypothesen per onderzoeksvraag beschreven.

*Wat is de mate van de interne consistentie van de ANTAT-CU4?*

Voordat de test-hertestbetrouwbaarheid gemeten wordt, is het van belang om na te gaan wat de mate van interne consistentie van de ANTAT-CU4 is. Er wordt onderzocht in hoeverre de verschillende items van de ANTAT-CU4 hetzelfde kenmerk beogen te meten. De interne consistentie wordt onderzocht door middel van *reliability analyses* en uitgedrukt in Cronbachs alfa. De verwachting is dat de Cronbachs alfawaarde .80 of hoger is, wat volgens Van Wijk (2000) goed is. Bij een alfawaarde van .80 of hoger betekent dit dat de items van de ANTAT-CU4 hetzelfde kenmerk meten.

*Wat is de mate van de test-hertestbetrouwbaarheid van de ANTAT-CU4 bij gezonde sprekers?*

Om de mate van de test-hertestbetrouwbaarheid te onderzoeken, wordt nagegaan of de behaalde scores van 31 gezonde proefpersonen op de ANTAT-CU4 stabiel blijven over de tijd heen, ofwel of de deelnemers bij de tweede testafname ongeveer dezelfde resultaten zullen behalen als bij de eerste testafname. Om de test-hertestbetrouwbaarheid te toetsen wordt nagegaan of er een verschil is in de mate van verbale effectiviteit en verbale efficiëntie tussen de eerste en de tweede testafname.

Tijdens het onderzoek zal worden nagegaan of een verandering optreedt in de mate van effectiviteit, ofwel het percentage geproduceerde Content Units, wanneer de ANTAT-CU4 tweemaal wordt afgenomen. Het percentage geproduceerde Content Units op beide testafnamen zullen met elkaar vergeleken worden. Er wordt geen significant verschil verwacht.

Om de verschillen wat betreft de verbale efficiëntie, ofwel het gemiddeld aantal geproduceerde Content Units per minuut, na te gaan, zullen deze scores op beide toetsmomenten met elkaar vergeleken worden. Hierbij wordt verwacht dat de proefpersonen bij de tweede testafname gemiddeld meer Content Units per minuut zullen produceren dan bij de eerste testafname.

De verwachting is dus dat de verbale effectiviteit hetzelfde blijft, terwijl de verbale efficiëntie toeneemt.

*Wat is de mate van interbeoordelaarsbetrouwbaarheid van de ANTAT-CU4?*

Om de interbeoordelaarsbetrouwbaarheid te onderzoeken zullen verschillende beoordelaars punten toekennen aan de prestaties van zes proefpersonen zonder afasie op de eerste en tweede testafname. De gegeven scores zullen dan met elkaar vergeleken worden. De verwachting is dat de verschillende beoordelaars ongeveer gelijke puntentaallen zullen toekennen. Er wordt geen significant verschil verwacht.

*Wat is de mate van intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4?*

Om de intrabeoordelaarsbetrouwbaarheid te onderzoeken, zal de auteur van deze scriptie op twee momenten, met een tussentijd van ongeveer acht weken, de prestaties op de eerste testafname van 18 proefpersonen zonder afasie beoordelen. Hiermee wordt bekeken hoe consistent de beoordelaar is in het geven van CU-scores. Hier wordt geen significant verschil verwacht.



## **1.1 Indeling van de scriptie**

In hoofdstuk 2 “Theoretische achtergrond” wordt verdiepende achtergrondinformatie gegeven over de voorgaande onderzoeken en de ontwikkeling van de ANTAT (Blomert et al., 1995) tot de huidige ANTAT-CU4. De term ‘Content Units’ zal uitgebreider besproken worden en ook wordt aandacht besteed aan enkele componenten van betrouwbaarheid. In hoofdstuk 3 “Methode” zal ingegaan worden op de onderzoeksopzet en de uitvoering van het onderzoek. Aansluitend zullen in hoofdstuk 4 “Resultaten” de uitkomsten van het onderzoek weergegeven worden. In hoofdstuk 5 “Discussie/Conclusie” worden de resultaten van het onderzoek geïnterpreteerd. Tevens volgt een algemene conclusie met kritische terugblik op het huidige onderzoek en aanbevelingen voor toekomstig onderzoek.

## **2. Theoretische achtergrond**

In dit hoofdstuk wordt kort ingegaan op de ontwikkeling van de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (ANTAT; Blomert et al., 1995) tot de ANTAT-CU4 in zijn huidige vorm. Daarnaast worden verschillende termen behandeld, die van belang zijn bij het onderzoek.

### **2.1 Waarom is een nieuwe ANTAT nodig?**

In de inleiding werd kort beschreven waarom de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden (ANTAT; Blomert et al., 1995) onvoldoende geschikt is om de verbale communicatievaardigheden bij cliënten met afasie te onderzoeken en in kaart te brengen. Hieronder worden de punten nogmaals kort toegelicht.

Met de ANTAT kan, wat betreft de functionele communicatie, alleen een uitspraak gedaan worden over de verbale effectiviteit. Er wordt geen score berekend wat betreft de communicatieve efficiëntie (Ruiter et al., 2011). Hierdoor kan de functionele communicatie van de spreker onvoldoende in kaart gebracht worden (Yorkston & Beukelman, 1980; Nicholas & Brookshire, 1993; 1995). Daarnaast kan met de huidige 5-puntsschaal geen objectief oordeel gegeven worden over de communicatieve effectiviteit. De 5-puntsschaal wordt gebruikt om na te gaan in hoeverre een persoon met afasie te begrijpen is (hoe begrijpelijk is de boodschap, die de persoon overbrengt) en in hoeverre de persoon verstaanbaar is (Blomert et al., 1994). Dassek (2016) beschrijft dat een objectieve beoordeling met deze 5-puntsschaal niet mogelijk is, omdat alleen een beschrijving gegeven wordt en er geen telbare eenheden gebruikt worden. Volgens Stark en Viola (2006) is het voornaamste punt van kritiek dat Blomert et al. (1994) geen adequate beschrijving geven voor het interpreteren van de vijf classificaties.

Naast het onvoldoende in kaart kunnen brengen van de communicatieve efficiëntie en het ontbreken van een objectieve beoordelingsmaat blijkt de ANTAT in zijn huidige vorm volgens Aan de Stegge (2015) niet meer representatief te zijn voor deze generatie. Ze beschrijft dat bepaalde voorwerpen of handelingen, die in de tijd van de ontwikkeling van de ANTAT gewoon waren, in de huidige maatschappij zijn vervangen of veranderd. Verder heeft Filipinski (2014) aangetoond dat de scenario's niet meer voldoen aan de criteria, die door Blomert et al. (1994) waren opgesteld. Filipinski beschrijft dat de scenario's van de ANTAT tegenwoordig niet meer actueel, duidelijk en inleefbaar zijn. Een van de scenario's uit de ANTAT is als volgt: "*U bent op het station. U moet naar Zwolle. U staat voor het loket. Wat zegt u?*". In het verleden kon een treinkaartje gekocht worden aan het loket. Tegenwoordig maakt men gebruik van een OV-chipkaart of koopt men een kaartje via de automaat. Dit genoemde scenario is daarom niet (meer) actueel en mogelijk is het voor jongeren tevens minder goed inleefbaar, omdat zij een dergelijke situatie misschien niet meegemaakt hebben.

Samengevat zijn er drie redenen waarom het wenselijk is om de ANTAT (Blomert et al., 1995) aan te passen. Ten eerste kan met de ANTAT in zijn huidige vorm niet de volledige functionele vaardigheid van iemand in kaart gebracht worden. Ten tweede kan geen objectieve beoordeling gegeven worden over iemands verbale effectiviteit en als laatste zijn de scenario's niet meer van deze tijd.

### **2.2 Ontwikkeling van ANTAT naar ANTAT-CU4**

Voordat Ruiter et al. (2011) een kwantitatieve maat ontwikkelden voor de ANTAT (Blomert et al., 1995), heeft Dijkstra (2006) de responsen op de ANTAT van 24 proefpersonen geanalyseerd.

Hierbij werd gekeken naar welke informatie per scenario noodzakelijk was om de boodschap over te brengen. Dit was in lijn met de werkwijze van Blomert, Koster, Van Mier en Kean (1987). Zij benoemden dat per situatie, in dit geval per scenario van de ANTAT, onderscheid gemaakt kan worden tussen twee aspecten: (a) de noodzakelijke elementen, die essentieel zijn voor het begrijpelijk overbrengen van de boodschap en (b) sociaal gewenste elementen, die niet noodzakelijk zijn voor het overbrengen van de boodschap, maar die wel deel zijn van sociaal gewenst gedrag in een bepaalde situatie. De responsen van 24 gezonde proefpersonen werden door Dijkstra (2006) onderverdeeld in een aanleiding en een verzoek. De aanleiding is de reden waarom een conversatie gestart wordt. Het verzoek is het communicatieve doel dat de spreker wil bereiken. Voorbeeld (1) zal het verschil tussen aanleiding en verzoek verduidelijken. Het voorbeeld is een respons op scenario 8 “hond buurman”.

De instructie bij scenario 8 is als volgt: *“De hond van uw buurman blaft de hele dag. U heeft er echt genoeg van. U wilt het er met hem over hebben. Wat zegt u?”*. Een voorbeeld van een reactie van een deelnemer kan zijn: *“Nou buurman. Ik heb erg last van je hond, hij blaft de hele dag door. Kunnen we daar iets aan doen?”*. Wanneer deze reactie wordt gesplitst in een aanleiding en een verzoek, ziet dat er als volgt uit:

- (1) Aanleiding: *“Ik heb erg last van je hond, hij blaft de hele dag door.”*  
 Verzoek: *“Kunnen we daar iets aan doen?”*

De onderdelen aanleiding en verzoek kunnen nog verder uitgediept worden door middel van de kwantitatieve maat die Ruiter et al. (2011) hebben ontwikkeld, gebaseerd op de Content Unit Analyse (CU-analyse) van Yorkston en Beukelman (1980). Een Content Unit (CU) is een groep van informatie die door gezonde sprekers altijd als een eenheid wordt uitgesproken (Yorkston & Beukelman, 1980). Yorkston en Beukelman haalden hierbij een voorbeeld aan van een zin die een gezonde spreker zou kunnen produceren bij het beschrijven van een afbeelding uit de Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972). Op deze afbeelding probeert een jongen een koekje te pakken. De afbeelding is te vinden in Bijlage A. De geproduceerde zin is: *“The little boy is on the stool and reaching up for a cookie and he’s going to fall over”* (“De kleine jongen is op de kruk en reikt omhoog voor een koekje en hij gaat omvallen”). In voorbeeld (2) is nogmaals de zin te zien, nu met de Content Units onderstreept.

- (2) The little boy is on the stool and reaching up for a cookie and  
 he's going to fall over  
 1 2 3 4 5  
 6

De woorden ‘*on the stool*’ (op de kruk) werden door Yorkston en Beukelman gezien als een Content Unit, omdat in het taalgebruik van gezonde sprekers het woord ‘*on*’ (op) niet voorkwam zonder ‘*the stool*’ (de kruk). De woorden ‘*little*’ (kleine) en ‘*boy*’ (jongen) werden gezien als afzonderlijke Content Units, omdat bij veel van de taalsamples van gezonde sprekers het concept ‘*boy*’ voorkwam, zonder ‘*little*’. In paragraaf 2.5 wordt uitgebreid ingegaan op de CU-analyse. Een kwantitatieve beoordelingsmaat, zoals de CU-analyse, is volgens Wolthuis et al. (2014) betrouwbaarder dan een kwalitatieve beoordelingsmaat, zoals de 5-puntsschaal. Daarnaast kan

men met een kwantitatieve analyse nauwkeuriger veranderingen vaststellen dan met een kwalitatieve analyse (Grande et al., 2008). Een kwantitatieve beoordelingsmaat is daarom volgens Ruiter et al. (2011) mogelijk ook gevoeliger als het gaat om het opsporen van veranderingen in de functionele communicatie over de tijd. Voor het ontwikkelen van een kwantitatieve beoordelingsmaat (CU-schema) hebben Ruiter et al. (2011) van 24 sprekers, zonder talige problemen, de responsen op de scenario's van de ANTAT (Blomert et al., 1995) getranscribeerd en informatie-eenheden onderverdeeld in categorieën. Hierbij werd een norm van 30% gehanteerd. Dat wil zeggen: wanneer 30% van de sprekers een bepaald CU noemden, werd dat CU gezien als belangrijk communicatief doel van het betreffende scenario. Ruiter et al. (2011) maakten evenals Dijkstra (2006) onderscheid tussen aanleiding en verzoek. Het scoringsschema van Ruiter et al. (2011) is daarna verder ontwikkeld in verschillende vervolgonderzoeken (o.a. Filipinski, 2014; Aan de Stegge, 2015; Giessen, 2015). Dassek (2016) heeft naar aanleiding van onderzoek van Aan de Stegge en Giessen een groot aantal scenario's van de ANTAT (Blomert et al., 1995) aangepast of vervangen. Deze nieuwe versie wordt ANTAT-CU4 genoemd. Voor de ANTAT-CU4 heeft Dassek tevens het CU-scoringsschema aangepast. Ze heeft per aangepast scenario geanalyseerd hoeveel proefpersonen (n=60) een aanleiding en/of verzoek produceerden. Daarna onderzocht ze de geproduceerde Content Units. Op basis van drempelwaarden werd besloten om genoemde aanleidingen, verzoeken en Content Units al dan niet op te nemen in het CU-schema.

Ten tijde van de ontwikkeling van de ANTAT beschreven Blomert et al. (1994) dat de ANTAT een betrouwbaar en valide instrument is voor het meten van de taalvaardigheden bij personen met een hersenbeschadiging. Ze benoemden dat de test zowel voor diagnostisch onderzoek als evaluatieonderzoek gebruikt kan worden. Het doel van de ANTAT is volgens Blomert et al. (1995) het in kaart brengen van de functionele communicatievaardigheden van sprekers met afasie. Om de functionele communicatie te onderzoeken, dient gebruik gemaakt te worden van alledaagse situaties (ASHA, 1990, in Frattali, 1992; Blomert et al., 1995). Door het toevoegen van een kwantitatieve maat aan de ANTAT en het aanpassen van de scenario's, is het met de ANTAT-CU4 mogelijk om de functionele communicatie objectiever en mogelijk ook betrouwbaarder in kaart te brengen. Het is belangrijk dat de ANTAT-CU4, net als de originele ANTAT, een betrouwbaar en valide meetinstrument is waarmee de taalvaardigheden onderzocht kunnen worden.

### **2.3 Een kwantitatief scoringssysteem**

Zoals eerder benoemd, maakt de ANTAT (Blomert et al., 1995) gebruik van een subjectieve schaal, waarbij men uitspraken doet aan de hand van een beschrijving. Dit is een voorbeeld van een kwalitatieve scoringsmethode. Bij kwalitatief onderzoek, worden persoonlijke kwaliteiten van de onderzoeker, zoals waarneming en interpretatie, gericht ingezet om tot een conclusie te komen (Van Zwieten & Willems, 2004). Bij de ANTAT geeft de onderzoeker een oordeel over de begrijpelijkheid en verstaanbaarheid van iemand met afasie. Bij kwantitatief onderzoek spelen persoonlijke waarnemingen juist geen rol of is deze rol tot een minimum beperkt. Bij kwantitatief onderzoek wordt op basis van telbare eenheden, die als extern criterium dienen, een uitspraak gedaan over bepaalde taalvaardigheden (Boxum et al., 2013).

Volgens Blomert et al. (1990) leent de ANTAT zich goed voor het ontwikkelen van een kwantitatieve maat. Ze benoemden dat de verbale communicatie, gemeten met de ANTAT,

betrouwbaar kan worden uitgedrukt in kwantitatieve scores. Dit is een belangrijke constatering, want Yorkston en Beukelman (1977) gaven aan dat er een systeem nodig is voor het kwantificeren van de effectiviteit van verbale output bij personen met afasie. Daarnaast benoemden Yorkston en Beukelman dat, wanneer de verbale output gekwantificeerd wordt, het noodzakelijk is om een taak te vinden waarop gezonde sprekers relatief consistent en voorspelbaar scoren. Bij een onderzoek gericht op semi-spontane taal, waarbij gerichte vragen gesteld worden, zijn de antwoordmogelijkheden tot op een bepaalde hoogte beperkt. Hierdoor kan een goede vergelijking gemaakt worden tussen gezonde sprekers en sprekers met afasie (Yorkston & Beukelman, 1977). Tevens kan met instrumenten die kwantitatief de semi-spontane taal meten, zowel de verbale effectiviteit als de verbale efficiëntie gemeten worden (Yorkston & Beukelman, 1980; Nicholas & Brookshire, 1993; 1995).

Ruiter et al. (2011) hebben voor de ANTAT een systeem ontwikkeld waarmee het mogelijk is om op een kwantitatieve manier de semi-spontane taal te meten. Met het ANTAT-CU-schema ontwikkelden zij een kwantitatieve maat voor verbale effectiviteit en efficiëntie, waarbij ze uitgingen van essentiële informatie-eenheden die nodig zijn om het communicatieve doel van de ANTAT-scenario's te bereiken. Een kwantitatieve analyse van de (semi-)spontane spraak is een onmisbare methode om het herstel van de functionele taalvaardigheden objectief te beoordelen (Prins & Bastiaanse, 2004). Hierdoor krijgen klinici een vollediger beeld van de functionele taalvaardigheden van afatische sprekers (Ruiter et al., 2011; Prins & Bastiaanse, 2004).

## **2.4 Functionele communicatie**

Het huidige onderzoek bouwt voort op andere onderzoeken die gericht waren op het ontwikkelen van een kwantitatief scoringssysteem voor de ANTAT (Ruiter et al., 2011; Filipinski, 2014; Aan de Stegge, 2015; Giessen, 2015; Dassek, 2016). In deze voorgaande onderzoeken is de term functionele communicatie al eens gebruikt. Om deze reden wordt in het huidige onderzoek dezelfde definitie aangehouden, namelijk: functionele communicatie is "*the ability to receive a message or to convey a message, regardless of the mode, to communicate effectively and independently in a given environment*" (American Speech-Language-Hearing Association; ASHA; 1990, zoals geciteerd door Frattali, 1992, p. 64). Vrij vertaald houdt dit in dat functionele communicatie het vermogen is om een boodschap te begrijpen of produceren, ongeacht op welke manier, om effectief en onafhankelijk in een bepaalde omgeving te communiceren.

Om een uitspraak te kunnen doen over de functionele communicatie is het van belang om zowel de communicatieve effectiviteit als de communicatieve efficiëntie te beoordelen (Nicholas & Brookshire, 1995). Volgens Yorkston en Beukelman (1980) zijn de communicatieve effectiviteit en de communicatieve efficiëntie namelijk twee onderliggende dimensies van functionele communicatie. Daarnaast beschrijven Nicholas en Brookshire (1995) dat een spreker van de ene kant efficiënt kan zijn (een bepaalde hoeveelheid informatie in korte tijd kan overbrengen), maar niet de gewenste informatie over kan brengen. De spreker is dan niet effectief. Dit kan ook andersom voorkomen, namelijk dat een spreker wel de essentiële informatie over kan brengen, maar dat hij niet efficiënt is. Het is dus van belang om informatie te verkrijgen over beide aspecten van functionele communicatie.

Filipinski (2014) beschreef dat er in de literatuur door Nicholas & Brookshire (1995) en Ruiter et al. (2011) twee verschillende methoden benoemd worden, die gebruikt kunnen worden om de effectiviteit en efficiëntie van de functionele communicatie van sprekers met afasie te meten. Dit

zijn de “Content-Unit Analysis” (CU-Analysis) van Yorkston en Beukelman (1980) en de “Correct Information Unit-Analysis” (CIU-analysis) van Nicholas en Brookshire (1993). Bij beide analysemethoden wordt gebruikt gemaakt van kwantificeerbare eenheden. Hierdoor zijn beide methoden geschikt voor het meten van effectiviteit en efficiëntie (Filipinski, 2014).

In het huidige onderzoek, zal in lijn met voorgaande onderzoeken (Ruiter et al., 2011; Filipinski, 2014; Aan de Stegge, 2015; Giessen, 2015; Dassek, 2016) gewerkt worden met Content Units, gebaseerd op de Content-Unit Analyse van Yorkston en Beukelman (1980). Er zijn twee redenen waarom voor deze methode gekozen is. Ten eerste wordt er bij de CU-analyse onderscheid gemaakt tussen essentiële en relevante informatie-eenheden. Voor het scoren van informatie-eenheden wordt alleen die informatie, die essentieel is voor het begrijpen van de boodschap, meegerekend. Bij de CIU-analyse worden niet alleen essentiële informatie-eenheden meegeteld, maar ook de relevante informatie-eenheden. Relevante informatie heeft wel te maken met de boodschap en kan aanvullende informatie geven, maar is niet noodzakelijk voor het begrijpen van deze boodschap. Uiteindelijk is het de bedoeling om de ANTAT-CU4 als onderzoeksinstrument te gebruiken om na te gaan in hoeverre een spreker met afasie een intentie begrijpelijk over kan brengen. Het is dus van belang om te kijken naar de hoeveelheid essentiële informatie die een persoon met afasie over kan brengen. Ten tweede zien Yorkston en Beukelman (1980) een Content Unit als een groep informatie-eenheden, deze kan bestaan uit een enkel woord of een groep woorden, die door gezonde sprekers altijd als een eenheid worden uitgesproken. Nicholas en Brookshire (1993) beschouwen CIUs als losse woorden. Elk relevant, maar niet per se essentieel woord, wordt als een CIU gezien en gescoord. Voor het analyseren van de geproduceerde boodschap van de spreker met afasie is het, naar mijn mening, belangrijker om te kijken naar de hoeveelheid groepen essentiële van informatie een spreker met afasie over kan brengen, dan te kijken naar de hoeveelheid woorden omtrent het onderwerp hij produceert. Volgens Shewan (1988) blijken Content Units nuttig te zijn om te kunnen differentiëren tussen gezonde sprekers en sprekers met afasie. Daarnaast beschrijft Armstrong (2000) dat er metingen van verbale efficiëntie gedaan kunnen worden wanneer Content Units worden gekwantificeerd en aan een tijdsduur worden gerelateerd. Dit kan volgens Armstrong nuttig zijn bij onderzoek gericht op de inspanning die het vergt van de spreker om uitingen te produceren.

Nicholas en Brookshire (1993) merken op dat het meten van communicatieve vaardigheden in conversatie-omstandigheden mogelijk het meest ecologisch valide middel is om verbale communicatievaardigheden van volwassenen met afasie te bepalen. De ANTAT (Blomert et al., 1995) en de ANTAT-CU4 lokken taal uit door middel van scenario's, wat overeenkomsten vertoont met converseren. Zodoende lijken de ANTAT en de ANTAT-CU4 geschikte middelen om de functionele communicatie bij sprekers met afasie te onderzoeken.

## **2.5 Propositions en Content Units**

Om onze mentale toestand en gesproken uitingen te identificeren, classificeren en individualiseren gebruiken we proposities (Hanks, 2015). Propositions zijn een weerspiegeling van pre-verbale boodschappen. Ze staan voor de structuur van lexicale concepten (Ruiter et al., 2011). Turner en Greene (1977) benoemen dat proposities eenheden zijn die een relatie en argument (meestal subject of object) met elkaar verbinden, zodat het resultaat een enkele intentie is. Turner en Greene beschrijven dat er drie hoofdtypen van verbindende proposities zijn, die een relatie uitdrukken, namelijk predicaat, modificeerder en proposities die een connectie uitdrukken. Bij een

predicaat wordt de actie of staat uitgedrukt. Smolik et al. (2016) leggen dit uit aan de hand van het volgende voorbeeld. De zin “*John is running*” (John rent) bevat de predicatieve propositie *run* (John). Dit betekent dat John het onderwerp van het predicaat rennen is. Modificeerders omvatten relaties die een eigenschap toekennen aan het argument. Zo drukt de zin “*Peter is smart*” (Peter is slim) de propositie *smart* (Peter; slimme Peter) uit. Propositionen die een connectie uitdrukken, geven volgens Turner en Greene (1977) de samenhang tussen verschillende propositionen weer. Ze zorgen voor cohesie in de geproduceerde tekst. Propositionen bestaan dus eigenlijk uit de belangrijkste informatie-eenheden waarmee we onze gedachten en intenties uit kunnen drukken. Het gebruik van propositionen is volgens Hanks (2015) een voorbeeld van een strategie om objecten te beschrijven en classificeren. Op deze manier kunnen concrete en abstracte objecten en eigenschappen aan elkaar gekoppeld worden (Hanks, 2015). Ruiter et al. (2011) benoemen dat hoewel propositionen met woorden corresponderen, de propositionen in verschillende bewoordingen uitgedrukt kunnen worden. Stel dat de inhoud van de propositie ‘iemand's pad kruisen’ voorstelt. Hiervoor zouden onder andere de woorden ‘ontmoeten’, ‘treffen’ en ‘tegengekomen’ gebruikt kunnen worden om de situatie te beschrijven. Hoewel de woorden verschillend zijn, blijft de intentie gelijk, namelijk het beschrijven dat een persoon het pad van een andere persoon heeft gekruist.

Samengevat is een propositie een manier om ideeën uit te drukken. Om een idee uit te drukken op verbale wijze, gebruiken we Content Units. Een Content Unit (CU) is een groep van informatie die door gezonde sprekers altijd als een eenheid wordt uitgesproken (Yorkston & Beukelman, 1980). Een CU kan bestaan uit een woord (‘nieuw’, ‘shirt’) of een frase (‘op de tafel’) (Yorkston & Beukelman, 1980; Ruiter et al., 2011). Synoniemen van een woord (bijvoorbeeld “trap” en “ladder”) vallen in dezelfde CU-categorie (Ruiter et al., 2011).

Zoals hierboven is besproken, is in het ANTAT-CU4-schema een onderverdeling gemaakt in ‘aanleiding’ en ‘verzoek’. Bij het vragen naar een reactie op een scenario, kan de spreker een aanleiding voor het gesprek geven. Hierbij geeft de spreker aan wat de reden is voor het gesprek. De propositie is in dit geval de motivatie voor het aangaan van de conversatie (aanleiding). Voor het verbaliseren van deze propositie zal gebruik gemaakt worden van een of meerdere Content Units (CUs). Hetzelfde geldt voor het onderdeel ‘verzoek’. Bij het verzoek wordt door de spreker benoemd wat hij met het gesprek wil bereiken. De propositie, of achterliggende gedachte, is in dit geval het beoogde gespreksdoel (verzoek). Om dit te verbaliseren, kan de spreker evenals bij de aanleiding gebruik maken van een of meerdere Content Units. Eerder werd bij voorbeeld (1) scenario 8 aangehaald, waarbij de respons van een spreker werd onderverdeeld in een aanleiding en een verzoek. Wanneer deze reactie verder onderverdeeld wordt in Content Units, ziet dat er als volgt uit:

- |                 |  |
|-----------------|--|
| (3) Aanleiding: | “ <u>Ik heb erg last</u> van je hond, hij blaft de hele dag door.” |
| CU's:           | ik – last – hebben   |
| <br>            |  |
| Verzoek:        | “Kunnen <u>we daar iets aan doen</u> ?”                            |
| CU's:           | we – iets – aan doen   |

In voorbeeld (3) is te zien dat de aanleiding bestaat uit drie Content Units (ik – last- hebben), waarmee de spreker de reden voor het aangaan van het gesprek benoemt. Het verzoek, het

beoogde gespreksdoel, bestaat eveneens uit drie Content Units (we – iets – aan doen). Een propositie kan op verschillende manieren tot uiting komen. Dit is ook te zien in de respons van de spreker bij scenario 8. In de aanleiding geeft de spreker niet alleen aan dat hij last heeft, maar ook waarvan hij last heeft, namelijk het blaffen van de hond. De Content Units die hierbij van belang zijn, zijn te zien in voorbeeld (4).

(4) Aanleiding: *“Ik heb erg last van je hond, hij blaft de hele dag door.”*  
CUs: hond – blaffen – hele dag

In dit geval benoemt de spreker beide propositieformuleringen van mogelijke aanleidingen voor het gesprek, maar ook wanneer er voor een van deze twee formuleringen was gekozen, zou de aanleiding nog steeds de essentiële informatie-eenheden bevatten, die de luisteraar nodig heeft om de boodschap van de spreker te begrijpen. In dit geval blijft het aantal Content Units gelijk, beide formuleringen bestaan uit drie Content Units. Bij enkele scenario's waarbij verschillende formuleringen mogelijk zijn, is dit niet het geval. Bijvoorbeeld bij scenario 15 “huisarts”, waarbij bij het verzoek verschillende uitingvormen mogelijk zijn, hangt het aantal Content Units af van de gekozen formulering. De instructie van scenario 15 is als volgt: *“Er is bij u een kind gevallen. Het ziet er niet zo best uit. U belt naar de huisarts en zegt?”*.

Na het geven van de aanleiding voor het gesprek, kan het verzoek bestaan uit verschillende formuleringen.

(5) Verzoek 1: *“Kan ik met het kind naar u komen? ” / “Kunt u komen?”*  
CUs: ik/u – komen

Verzoek 2: *“Kunt u naar het kind kijken?”*  
CUs: u – ernaar - kijken

In voorbeeld (5) is te zien dat de gekozen formulering invloed heeft op het aantal Content Units. Dit is ook terug te zien in Figuur 1, waarin het ANTAT-CU4-schema (naar voorbeeld van Dassek, 2016) van scenario 15 “huisarts” is weergegeven. Bij het scoren van de ANTAT-CU4 is het van belang rekening te houden met variabele maximale scores, omdat bij het berekenen van de effectiviteit het totale aantal behaalde CU-scores gedeeld wordt door het aantal Content Units die de spreker maximaal had kunnen behalen. Wanneer niet het juiste aantal maximaal haalbare Content Units wordt genoteerd, kan het zijn dat de uiteindelijke testresultaten niet overeenkomen met de verbale communicatievaardigheden van de spreker.

Stel dat spreker A alle Content Units (3 stuks) van de aanleiding correct benoemde en alle Content Units van verzoek 1 (2 stuks) correct benoemde. Dan zag de berekening er als volgt uit:  $5 \text{ behaalde CUs} / 5 \text{ maximaal haalbare CUs} * 100 = 100\%$ .

Stel dat spreker B alle Content Units (3 stuks) van de aanleiding correct benoemde en twee van de drie Content Units van verzoek 2 correct benoemde, dan zag de berekening er als volgt uit:  $5 \text{ behaalde CUs} / 6 \text{ maximaal haalbare CUs} * 100 = 83.33\%$ . Hoewel spreker A en B beiden vijf Content Units correct benoemden, was spreker A effectiever, omdat hij alle essentiële elementen benoemde die nodig waren voor het overbrengen van de geformuleerde boodschap.



Scenario 15: Huisarts							Score	5/6
<b>A</b>	90 <sup>1</sup>	91	92	EN V	93	94	95	
A1	kind	gevallen	toelichting X	V1	ik, u	-	komen	
				OF V2	u	ernaar, advies	kijken, geven	
<p>Opmerkingen: Bij dit scenario kunnen maximaal 5 of 6 punten behaald worden afhankelijk van het gebruikte verzoek. Verzoek 1 = maximaal 2 CUs, Verzoek 2 = maximaal 3 CUs.  A1: mogelijke aanleiding, die de spreker kan formuleren  V1/V2: mogelijke verzoeken, die de spreker kan formuleren  <sup>1</sup> nummer van Content Unit</p>								

Figuur 1. ANTAT-CU4-schema behorend bij scenario 15 "huisarts", naar voorbeeld van Dassek (2016).

## 2.6 Verbale effectiviteit (% Content Units)

De term (non-)verbale effectiviteit zegt iets over de hoeveelheid essentiële informatie die de spreker overbrengt op de luisteraar (Yorkston & Beukelman, 1980; Nicholas & Brookshire, 1993, 1995; Blomert et al., 1994; Ruiters et al., 2011; Ruiters et al., 2015). Hiermee wordt dus bedoeld hoe goed een spreker de essentiële informatie over kan brengen, die de luisteraar nodig heeft om de intentie van de spreker te begrijpen (Blomert et al., 1994; Ruiters et al., 2015). Blomert et al. (1994) benoemden daarnaast ook dat het overbrengen van deze essentiële informatie onafhankelijk is van de linguïstische vorm, zolang de luisteraar de boodschap maar begrijpt. Om de verbale effectiviteit te kwantificeren, kunnen deze informatie-eenheden geteld en gescoord worden. Met het ANTAT-CU4-schema kan onderzocht worden hoeveel essentiële informatie-eenheden een spreker produceert. In lijn met voorgaande onderzoeken (o.a. Filipinski, 2014; Ruiters et al., 2015) wordt met de verbale effectiviteit het percentage geproduceerde Content Units bedoeld.

## 2.7 Verbale efficiëntie (aantal Content Units per minuut)

Om een volledig beeld te krijgen van de verbale functionele communicatie is het van belang om naast het meten van de verbale effectiviteit ook de verbale efficiëntie te meten (Nicholas & Brookshire, 1995). Effectiviteit en efficiëntie zijn samen namelijk de belangrijkste eigenschappen van functionele communicatie (Ruiters et al. 2011).

Met verbale efficiëntie wordt bedoeld hoeveel essentiële informatie in een bepaalde tijd overgebracht wordt (Yorkston & Beukelman 1980; Nicholas & Brookshire, 1993; Blomert et al., 1994, Ruiters et al., 2011; Ruiters et al., 2015). Er wordt dus bekeken hoelang het duurt om het communicatieve doel te bereiken (Ruiters et al., 2015). De snelheid van informatieoverdracht kan op verschillende manieren gemeten worden. Zo benoemden Yorkston en Beukelman (1977) dat efficiëntie gedefinieerd kan worden als het aantal concepten of het aantal syllaben per minuut. Door het aantal syllaben per minuut te berekenen kan echter alleen iets gezegd worden over de spreeknelheid, niet over de snelheid van informatieoverdracht. Door het aantal concepten per minuut in kaart te brengen kan wel iets gezegd worden over de snelheid van informatieoverdracht in het algemeen, maar onder concepten vallen zowel de noodzakelijke (essentiële) als relevante elementen. Er kan dus niets gezegd worden over de snelheid waarmee essentiële informatie

overgebracht wordt. Op deze manier kan geen uitspraak gedaan worden over de verbale efficiëntie.

Nicholas en Brookshire (1993) poogden de spreesnelheid te berekenen aan de hand van het aantal woorden en aantal *Correct Information Units* (relevante informatie-eenheden; CIUs) per minuut. Net als bij het in kaart brengen van het aantal syllaben en concepten per minuut, kan met het nagaan van het aantal woorden of aantal CIUs per minuut geen uitspraak gedaan worden over de verbale efficiëntie. Het aantal woorden per minuut geeft net als het aantal syllaben per minuut alleen informatie over de spreesnelheid. Door het aantal CIUs per minuut te berekenen kan wel iets gezegd worden over de snelheid van relevante informatieoverdracht, maar niets over de snelheid van essentiële informatieoverdracht. Dit laatste is juist wat we willen meten met het huidige onderzoek. Om deze reden wordt in het huidige onderzoek de verbale efficiëntie gemeten aan de hand van het aantal Content Units per minuut, zoals beschreven in het onderzoek van Yorkston en Beukelman (1980). Dit is tevens in lijn met eerder onderzoek (Ruiter et al., 2015; Ruiter et al., 2011) waarbij verbale efficiëntie omschreven wordt als het aantal geproduceerde essentiële informatie-eenheden, oftewel het aantal Content Units per minuut.

## **2.8 Componenten van betrouwbaarheid**

In paragraaf 2.2 werden de stappen beschreven, die zijn doorlopen om de ANTAT-CU4 te ontwikkelen. Bij het aanpassen van de ANTAT (Blomert et al., 1995) naar de ANTAT-CU4 is het van belang dat de ANTAT-CU4 uiteindelijk een betrouwbare en valide test is, die in de praktijk gebruikt kan worden. Volgens Boeije (2014) zijn betrouwbaarheid en validiteit de klassieke manieren om te streven naar objectiviteit bij onderzoek. Met validiteit wordt bedoeld of het onderzoeksinstrument daadwerkelijk meet wat het beoogt te meten (Field, 2013). Betrouwbaarheid geeft weer of het onderzoeksinstrument consistent kan worden geïnterpreteerd in verschillende situaties (Field, 2013). Hieronder wordt uitgebreider ingegaan op deze termen en in hoeverre de ANTAT (Blomert et al., 1995) hieraan voldoet. Ook wordt een link gelegd naar de mate van validiteit en betrouwbaarheid van de ANTAT-CU4.

### **2.8.1 Validiteit**

Een valide test zorgt ervoor dat passende en betekenisvolle conclusies getrokken kunnen worden naar aanleiding van het onderzoek (Van der Meulen, Sandt-Koenderman, Duivenvoorden, & Ribbers, 2010). Het gaat er bij validiteit om dat we daadwerkelijk meten wat we willen meten (Field, 2013; Boeije, 2014). Met de ANTAT (Blomert et al., 1995) wordt gepoogd de taalvaardigheden van een persoon met afasie in alledaagse situaties te meten. Om dit na te gaan, zou het wellicht het beste zijn om een persoon een dag te volgen en zodoende te zien wat voor soort situaties de cliënt meemaakt en hoe hij hierop reageert. Dit is in de praktijk echter niet haalbaar, maar ook als dit wel mogelijk zou zijn, is het geen methode die de voorkeur heeft. In praktijksituaties kan een cliënt bepaalde uitingen ontwijken. Hij wordt immers niet gedwongen om een bepaalde taaluiting te produceren. Op deze manier krijgt men geen realistisch beeld van de verbale taalvaardigheden die de spreker met afasie heeft. Daarnaast kan de persoon met afasie gebruik maken van hulpmiddelen en compensatie- en aanpassingstechnieken om de communicatie te vergemakkelijken. Zo kan een cliënt bijvoorbeeld gebaren maken of dingen aanwijzen. Er wordt dan geen onderzoek gedaan naar de verbale taalvaardigheden, maar naar de communicatievaardigheden in het algemeen.

De ANTAT is een meetinstrument waarmee zo goed mogelijk een echt voorkomende situatie wordt nagebootst. Blomert et al. (1994) beschrijven dat de ANTAT bestaat uit korte communicatieve situaties die in het dagelijks leven voor zouden kunnen komen en daardoor realistisch zijn. Met de situaties wordt een monoloog uitgelokt. Dit komt niet volledig overeen met spontane taal, die wenselijk is voor een spontane taalanalyse, maar volgens Blomert et al. (1994) zou een echte dialoog tussen de testleider en spreker ervoor zorgen dat het een stuk moeilijker is om een standaardprocedure te verkrijgen voor het beoordelen van de verbale taalvaardigheden. Bij validiteit gaat het volgens Boeije (2014) voornamelijk over de interpretatie van de waarnemingen en over beïnvloeding van het onderzoek door systematische fouten. Een voorbeeld van een systematische fout is dat onderzoekers de verkregen data bewust of onbewust in een bepaalde richting interpreteren, die niet overeenkomt met de verkregen informatie (Boeije, 2014). Wanneer je bijvoorbeeld een te korte liniaal gebruikt, krijg je bij alle metingen een verkeerd resultaat, terwijl de meting zelf wel goed uitgevoerd is. Om ervoor te zorgen dat de onderzoeksgegevens juist geïnterpreteerd kunnen worden, is het van belang dat een onderzoeksinstrument valide is. Er zijn verschillende soorten validiteit, waaronder 'inhoudsvaliditeit', 'begripsvaliditeit' en 'criteriumvaliditeit'. Deze verschillende vormen worden hieronder besproken.

### **2.8.1.1 Inhoudsvaliditeit**

Met de term inhoudsvaliditeit wordt bedoeld in hoeverre de individuele onderdelen van de test het volledige toetsinstrument vertegenwoordigen (Patterson & Chapey, 2008; Field, 2013). In dit geval wordt bekeken in hoeverre de testscenario's van de ANTAT (Blomert et al., 1995) representatief zijn voor alledaagse situaties. Om inhoudsvalide te zijn, moeten de scenario's aan een aantal eisen voldoen. Blomert et al. (1994) hebben zelf een aantal eisen beschreven waar de ANTAT aan moet voldoen. Ten eerste moeten de scenario's aannemelijk en voorstelbaar zijn. Dit houdt in dat de scenario's bij iedere cliënt daadwerkelijk in het dagelijkse leven voor zouden kunnen komen, zodat de cliënt zich gemakkelijk kan inleven in de situatie. Ten tweede mogen de scenario's niet te belastend zijn voor het werkgeheugen. Blomert et al. (1994) benoemen dat mensen met afasie, naast taalproblemen, vaak ook andere cognitieve problemen hebben, die mogelijk de testafname kunnen verstoren. Door scenario's te gebruiken, die de geheugencapaciteit zo min mogelijk te belasten, kan de verbale communicatie beter onderzocht worden, omdat de cliënt minder vermoeid wordt. Als laatste geven Blomert et al. (1994) aan dat de voorwerpen die gebruikt worden, om de scenario's te ondersteunen, bekend moeten zijn voor de cliënt. Ook moeten de materialen van nature eenvoudig zijn en door de cliënt te manipuleren zijn. Dit betekent dat het materiaal makkelijk te herkennen moet zijn en eenvoudig in het gebruik. Toen de ANTAT ontwikkeld werd, voldeden de scenario's en bijbehorende materialen aan deze genoemde eisen. Recent heeft Filipinski (2014) laten zien dat de meeste scenario's in de huidige maatschappij niet meer voldoen aan een of meerdere van deze eisen. Om deze reden is een groot aantal scenario's aangepast of vervangen (Aan de Stegge, 2015; Giessen, 2015; Dassek, 2016). Deze scenario's voldoen beter aan de drie eisen, die Blomert et al. (1994) hebben opgesteld. Hiermee wordt de inhoudsvaliditeit van de ANTAT-CU4 gewaarborgd.

### **2.8.1.2 Begripsvaliditeit**

De term begripsvaliditeit verwijst naar de mate waarin een onderzoeksinstrument meet wat het beoogt te meten (Van der Meulen et al., 2010) in vergelijking met andere soortgelijke onderzoeksinstrumenten (Patterson & Chapey, 2008). Het heeft dus te maken met de mate waarin een test een bepaald theoretisch begrip meet (Tilburg University, 2017). Bij het analyseren van de begripsvaliditeit worden onderliggende kenmerken van het onderzoeksinstrument geïdentificeerd. Er wordt nagegaan of deze kenmerken overeenkomen met het construct waaraan de test volgens eisen moet voldoen (Van der Meulen et al., 2010). Door te vergelijken met andere onderzoeksinstrumenten, die dezelfde (deel)aspecten trachten te meten, kan de begripsvaliditeit onderzocht worden. De oorspronkelijke ANTAT (Blomert et al., 1995) beoogt de verbale communicatie van mensen met afasie te meten. Dit wordt gedaan door de begrijpelijkheid en verstaanbaarheid van de gesproken boodschap te beoordelen.

De begripsvaliditeit van de ANTAT (Blomert et al., 1995) werd op vier manieren onderzocht. Zo heeft Blomert bijvoorbeeld een Afasie Partner Vragenlijst (APV; 1993, zoals beschreven in Blomert, Koster, & Kean, 1995) ontwikkeld. Op deze manier werd een meetinstrument ontwikkeld dat net als de ANTAT beoogt inzage te krijgen in de communicatieve vaardigheden en konden de uitkomsten van de ANTAT en de APV naast elkaar gelegd worden om de overeenkomsten na te gaan. Daarnaast werd het verbale communicatieniveau van de cliënten vergeleken met de verschillende afasiesyndromen. Ook werd de relatie tussen communicatieniveau en type afasie onderzocht. Als laatste werd de relatie tussen begrijpelijkheid, verstaanbaarheid en verbale communicatie onderzocht.

Voor de ANTAT-CU4 zijn ook verschillende mogelijkheden om de begripsvaliditeit te onderzoeken en te waarborgen. Zo beschrijft Dassek (2016) dat de begripsvaliditeit door Ruiters et al. (2011) al positief verbeterd is door de kwalitatieve beoordelingsmethode van Blomert et al. (1994) te vervangen door een kwantitatieve beoordelingsmethode. Daarnaast kan de ANTAT-CU4 ook vergeleken worden met andere onderzoeksinstrumenten die hetzelfde beogen te meten. Hiervoor zou een vragenlijst als de APV (Blomert, 1993) gebruikt kunnen worden of het semigestructureerde interview van de AAT (Graetz et al., 1992). Deze beide onderzoeksinstrumenten zijn echter kwalitatief van aard en de vaardigheden worden beoordeeld op een 5-puntsschaal (APV) of 6-puntsschaal (semigestructureerd interview AAT). Hierdoor kunnen de verkregen scores niet direct met elkaar vergeleken worden. Om toch een vergelijking te kunnen maken tussen een van deze meetinstrumenten en de ANTAT-CU4, kunnen de behaalde scores op de ANTAT-CU4 mogelijk omgezet worden naar ordinale variabelen. Tevens zou gebruik gemaakt kunnen worden van de Scenario Test (van der Meulen et al., 2008). Hierbij moet wel in de gaten gehouden worden dat met de Scenario Test niet alleen de verbale, maar ook de non-verbale communicatievaardigheden gemeten worden. Tot op heden is de begripsvaliditeit van de ANTAT-CU4 niet onderzocht, omdat de test nog in ontwikkeling is. Wanneer het onderzoeksinstrument gereed is, is het noodzakelijk om de begripsvaliditeit te onderzoeken.

### **2.8.1.3 Criteriumvaliditeit**

De term criteriumvaliditeit gaat evenals begripsvaliditeit over de vraag of het onderzoeksinstrument meet wat het beoogt te meten (Field, 2013). Het verschil met

begripsvaliditeit is dat bij criteriumvaliditeit niet alleen wordt nagegaan of een instrument meet wat hij beweert te meten, maar wordt dit onderzocht door het te vergelijken met objectieve criteria (Field, 2013). Het uitgangspunt van de ANTAT (Blomert et al., 1995) is dat de score van een cliënt gebruikt kan worden als een maatstaf voor zijn communicatieve verbale vaardigheden in het dagelijks leven. Om te onderzoeken of de ANTAT inderdaad als maatstaf gebruikt kon worden, werden delen van de ANTAT vergeleken met de AAT (Graetz et al., 1992). De 'Begrijpelijkheid' van de ANTAT werd vergeleken met de Spontane Spraakschaal 'Communicatief gedrag' van de AAT. De correlaties bedroegen respectievelijk .81 voor de ANTAT I en .83 voor de ANTAT II, de twee parallelle testversies van de ANTAT. Hieruit bleek volgens Blomert et al. (1994) dat het construct 'verbale communicatie' een hoge validiteit heeft.

Verder werd de ANTAT (Blomert et al., 1995) door de Commissie Testaangelegenheden Nederland (COTAN; commissie die in Nederland de kwaliteit van testen beoordeelt) als voldoende beoordeeld wat betreft de criteriumvaliditeit. De COTAN beoordeelde de ANTAT dus als voldoende valide aangaande het meten van verbale communicatieve vaardigheden.

Met de ANTAT-CU wordt ook beoogd de verbale functionele communicatie te meten. Ruiters et al. (2011) suggereren dat, bij een kleine groep sprekers met afasie (n=10), de kwantitatieve scoringsmethode de verbale effectiviteit en efficiëntie kan meten. Het is zeer waarschijnlijk dat het ANTAT-CU4-schema ook bij een grotere onderzoeksgroep gebruikt kan worden om de verbale effectiviteit en efficiëntie te meten. Als dat mogelijk is, zou de nieuwe ANTAT-CU4 een valide meetinstrument zijn.

## **2.8.2 Betrouwbaarheid**

Hoewel validiteit een noodzakelijk goed is voor een testinstrument, is een hoge mate van validiteit alleen niet voldoende (Field, 2013). Om valide te zijn, moet een testinstrument volgens Field (2013) op de eerste plaats betrouwbaar zijn. Volgens Boeije (2014) betekent betrouwbaarheid dat de waarnemingen die gedaan worden, zo min mogelijk worden beïnvloed door toevallige of niet-systematische fouten. Hoe betrouwbaarder een testinstrument is, hoe minder meetfouten erin voorkomen. Betrouwbaarheidsanalyses worden gebruikt om de consistentie van een meting na te gaan. Betrouwbaarheid verwijst volgens Tompkins, Gibbs-Scott en Sharp (2008) naar de stabiliteit, consistentie en herhaalbaarheid van resultaten. In het huidige onderzoek wordt ook gekeken naar de interne consistentie van de ANTAT-CU4. Wanneer gesproken wordt over de betrouwbaarheid van de ANTAT-CU4 wordt hiermee de interne consistentie van de test bedoeld. Voordat een bepaalde test bij cliënten met een stoornis gebruikt kan worden, moet eerst bij gezonde proefpersonen worden nagegaan of een testinstrument betrouwbaar is (Potkamp-van den Belt, 2011). Er zijn verschillende soorten betrouwbaarheid, die hiervoor gebruikt kunnen worden, waaronder test-hertestbetrouwbaarheid en interbeoordelaarsbetrouwbaarheid. Deze vormen van betrouwbaarheid worden hieronder uitgebreid besproken.

### **2.8.2.1 Test-hertestbetrouwbaarheid**

Field (2013) omschrijft test-hertestbetrouwbaarheid als het vermogen om consistente resultaten te meten wanneer dezelfde entiteiten op twee verschillende momenten gemeten worden. Dit houdt in dat dezelfde groep personen op twee momenten in de tijd getoetst wordt over hetzelfde onderwerp. Als het testinstrument betrouwbaar is, zal de groep, op gelijke items op beide momenten in de tijd, (bijna) dezelfde scores behalen (Field, 2013).

Volgens Potkamp-van den Belt (2011) heeft een test bij een hoge betrouwbaarheid een goede kwaliteit. Hierdoor kan verantwoord een conclusie getrokken worden uit de behaalde testresultaten, mits een test daarnaast ook valide is. De mate van test-hertestbetrouwbaarheid van een meetinstrument wordt weergegeven door een correlatiecoëfficiënt. De samenhang tussen de behaalde scores op de verschillende meetmomenten vormt een graadmeter van de teststabiliteit (Baarda & De Goede, 2001). De test-hertestbetrouwbaarheid kan het beste gemeten worden met een groep proefpersonen waarvan de communicatieve taalvaardigheden relatief stabiel zijn. Daarnaast is het van belang dat de te meten eigenschappen van de groep proefpersonen niet veranderen over de tijd heen (Wright, Capilouto, Wagovich, Cranfill, & Davis, 2005; Potkamp-van den Belt, 2011). In het geval van de ANTAT-CU4 betekent dit dat de taalvaardigheden van de proefpersonen tussen de eerste en tweede afname niet veranderd (verbeterd of verslechterd) mogen zijn. Een groep proefpersonen zou kunnen bestaan uit gezonde proefpersonen of een groep mensen met afasie, die in de chronische fase zitten en tussen de meetmomenten in geen therapie krijgen.

Blomert et al., (1994) geven aan dat de test-hertestbetrouwbaarheid van de ANTAT (Blomert et al., 1995) goed is. Ze hebben dit onderzocht door 30 chronische afasiepatiënten tweemaal te testen met een tussenperiode van drie maanden. De resultaten van beide afnamen werden met elkaar vergeleken. Tussen beide momenten werd geen significant verschil gevonden wat betreft het patroon (factoranalyse) en niveau (t-toets) van de testcores (Blomert et al., 1995). Dassek (2016) heeft met zes gezonde proefpersonen de test-hertestbetrouwbaarheid voor de ANTAT-CU4 berekend. Hiervoor heeft ze gebruik gemaakt van gepaarde t-toetsen om de gemiddelde behaalde scores op de testafnamen met elkaar te vergelijken. Uit het onderzoek van Dassek (2016) kwam naar voren dat er geen significante verschillen waren in de scores over de tijd heen. De ANTAT-CU4 lijkt hiermee stabiel te zijn. In het huidige onderzoek wordt nogmaals de test-hertestbetrouwbaarheid van de ANTAT-CU4 onderzocht, maar bij een grotere onderzoeksgroep.

### **2.8.2.2 Begrip van testinstructie**

De ANTAT-CU4 bestaat uit verschillende scenario's waarin de spreker een bepaalde boodschap moet overbrengen naar de luisteraar. Het is daarom van belang dat voor de spreker duidelijk is welke boodschap hij over moet brengen. De spreker krijgt om deze reden bij elk scenario een instructie horen. Dassek (2016) gaf aan dat de instructies bij drie scenario's door de proefpersonen mogelijk minder goed werden begrepen. Het gaat hierbij om scenario 11 "bril", scenario 16 "gips" en scenario 20 "gemeente". Om deze reden wordt in het huidige onderzoek een tweede versie van deze scenario's opgenomen in de ANTAT-CU4. Doordat bij de proefpersonen beide versies worden afgenomen, kan worden nagegaan welke versie van de scenario's beter wordt begrepen door de proefpersonen en of de behaalde scores op de scenarioversies van elkaar afwijken.

### **2.8.2.3 Interbeoordelaarsbetrouwbaarheid**

De interbeoordelaarsbetrouwbaarheid geeft net als de test-hertestbetrouwbaarheid informatie over de kwaliteit van een testinstrument (Potkamp-van den Belt, 2011). Interbeoordelaarsbetrouwbaarheid is de mate van overeenstemming tussen de beoordeling van de uitkomsten door verschillende onderzoekers (Appels & Ponds, 2012). De mate van interbeoordelaarsbetrouwbaarheid kan onder andere worden weergegeven in Cohens kappa en

Cronbachs alfa. De mate van interbeoordelaarsbetrouwbaarheid van de ANTAT (Blomert et al., 1995) werd door Blomert et al. (1994) onderzocht. Zij vroegen zes onafhankelijke beoordelaars om de prestaties van 14 cliënten op de test te beoordelen. Deze zes beoordelaars waren geen deskundigen op het gebied van afasie, maar waren wel bekend met het scoringssysteem van de ANTAT. Ze moesten de taaluitingen scoren aan de hand van een 5-puntsschaal op begrijpelijkheid en verstaanbaarheid. Na het analyseren van de verkregen data volgens Krippendorff (1970), bleek dat het niveau van de interbeoordelaarsbetrouwbaarheid van de ANTAT voldoende is. De betrouwbaarheidswaarden bedroegen respectievelijk .92 en .94 voor de twee parallelle testversies op het onderdeel 'Begrijpelijkheid' (goed; Van Wijk, 2000) en .72 en .70 (adequaat; Van Wijk, 2000) voor de twee parallelle testversies op het onderdeel 'Verstaanbaarheid'.

Wat betreft de ANTAT-CU4 kunnen nog geen uitspraken gedaan worden over de interbeoordelaarsbetrouwbaarheid. In voorgaand onderzoek heeft Dassek (2016) voor zes proefpersonen de CU-scores uitgewerkt. In het huidige onderzoek zullen deze resultaten vergeleken worden met de scores die onafhankelijk van die van Dassek (2016) gegeven zijn. Doordat de kwalitatieve beoordelingsschaal is vervangen door een kwantitatieve beoordelingsschaal, door gebruik te maken van Content Units, is het mogelijk dat de interbeoordelaarsbetrouwbaarheid van de ANTAT-CU4 net als de ANTAT (Blomert et al., 1995) voldoende is.

#### **2.8.2.4 Intrabeoordelaarsbetrouwbaarheid**

Intrabeoordelaarsbetrouwbaarheid is de mate van overeenstemming tussen meerdere herhaalde metingen van een beoordelaar (Rousson, Gasser, & Seifert, 2002). Hierbij wordt nagegaan hoe groot de overeenkomst is tussen de uitkomsten van herhaalde metingen. De mate van intrabeoordelaarsbetrouwbaarheid kan net als de interbeoordelaarsbetrouwbaarheid onder andere worden weergegeven in Cohens kappa of Cronbachs alfa. Bij het ontwikkelen en testen van de ANTAT (Blomert et al., 1995) werd de intrabeoordelaarsbetrouwbaarheid niet onderzocht. In dit onderzoek zal de intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4 wel onderzocht worden. Het doel is om de ANTAT-CU4 uiteindelijk in de praktijk te gaan gebruiken. De therapeut die de ANTAT-CU4 afneemt, zal dan voor en na de interventie waarschijnlijk dezelfde beoordelaar zijn. Het is dus van belang dat de therapeut consistent scoort om meetfouten te voorkomen.

#### **2.8.3 Sensitiviteit en specificiteit**

Sensitiviteit en specificiteit zeggen iets over hoe accuraat een testinstrument is. Volgens Dollaghan (2007) betreft sensitiviteit alleen de mate waarin personen, die een stoornis of aandoening hebben als zodanig geïdentificeerd worden. Specificiteit betreft alleen de mate waarin personen, die geen stoornis of aandoening hebben als zodanig geïdentificeerd worden. Sensitiviteit kan worden gedefinieerd als het aantal terecht geïdentificeerde positieven (McNett, Omato, & Olson, 2017). Voor de ANTAT (Blomert et al., 1995) en de ANTAT-CU4 houdt dit in dat de testinstrumenten sprekers met een afatische stoornis moeten kunnen onderscheiden van gezonde sprekers. Specificiteit kan worden gedefinieerd als het aantal terecht geïdentificeerde negatieven. Voor de ANTAT en de ANTAT-CU4 betekent dit dat gezonde sprekers niet gezien mogen worden als sprekers met een afasie. Bij een onderzoeksinstrument is het belangrijk dat een test zo sensitief en tegelijk zo specifiek mogelijk is. Dat wil voor de ANTAT en ANTAT-CU4 zeggen, dat in een groep met gezonde en afatische sprekers er zoveel mogelijk afatische sprekers

geïdentificeerd worden als zijnde afatische sprekers en zoveel mogelijk gezonde sprekers gezien moeten worden als niet-afatische sprekers. Blomert et al. (1994) bespreken niet de termen sensitiviteit en specificiteit voor de ANTAT, maar ze spreken van *pathology criterion* (p. 384). Blomert et al. beschrijven dat gezonde sprekers bij de testafname op elk scenario een (bijna) maximale score moeten halen. Met de ANTAT-CU4 wordt in het huidige onderzoek alleen gekeken naar gezonde proefpersonen. Op basis hiervan kan een begin gemaakt worden met het instellen van een norm, hoe hoog gezonde sprekers moeten scoren per scenario en/of in zijn totaliteit.

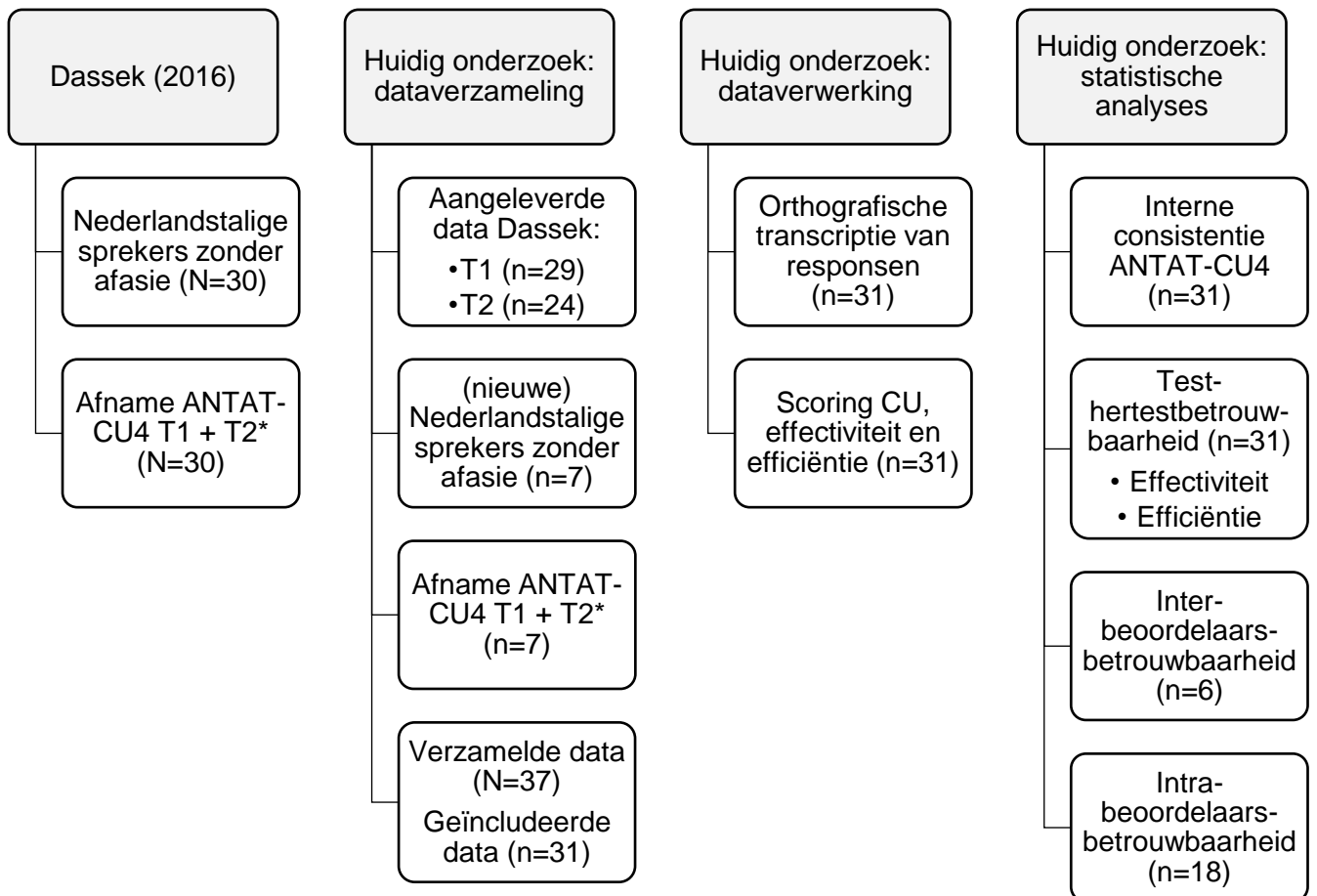


### 3. Methode

Zoals benoemd in de inleiding was het doel van dit onderzoek het in kaart brengen van de betrouwbaarheid van de ANTAT-CU4. In dit hoofdstuk wordt de methodische verantwoording rondom dit onderzoek besproken. Eerst wordt uitleg gegeven over de groep proefpersonen die deelnam aan het onderzoek en welke proefpersoonsgegevens zijn gebruikt voor het onderzoeken van de test-hertestbetrouwbaarheid en de inter- en intrabeoordelaarsbetrouwbaarheid. Daarna wordt ingegaan op de verwerking van de verkregen data en worden de analysemethoden besproken.

#### 3.1 Onderzoeksopzet

In Figuur 2 staat schematisch weergegeven welke data in het onderzoek van Dassek (2016) en het huidige onderzoek zijn verzameld en dat tussen de twee testafnamen (T1 en T2) ongeveer acht weken zat. Eveneens staat beschreven welke stappen in het huidige onderzoek zijn genomen om de data te verwerken en te analyseren.



Figuur 2. Schematische weergave onderzoeksontwerp.

\* De ANTAT-CU4 werd bij elke proefpersoon twee keer afgenomen met een tussentijd van ongeveer acht weken.

### **3.2 Proefpersonen**

Dassek (2016) heeft bij haar onderzoek een groep van 30 gezonde sprekers onderzocht en in het huidige onderzoek werden zeven gezonde sprekers onderzocht. In totaal deden dus 37 proefpersonen mee aan het onderzoek naar de betrouwbaarheid van de ANTAT-CU4 wat betreft gezonde sprekers. Echter, niet alle deelnemers konden worden geïnccludeerd in het huidige onderzoek. De reden hiervoor was dat niet alle verzamelde data van Dassek beschikbaar was voor nadere analyse. De gegevens van zes participanten konden niet worden meegenomen, omdat van hen de audio-opnamen van de tweede testafname niet voorhanden was. Van een van deze deelnemers was tevens de audio-opname van de eerste testafname niet beschikbaar. Om de onderzoeksgroep uit te breiden tot een groep van minimaal 30 personen, werd besloten om bij zes andere proefpersonen alsnog de ANTAT-CU4 tweemaal af te nemen. Uiteindelijk werden zeven nieuwe proefpersonen onderzocht, zodat ook de man-vrouwverhouding meer gelijk werd getrokken. Zodoende werd de data van 31 proefpersonen in het huidige onderzoek geïnccludeerd. Dit is tevens te zien in Figuur 2: verzamelde data (N=37) en geïnccludeerde data (n=31). Een schematisch overzicht van de 31 proefpersonen is te vinden in Bijlage B.

Om deel te mogen nemen aan het onderzoek moesten de proefpersonen aan een aantal voorwaarden voldoen. Zo moesten ze Nederlands als moedertaal hebben en minimaal 18 jaar oud zijn. Een andere vereiste was dat ze geen hersenbeschadiging of (neuro)degeneratieve aandoeningen mochten hebben. Ook mocht er geen sprake zijn van mentale retardatie. Ze mochten geen visus- of gehoorproblemen hebben, die niet door bril of lenzen of een gehoorapparaat gecorrigeerd konden worden. Daarnaast mochten de deelnemers geen beperking in de spraakverstaanbaarheid hebben.

In de volgende paragrafen wordt beschreven van welke proefpersonen de gegevens zijn gebruikt om de interne consistentie van de ANTAT-CU4, test-hertestbetrouwbaarheid en de inter- en intrabeoordelaarsbetrouwbaarheid te analyseren. In Figuur 2 wordt deze informatie schematisch weergegeven.

#### **3.2.1 Proefpersonen test-hertestbetrouwbaarheid**

De onderzoeksgroep voor de test-hertestbetrouwbaarheid bestond uit 31 proefpersonen tussen de 25 en 86 jaar met een gemiddelde leeftijd van 49,4 jaar en een standaarddeviatie van 17,1 jaar. Deze onderzoeksgroep kwam grotendeels overeen met de proefpersonen uit het onderzoek van Dassek (2016). Dat wil zeggen: van 24 proefpersonen, van de in de totaal 30 proefpersonen die Dassek onderzocht heeft, werden de gegevens meegenomen in het huidige onderzoek. De overige zeven proefpersonen kwamen zoveel mogelijk overeen met hetzelfde geslacht en dezelfde leeftijd als de zes proefpersonen uit het onderzoek van Dassek, van wie de data niet beschikbaar waren voor het huidige onderzoek.

De onderzoeksgroep bestond uit 16 mannen en 15 vrouwen. Alle proefpersonen waren rechtshandig op één man en twee vrouwen na. Alle proefpersonen hadden Nederlands als moedertaal en het opleidingsniveau varieerde van meer dan zes klassen lager onderwijs tot een universitaire opleiding.

Een deelnemer gaf aan een neurologische aandoening te hebben gehad in het verleden. Het betrof hier een hernia. Een hernia valt onder neurologische aandoeningen, maar het is een aandoening van het ruggenmerg. Hierbij drukt een afwijkende tussenwervelschijf op de

zenuwwortels (Kirchman, 2003). Een hernia heeft geen invloed op de hersenfuncties. Om deze reden werd deze deelnemer toch geïncludeerd in het onderzoek.

In Tabel I staat een overzicht van alle proefpersonen voor het onderdeel test-hertest betrouwbaarheid. In de tabel staat informatie over het geslacht, de leeftijd en het opleidingsniveau.

*Tabel I Gegevens proefpersonen test-hertestbetrouwbaarheid; geslacht leeftijd en opleidingsniveau (n=31).*

<b>N=31</b>		<b>Aantal (%)</b>
<b>Geslacht</b>	Vrouwelijk 15 (48.4%); Mannelijk 16 (51.6%)	
<b>Leeftijd<sup>1</sup></b>	Mean 49.4; SD 17.1; Range 61; Min 25; Max 86	
<b>Leeftijdscategorieën</b>	20-30	5 (16.1%)
	30-40	5 (16.1%)
	40-50	4 (12.9 %)
	50-60	10 (32.3 %)
	60-70	3 (9.7 %)
	70-80	2 (6.5 %)
	80-90	2 (6.5 %)
<b>Opleidingsniveau<sup>2</sup></b>	Mean 5.77, SD 1.2	
	1. Minder dan zes klassen lager onderwijs	0 (0 %)
	2. Zes klassen lager onderwijs	0 (0 %)
	3. Meer dan zes klassen lager onderwijs, zonder verder afgesloten speciale opleiding	1 (3.2 %)
	4. Gediplomeerde schoolopleiding lager dan MULO	3 (9.7 %)
	5. MULO	7 (22.6 %)
	6. VHMO	10 (32.3 %)
	7. WO	10 (32.3 %)

<sup>1</sup> in jaren

<sup>2</sup> volgens de indeling van Verhage (1964) aangepast door het Centraal Bureau voor de Statistiek (CBS; 2004)

### 3.2.2 Proefpersonen interbeoordelaarsbetrouwbaarheid

Om de interbeoordelaarsbetrouwbaarheid te testen had Dassek (2016) random zes proefpersonen uit verschillende leeftijdscategorieën uit het databestand gekozen. Het gaat hierbij om proefpersonen 1, 6, 10, 12, 16 en 22 (zie tevens Bijlage C). Een overzicht van deze proefpersonen is te vinden in Tabel II. De groep bestond uit vier mannen en twee vrouwen en de leeftijden van de proefpersonen liepen uiteen van 25 tot 86 jaar met een gemiddelde van 47,8 jaar.

Tabel II Gegevens proefpersonen interbeoordelaarsbetrouwbaarheid; geslacht leeftijd en opleidingsniveau (n=6).

N=6		Aantal (%)
<b>Geslacht</b>	Vrouwelijk 2 (33.3%); Mannelijk 4 (66.6%)	
<b>Leeftijd<sup>1</sup></b>	Mean 47.8; SD 21.8; Range 61; Min 25; Max 86	
<b>Leeftijdscategorieën</b>	20-30	1 (16.7 %)
	30-40	1 (16.7 %)
	40-50	2 (33.3 %)
	50-60	1 (16.7 %)
	60-70	0 (0 %)
	70-80	0 (0 %)
	80-90	1 (16.7 %)
<b>Opleidingsniveau<sup>2</sup></b>	1. Minder dan zes klassen lager onderwijs	0 (0 %)
	2. Zes klassen lager onderwijs	0 (0 %)
	3. Meer dan zes klassen lager onderwijs, zonder verder afgesloten speciale opleiding	0 (0 %)
	4. Gediplomeerde schoolopleiding lager dan MULO	0 (0 %)
	5. MULO	0 (0 %)
	6. VHMO	3 (50 %)
	7. WO	3 (50 %)

<sup>1</sup> in jaren

<sup>2</sup> volgens de indeling van Verhage (1964) aangepast door het CBS (2004)

Tabel III Gegevens proefpersonen Intrabeoordelaarsbetrouwbaarheid; geslacht leeftijd en opleidingsniveau (n=18).

N=18		Aantal (%)
<b>Geslacht</b>	Vrouwelijk 9 (50%); Mannelijk 9 (50%)	
<b>Leeftijd<sup>1</sup></b>	Mean 45.1; SD 15.1; Range 43; Min 26; Max 69	
<b>Leeftijdscategorieën</b>	20-30	4 (22.2 %)
	30-40	4 (22.2 %)
	40-50	1 (5.6 %)
	50-60	6 (33.3 %)
	60-70	3 (16.7 %)
	70-80	0 (0 %)
	80-90	0 (0 %)
<b>Opleidingsniveau<sup>2</sup></b>	Mean 5.78, SD 1.4	
	1. Minder dan zes klassen lager onderwijs	0 (0 %)
	2. Zes klassen lager onderwijs	0 (0 %)
	3. Meer dan zes klassen lager onderwijs, zonder verder afgesloten speciale opleiding	1 (5.6 %)
	4. Gediplomeerde schoolopleiding lager dan MULO	2 (11.1 %)
	5. MULO	3 (16.7 %)
	6. VHMO	5 (27.8 %)
7. WO	7 (38.9 %)	

<sup>1</sup> in jaren

<sup>2</sup> volgens de indeling van Verhage (1964) aangepast door het CBS (2004)

### 3.2.3 Proefpersonen intrabeoordelaarsbetrouwbaarheid

Voor het berekenen van de intrabeoordelaarsbetrouwbaarheid werden de gegevens gebruikt van 18 proefpersonen bij wie Dassek (2016) de ANTAT-CU4 had afgenomen en van wie de data niet gebruikt werd voor het berekenen van de interbeoordelaarsbetrouwbaarheid. Het gaat om de proefpersonen 2, 3, 4, 5, 7, 8, 11, 13, 14, 15, 17, 18, 19, 20, 21, 25, 26, 27 (zie eveneens Bijlage D). Deze groep bestond uit negen mannen en negen vrouwen. De leeftijden liepen uiteen van 26 tot 69 jaar met een gemiddelde leeftijd van 45,1 jaar. In Tabel III staan de gegevens van deze proefpersonen verder beschreven.

### 3.3 Materiaal

Bij alle 31 deelnemers werd de ANTAT-CU4 afgenomen, zoals beschreven in Dassek (2016). De ANTAT-CU4 bestaat uit 20 scenario's. Deze scenario's werden afgenomen inclusief drie aanvullende scenario's, namelijk een tweede versie van scenario 11 "bril", 16 "gips" en 20 "gemeente", zoals Dassek had aanbevolen. De in totaal 23 scenario's werden door Dassek verdeeld over twee parallelle versies. Zie Bijlage E voor een complete instructie van de testafname en de scenario's.

Bij zeven deelnemers werd nog een aanvullend scenario afgenomen. Het ging hierbij om een versie van scenario 20 (gemeente) waar al twee versies van bestonden (20.1 en 20.2), maar waarvan de instructies, volgens Dassek (2016), door de proefpersonen niet altijd goed begrepen werden. Dassek adviseerde om de formulering van dit scenario aan te passen. Het toegevoegde scenario betreft scenario 20.3 "gemeente" met de instructie: *"U belt naar de gemeente. U wilt iemand van de gemeente spreken, maar blijkbaar klopt het nummer niet. Per ongeluk krijgt u de tandarts aan de lijn. Wat zegt u?"*. In Tabel IV is overzichtelijk weergegeven bij welk aantal proefpersonen welke scenario's zijn afgenomen.

Het merendeel van de scenario's van de ANTAT-CU4 bestond alleen uit een mondelinge instructie. Zeven scenario's werden echter visueel ondersteund met een of twee foto's van een voorwerp. Acht van deze afbeeldingen zijn ontwikkeld door Aan de Stegge (2015) en Giessen (2015), een afbeelding is ontworpen door Dassek (2016). Zie Bijlage F voor een overzicht van alle afbeeldingen die gebruikt zijn bij de afname van de ANTAT-CU4.

*Tabel IV Weergave van het aantal afgenomen scenario's bij de proefpersonen uit het onderzoek van Dassek (2016) en de 'nieuwe' proefpersonen uit het huidige onderzoek.*

<b>Proefpersonen</b>	<b>Afgenomen scenario's</b>
24 proefpersonen (Dassek, 2016)	1-20, 11.2, 16.2, 20.2
7 proefpersonen (huidig onderzoek)	1-20, 11.2, 16.2, 20.2, 20.3

### 3.4 Procedure

Voorafgaand aan de deelname aan het onderzoek werden alle proefpersonen zowel mondeling als schriftelijk geïnformeerd over de testafname. De deelnemers kregen een informatiebrief waarin het onderzoek werd uitgelegd. In Bijlage G staat de informatiebrief die de deelnemers uit het onderzoek van Dassek (2016) kregen en in Bijlage H staat de informatiebrief die de zeven participanten kregen, die door de auteur van deze scriptie zijn onderzocht. Na de mondelinge en schriftelijke uitleg ondertekenden alle proefpersonen een toestemmingsverklaring voor deelname aan het onderzoek (zie Bijlage I voor de toestemmingsverklaring voor de deelnemers van het

onderzoek van Dassek (2016) en Bijlage J voor de toestemmingsverklaring voor de zeven overige deelnemers). Daarnaast vulden de deelnemers een vragenlijst in waarin werd gevraagd naar enkele persoonlijke gegevens, die van belang waren voor het onderzoek (zie Bijlage K voor de vragenlijst voor de deelnemers van het onderzoek van Dassek en Bijlage L voor de vragenlijst voor de zeven overige deelnemers). De vragen werden ontleend uit de vragenlijst van Filipinski (2014). De ANTAT-CU4 werd bij elke proefpersoon twee keer afgenomen met een tussentijd van ongeveer acht weken. Zowel de eerste als de tweede testafname werden zoveel mogelijk afgenomen in een rustige ruimte zonder omstanders. Van de testafnamen werden audio-opnamen gemaakt, zodat deze naderhand beluisterd konden worden door de testleider.

Tijdens de afname van de test kregen de deelnemers eerst een instructie te horen (zie Bijlage E voor de testinstructie). Deze instructie was overgenomen uit de instructie van de originele versie van de ANTAT (Blomert et al., 1995) met als aanvulling dat enkele situaties op elkaar lijken. Dit kwam door de dubbele scenario's, die werden afgenomen. De instructie was een korte uitleg over wat er van de proefpersoon verwacht werd. Deze instructie werd zowel bij de eerste als de tweede afname gegeven. Na de instructie kregen de deelnemers twee oefenopgaven om na te gaan of de instructie duidelijk was en zodat ze gewend konden raken aan de opdracht. Deze oefenopgaven waren, net als de instructie, overgenomen uit de originele versie van de ANTAT (Blomert et al., 1995). Na de oefenopgaven werd aan de proefpersonen gevraagd of de opdracht duidelijk was. Indien de instructie helder was en er geen vragen waren, werd de daadwerkelijke testafname gestart. Tijdens de afname van de ANTAT-CU4 mocht de testleider, behalve het aangeven van de scenario's geen aanvullende instructies geven of de proefpersonen helpen. Soms kwam het voor dat de proefpersoon aangaf dat hij de instructie van een scenario niet goed verstaan of begrepen had. Wanneer dit voorkwam, werd de instructie van het betreffende scenario herhaald. Wanneer een proefpersoon een respons gaf dat naar de mening van de testleider niet passend was bij de instructie van het scenario, kon de testleider benoemen dat de deelnemer de instructie wellicht niet goed begrepen had en het scenario nogmaals voorleggen, eventueel met nadruk op bepaalde woorden om het scenario te verduidelijken. De instructie voor de testleider in een dergelijk geval staat beschreven in Bijlage E. Op het geven van een reactie op een scenario was geen tijdslimiet. De proefpersonen konden zelf bepalen hoeveel tijd ze nodig hadden voordat ze een antwoord gaven en konden het antwoord zo kort of zo lang maken als ze zelf wilden.

Bij alle proefpersonen (n=31) werden op beide momenten alle 23 scenario's van de ANTAT-CU4 afgenomen, zoals beschreven in Dassek (2016). Bij zeven van deze proefpersonen werd een aanvullend scenario, namelijk scenario 20.3 (gemeente) afgenomen. Bij de 24 proefpersonen, onderzocht door Dassek (2016), werden de 23 scenario's verdeeld in twee parallelle versies van de test (zie Tabel V).

*Tabel V Volgorde van scenario's, zoals beschreven in Dassek (2016)*

<b>Versie 1 –</b>	1, 11.1, 2, 3, 4, 5, 6, 16.2, 7, 8, 9, 10,
<b>Versie 2</b>	20.1, 11.2, 12, 13, 14, 15, 16.1, 17, 18, 19, 20.2
<b>Versie 2 –</b>	20.1, 11.2, 12, 13, 14, 15, 16.1, 17, 18, 19, 20.2,
<b>Versie 1</b>	1, 11.1, 2, 3, 4, 5, 6, 16.2, 7, 8, 9, 10

De twee versies werden achter elkaar afgenomen. De volgorde van de afname (eerst versie 1 en dan versie 2 of andersom) verschilde per proefpersoon om te voorkomen dat de volgorde invloed

op de meetresultaten zou kunnen hebben. Ongeveer de helft van de proefpersonen kreeg bij de eerste afname eerst versie 1 en aansluitend versie 2 aangeboden. Bij de tweede afname kregen ze eerst versie 2 en aansluitend versie 1 van de test aangeboden. Voor de overige proefpersonen was dit andersom. Zie Tabel VI voor een overzicht van de aangeboden testvolgorde.

*Tabel VI Aangeboden testvolgorde*

<b>N=31</b>	<b>T1</b>	<b>Periode tussen T1 en T2</b>	<b>T2</b>
<b>Groep 1: 58.1% n=18</b>	versie 1 – versie 2	8 weken	versie 2 – versie 1
<b>Groep 2: 41.9% n=13</b>	versie 2 – versie 1	8 weken	versie 1 – versie 2

Voor de zeven proefpersonen bij wie ook scenario 20.3 afgenomen werd, werden deze parallelle versies, zoals Dassek (2016) had ingedeeld zoveel mogelijk aangehouden. De volgorde van de versies was gelijk aan de volgorde zoals die bedoeld was voor de proefpersonen van Dassek. Wel is scenario 20.3 toegevoegd en zijn de scenario's 20.1 en 20.2 verschoven, waardoor ze op een ander moment in de test werden afgenomen. Met scenario 20.3 bij de overige scenario's kwam het totale aantal scenario's op 24. Om de spreiding tussen de verschillende versies van scenario 20 zoveel mogelijk gelijk te houden is ervoor gekozen om tussen de verschillende versies van scenario 20 steeds zeven scenario's in te voegen. In Tabel VII is een overzicht te zien van de volgorde van de scenario's waar scenario 20.3 aan toegevoegd is.

*Tabel VII Volgorde van scenario's bij de zeven nieuwe proefpersonen, inclusief scenario 20.3*

<b>Versie 1 –</b>	1, 11.1, 2, 3, 4, 5, 6, 20.1, 16.2, 7, 8, 9, 10,
<b>Versie 2</b>	11.2, 12, 20.2, 13, 14, 15, 16.1, 17, 18, 19, 20.3
<b>Versie 2 –</b>	11.2, 12, 20.2, 13, 14, 15, 16.1, 17, 18, 19, 20.3
<b>Versie 1</b>	1, 11.1, 2, 3, 4, 5, 6, 20.1, 16.2, 7, 8, 9, 10,

### **3.5 Uitwerking audio-opnamen, meten en scoren**

De audiogegevens van de testafnames werden getranscribeerd en verwerkt voor de analyse. Tevens werd de spreekduur van de proefpersonen gemeten en werd het aantal geproduceerde woorden en Content Units (CUs) geteld. De gegevens werden onder andere gebruikt om de verbale effectiviteit en efficiëntie te meten.

#### **3.5.1 Transcriberen testafnames**

Op basis van de audio-opnamen, die gemaakt zijn tijdens het afnemen van de ANTAT-CU4, werden de uitingen van de proefpersonen getranscribeerd volgens het Protocol voor Orthografische Transcriptie (Corpus gesproken Nederlands, 2000). De uitingen van de testleider werden in principe niet getranscribeerd, omdat de testleider de instructies van de ANTAT-CU4 aan moest houden. Wel kwam het voor dat de testleider een scenario herhaalde voor de proefpersoon, nadat de proefpersoon al had gereageerd op de gegeven instructie. Dit gebeurde bijvoorbeeld wanneer de proefpersoon aangaf dat hij de instructie niet goed had begrepen. Het herhalen van de instructie werd in het transcript beschreven als 'herhaling instructie testleider'. Ook kwam het een enkele keer voor dat een participant om verduidelijking vroeg. Het antwoord

van de testleider werd dan getranscribeerd, zodat voor een onafhankelijke lezer duidelijk zou zijn wat zich tijdens de testafname had voorgedaan.

Na het transcriberen van de audio-opname werd dezelfde opname nogmaals beluisterd door de auteur van de transcripten, terwijl tegelijkertijd het transcript gelezen werd. Wanneer tijdens het terugluisteren onjuistheden in het transcript werden ontdekt, werden deze aangepast.

### **3.5.2 Woorden tellen aan de hand van transcripties**

Nadat de audio-opnamen waren getranscribeerd volgens het Protocol voor Orthografische Transcriptie (Corpus gesproken Nederlands, 2000) werd per persoon, per testafname, het totale aantal geproduceerde woorden geteld. Dit werd gedaan conform de Analyse voor Spontane Taal bij Afasie (ASTA; Boxum, van der Scheer & Zwaga, 2013). Om het tellen van het totale aantal woorden te vergemakkelijken werden uitingen, die volgens de ASTA niet als woord gescoord mogen worden, uit het transcript verwijderd.

Volgens het protocol van de ASTA (Boxum et al., 2013) mogen neologismen, stereotypen en echolalieën meegenomen worden in het totale aantal woorden. Ook herhalingen worden meegeteld bij het totale aantal woorden, mits minstens 50% van het aantal fonemen van het doelwoord is geproduceerd. Wanneer minder dan 50% van het doelwoord uitgesproken is, wordt dit niet gezien als een herhaling, maar als een valse start en wordt het niet gescoord als woord. Boxum et al. (2013) geven aan dat er ook uitzonderingen zijn. Wanneer het woord 'ja', 'nee' of 'nou' herhaald wordt, dan worden de herhalingen niet meegeteld. Tevens mogen interjecties, zoals 'hé' en 'eh' niet als woord gescoord worden. Vloeken zijn volgens Boxum et al. (2013) ook interjecties en worden om deze reden niet meegeteld bij het totale aantal woorden.

Bij het tellen van het aantal woorden was de keuze gemaakt om onderscheid te maken tussen het woord 'ja' als bevestiging op een vraag en het woord 'ja' als interjectie. Wanneer het woord 'ja' door de spreker werd geproduceerd als antwoord op een vraag van de testleider, werd dit als een woord gescoord. Wanneer het woord 'ja' door de proefpersoon niet als direct antwoord op een vraag werd gebruikt, werd dit gezien als een interjectie waardoor het niet als woord gescoord werd. Daarnaast werden woorden alleen meegeteld wanneer deze verstaanbaar waren uitgesproken door de proefpersoon.

Bij het tellen van het totale aantal woorden werden ook uitingen meegenomen die geen directe antwoorden waren op de gegeven scenario's. Wanneer de proefpersoon bijvoorbeeld om verheldering vroeg of een zijdelingse opmerking over een bepaald scenario maakte, werden deze uitingen meegenomen bij het bepalen van het totale aantal gesproken woorden.

### **3.5.3 CU-scores en effectiviteit**

Hoewel voor het berekenen van het totale aantal gebruikte woorden ook uitingen buiten de context meegenomen werden, was dit bij het berekenen van het totale aantal genoemde Content Units en het berekenen van de effectiviteit niet het geval. Hiervoor werden alleen de uitingen, of delen van uitingen, meegenomen die een concept uit het CU-schema uitdrukten. Voor het nagaan van de CU-scores moest het aantal genoemde Content Units worden bepaald. Om de geproduceerde Content Units te bepalen, werd gebruik gemaakt van het CU-schema voor de ANTAT-CU4, zoals opgesteld door Dassek (2016). Dit schema is te vinden in Bijlage M.

Om te bepalen wanneer verbale uitingen van een spreker gezien werden als een correcte Content Unit, werd in de eerste plaats nagegaan of deze uiting (of een synoniem ervan) voorkwam in het



CU-schema. Wanneer dit zo was, werd dit als een correct CU gescoord. Wanneer dit niet het geval was, werd bekeken of met de geproduceerde woorden dezelfde intentie werd uitgedrukt als met de woorden of woordgroepen in het CU-schema. Wanneer dit zo was, werd dit gescoord als een correct CU. Een voorbeeld zal dit mogelijk meer verduidelijken. In voorbeeld (6) staat een van de mogelijke aanleidingen van scenario 6 “slager” beschreven. Onder de aanleiding staan de Content Units weergegeven waarvoor de spreker punten kan krijgen. De instructie van scenario 6 is als volgt: “*U bent bij de slager en dit ligt op de grond. Wat zegt u?*”. Tijdens de instructie kreeg de proefpersoon een afbeelding van een portemonnee te zien.

(6) Aanleiding 1:        “*Ik heb een portemonnee gevonden*”  
 CUs:                    ik – portemonnee – gevonden

Wanneer een proefpersoon bijvoorbeeld het woord ‘beurs’ produceerde, gebruikte hij een synoniem van portemonnee en bedoelde hij dus hetzelfde voorwerp. Dit werd dan als een correct CU gescoord. Dit gold ook wanneer de spreker de zin produceerde in de tegenwoordige tijd in plaats van in de verleden tijd. Als de proefpersoon bijvoorbeeld zei: “*Ik vind een portemonnee*”, bleef de betekenis van het CU ‘gevonden’ gelijk en werd dit dus gezien als een correct CU.

*Tabel VIII Maximaal haalbare CU-score per scenario op de ANTAT-CU4*

<b>Scenario</b>	<b>Maximaal haalbare CU-score</b>
1	6
2	6/7
3	7
4	6
5	5
6	7
7	7
8	6
9	6
10	8
11	5
12	6
13	6
14	7
15	5/6
16	2
17	8
18	4/5/6
19	5
20	4
<b>Totaal</b>	<b>116-120</b>

Elk correct CU komt overeen met een punt. Het maximaal aantal haalbare Content Units op de scenario's loopt uiteen van 2 (scenario 16) tot 8 (scenario 10 en 17). Dit betekent dat per scenario een verschillend aantal Content Units nodig is om de gewenste boodschap over te brengen. Op de ANTAT-CU4 kunnen op 20 scenario's in totaal maximaal 116-120 punten<sup>2</sup> gescoord worden. In Tabel VIII staat weergegeven hoeveel Content Units maximaal per scenario behaald konden worden.

De variabele maximale totaalscore op de ANTAT-CU4 komt doordat er bij scenario 2 "voetballen", scenario 15 "huisarts" en scenario 18 "kat" verschillende formuleringmogelijkheden zijn. Voorbeeld (7), dat eerder in het hoofdstuk "Theoretische achtergrond" werd aangehaald, zal dit verduidelijken. Bij scenario 15 "huisarts" is de instructie als volgt: *"Er is bij u een kind gevallen. Het ziet er niet zo best uit. U belt naar de huisarts en zegt?"*. Het verzoek van de spreker kon dan bestaan uit twee formuleringen. In voorbeeld (7) staan twee mogelijke formuleringen beschreven.

- |                |  |
|----------------|--|
| (6) Verzoek 1: | <i>"Kan ik met het kind naar u <u>komen</u>?" / "Kunt u <u>komen</u>?"</i> |
| CU's:          | ik/u – komen   |
| Verzoek 2:     | <i>"Kunt u <u>naar</u> het kind <u>kijken</u>?"</i>                        |
| CU's:          | u – ernaar - kijken  |

Als de spreker verzoek 1 produceerde, kon hij maximaal 2 correcte Content Units behalen. Koos de spreker voor verzoek 2, kon hij maximaal 3 correcte Content Units behalen. Afhankelijk van de gekozen formulering werd dus nagegaan hoeveel Content Units de participant kon behalen. In Figuur 1 is het CU-schema behorend bij scenario 15 weergegeven. Scenario 2 "voetballen" en scenario 18 "kat" hebben een soortgelijke opbouw als scenario 15. Nadat de scores waren toegekend voor het aantal geproduceerde Content Units, werd het maximaal aantal haalbare Content Units per scenario bij elkaar opgeteld om uit te rekenen hoeveel Content Units de proefpersoon op de gehele ANTAT-CU4 kon behalen. De verbale effectiviteit, oftewel het percentage correct geproduceerde Content Units werd berekend door het totale aantal behaalde Content Units te delen door het maximale aantal Content Units dat behaald had kunnen worden (waarbij rekening werd gehouden met de formuleringkeuzes bij scenario 2, 15 en 18). Stel dat iemand 89 Content Units correct had geproduceerd en dat deze persoon maximaal 118 Content Units had kunnen formuleren, dan was de berekening als volgt:  $89 \text{ geproduceerde CUs} / 118 \text{ mogelijke CUs} * 100\% = 75.42\%$ .

### 3.5.4 Totale spreekduur meten

Voor het berekenen van de spreesnelheid werd eerst de spreekduur van de proefpersoon gemeten. Van de totale duur van de testafname, werd de spreekduur van de testleider afgetrokken. Pauzes tijdens de spreekbeurt van de testleider, vielen onder de spreekduur van de testleider. Wanneer de testleider klaar was met spreken, of wanneer de proefpersoon de testleider onderbrak en begon met spreken voordat de testleider uitgesproken was, ging de spreekduur van

---

<sup>2</sup> Wanneer de dubbele scenario's, namelijk 11.2, 16.2 en 20.2, werden geïnccludeerd in de ANTAT-CU4, waren er 23 scenario's en konden in totaal maximaal 127-131 punten worden gescoord.

de proefpersoon in. De spreekduur eindigde wanneer de proefpersoon klaar was met spreken. Pauzes die tijdens de spreekduur van de proefpersoon voorkwamen, werden in lijn met Oelschlaeger en Thorne (1999) meegenomen in de totale spreekduur van de proefpersoon. De spreekduur werd per scenario gemeten in seconden en daarna werden de spreekduren per scenario bij elkaar opgeteld.

### **3.5.5 Spreeksnelheid**

De spreesnelheid van de proefpersonen werd bepaald door het aantal geproduceerde woorden per minuut te berekenen. Hierbij werd uitgegaan van het totale aantal geproduceerde woorden dat overbleef na de Analyse voor Spontane Taal bij Afasie (ASTA; Boxum, van der Scheer & Zwaga, 2013). Uitingen die geen directe reactie waren op de scenario's werden hierbij meegeteld. Wanneer een proefpersoon bijvoorbeeld 571 woorden had geproduceerd en een spreekduur van 4.52 minuten had, was de berekening als volgt:  $571 \text{ woorden} / 4.52 \text{ minuten} = 126.33 \text{ woorden per minuut}$ .

### **3.5.6 Efficiëntie**

Na het opmeten van de spreekduur van de proefpersonen werd het percentage correcte informatie-eenheden (Content Units) berekend, dat wil zeggen het gemiddeld aantal Content Units per minuut per deelnemer, gemeten over de hele testafname. Wanneer een proefpersoon bijvoorbeeld 75 correcte Content Units had geproduceerd in een totale spreekduur van 2.4 minuten, dan was de berekening als volgt:  $75 \text{ CUs} / 2.4 \text{ minuten spreekduur} = 31.25 \text{ CUs per minuut}$ .

## **3.6 Analyse**

De data-analyses werden uitgevoerd in het programma SPSS 24 (Statistical Package of Social Sciences 24; IBM, 2016). Met deze analyses werden de interne consistentie, de test-hertestbetrouwbaarheid en inter- en intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4 berekend. In het huidige onderzoek werd voornamelijk gekeken naar de waarde van Cronbachs alfa als betrouwbaarheidsmaat. Cronbachs alfa is de meest gebruikte maat van betrouwbaarheid (Field, 2013). Hiermee kan worden vastgesteld of meerdere items samen een schaal mogen vormen. De ANTAT-CU4 bestaat uit verschillende scenario's die samen een geheel moeten vormen. De Cronbachs alfawaarde leek dus een geschikte manier om na te gaan in hoeverre de verschillende scenario's correleren op de verschillende betrouwbaarheidsanalyses.

### **3.6.1 Betrouwbaarheid van de ANTAT-CU4**

Om de betrouwbaarheid van de ANTAT-CU4 te onderzoeken, ofwel de consistentie tussen de scenario's na te gaan, werd een *reliability analysis* uitgevoerd. Om deze betrouwbaarheidsanalyse uit te voeren, werden zowel op het eerste als het tweede meetmoment alle scenario's met elkaar vergeleken. Hierbij werd gekeken naar de waarde van de Cronbachs alfa om de mate van betrouwbaarheid te bepalen en na te gaan welke scenario's goed en minder goed pasten in de gehele ANTAT-CU4. Hiervoor werd de classificatie volgens Van Wijk (2000) aangehouden.

### **3.6.2 Test-hertestbetrouwbaarheid**

De test-hertestbetrouwbaarheid van de ANTAT-CU4 werd getoetst door te onderzoeken of de behaalde scores van de proefpersonen stabiel waren over de tijd heen. Hierbij werd nagegaan of er een verschil was in de mate van verbale effectiviteit en verbale efficiëntie op de eerste en de tweede testafname. Er werd gebruik gemaakt van gepaarde t-toetsen (tweezijdig), waarbij de Cronbachs alfa als betrouwbaarheidsmaat werd gehandhaafd.

De effectiviteit werd berekend door na te gaan of er over de tijd heen verschillen waren in het aantal geproduceerde Content Units. De efficiëntie werd berekend om te onderzoeken of de proefpersonen bij de tweede testafname meer Content Units per minuut produceerden. Voor het berekenen van de efficiëntie, was het van belang om eerst de spreeknelheid te bepalen. De spreeknelheid werd nagegaan door de totale spreekduur en het totale aantal gebruikte woorden te bepalen en daarna het aantal geproduceerde woorden per minuut te berekenen. Om te controleren of er verschillen waren wat betreft de spreekduur, het totale aantal gebruikte woorden en het aantal geproduceerde woorden per minuut op de twee testafnamen, werden gepaarde t-toetsen (tweezijdig) uitgevoerd.

Tevens werd gekeken naar de correlaties van de effectiviteit en efficiëntie tussen de twee testafnamen. Een dergelijk correlatiecoëfficiënt geeft namelijk een schatting van de test-hertestbetrouwbaarheid en hangt tevens samen met de interne consistentie van een test (Van den Brink, 1998). Om deze reden was het van belang om bij de test-hertestanalyse de correlatie te onderzoeken.

### **3.6.3 Begrip van testinstructie**

Om na te gaan of proefpersonen bepaalde scenario's minder goed begrepen, werd nagegaan bij welke scenario's de proefpersonen zelf om herhaling van de testinstructie vroegen. Ook werd bekeken bij welke scenario's de testleider de instructie moest herhalen, omdat de deelnemer een onjuiste respons gaf. Dit werd voor alle scenario's nagegaan.

Er werd specifiek gekeken naar de testinstructies van scenario's 11 "bril", 16 "gips" en 20.2 "gemeente", omdat Dassek (2016) aangaf dat voornamelijk op deze scenario's incorrecte antwoorden werden gegeven. Zij heeft daarom in haar onderzoek twee versies van deze scenario's ontwikkeld. Bij scenario 11.1 "bril" en 11.2 "bril" is de testinstructie gelijk, maar worden verschillende afbeeldingen ter ondersteuning gebruikt. Bij scenario 11.1 krijgt de proefpersoon een afbeelding te zien van een bril met een scheef pootje en bij scenario 11.2 krijgt de deelnemer een afbeelding te zien van een bril die bij de neusbrug doormidden is gebroken (zie Bijlage F voor de afbeeldingen). Bij scenario 16.1 "gips" en 16.2 "gips" is een lichte variatie in de testinstructie. Ook bij scenario 20.1 "gips" en 20.2 "gips" is een lichte variatie in de testinstructie. Om te onderzoeken of de proefpersonen anders scoorden op de verschillende versies van de dubbele scenario's, werd voor beide testafnamemomenten de gemiddelde behaalde CU-score op deze scenario's berekend.

### **3.6.4 Interbeoordelaarsbetrouwbaarheid**

Om de interbeoordelaarsbetrouwbaarheid te berekenen werd gebruik gemaakt van een gepaarde t-toets (tweezijdig). Hierbij werden de totale behaalde ruwe CU-scores op beide meetmomenten volgens de interpretatie van Dassek (2016) en de auteur van deze scriptie met elkaar vergeleken. De reden dat naar de ruwe CU-scores werd gekeken en niet naar het percentage CU-scores, is

omdat de mate van overeenstemming tussen de beoordelaars onderzocht werd. Voor beide berekeningen van de interbeoordelaarsbetrouwbaarheid is daarnaast ook gekeken naar de Intraclass Correlation Coëfficiënt (ICC; Shrout & Fleiss, 1979). De ICC wordt gebruikt om de relatie tussen twee variabelen die hetzelfde meten, vast te stellen (Field, 2013). De ICC werd in dit geval dus, zoals Shrout en Fleiss (1979) zouden beschrijven, gebruikt om de conformiteit van de metingen, uitgevoerd door verschillende waarnemers, te beoordelen. Bij het uitvoeren van deze betrouwbaarheidsanalyse werd voor een *two-way random model* gekozen om de *absolute agreement* (mate van overeenstemming) te bepalen voor *average measures*. De reden dat gekozen werd voor een *two-way random model* is omdat de beoordelaars van het huidige onderzoek een representatie vormen van een grotere groep beoordelaars, namelijk logopedisten/taal-spraakpathologen. De keuze voor *absolute agreement* werd gemaakt omdat hiermee de mate van overeenstemming tussen verschillende beoordelaars onderzocht kon worden. Het is van belang dat iedere beoordelaar dezelfde punten toekent aan de geproduceerde uitingen, dat de gegeven scores identiek zijn aan elkaar. Wanneer namelijk een cliënt met afasie door twee verschillende personen wordt beoordeeld op de verbale communicatievaardigheden, zouden beiden tot dezelfde conclusie moeten komen. Er werd gekozen voor *average measures*, omdat hiermee de betrouwbaarheid van de overeenstemming tussen de beoordelaars kon worden nagegaan.

In paragraaf 4.5.1 zal worden beschreven dat er een grote discrepantie bestaat tussen de gegeven scores van Dassek (2016) en de auteur van deze scriptie. Om deze reden werd besloten om de gegevens van de zes proefpersonen te laten beoordelen door een derde onafhankelijke beoordelaar, Van den Berg (logopedist, afasietherapeut en Taal- en Spraakpatholoog). De scores van Van den Berg zullen door middel van gepaarde t-toetsen (tweezijdig) worden vergeleken met de scores gegeven door Dassek en de scores gegeven door de auteur van deze scriptie.

Van den Berg had geen ervaring met de ANTAT-CU4 en het ANTAT-CU4-schema. Om deze reden werd ze door de auteur van deze scriptie geïnformeerd over het gebruik van het ANTAT-CU4-schema. Van den Berg heeft met behulp van de getranscribeerde uitwerkingen van de testafnamen, uitgewerkt door de auteur van deze scriptie, de prestaties van het eerste en tweede testafnamemoment van zes proefpersonen beoordeeld. Ze gebruikte het ANTAT-CU4-schema, opgesteld door Dassek (2016), om CU-scores toe te kennen. Tevens kreeg ze enkele instructies. Zo werd aangegeven dat een synoniem van een Content Unit als correct gescoord mocht worden, dit stond namelijk niet expliciet benoemd in het CU-schema. Daarnaast moest ze bij twijfel bij zichzelf nagaan of de boodschap door de spreker duidelijk werd overgebracht.

### **3.6.5 Intrabeoordelaarsbetrouwbaarheid**

Om de intrabeoordelaarsbetrouwbaarheid te berekenen heeft de auteur van deze scriptie tweemaal, met een tussentijd van ongeveer acht weken, de prestaties van 18 proefpersonen op het eerste testafnamemoment gescoord. De gegeven ruwe CU-scores werden middels een gepaarde t-toets (tweezijdig) met elkaar vergeleken. Op deze manier werd onderzocht hoe consistent de beoordelaar was in haar beoordeling. De Cronbachs alfa werd als betrouwbaarheidsmaat aangehouden.

#### 4. Resultaten

In dit hoofdstuk worden de resultaten van de analyses besproken. Als eerste wordt ingegaan op de betrouwbaarheid van de ANTAT-CU4, ofwel de interne consistentie. Daaropvolgend worden de test-hertestbetrouwbaarheid en de inter- en intrabeoordelaarsbetrouwbaarheid van de ANTAT-CU4 besproken. Dit alles wordt gedaan voor de ANTAT-CU4, bestaande uit 20 scenario's, en de ANTAT-CU4 met drie dubbele scenario's, namelijk 11.2 "bril", 16.2 "gips" en 20.2 "gemeente".

##### 4.1 Betrouwbaarheid van de ANTAT-CU4; 20 scenario's

Op de eerste testafname behaalden de 31 proefpersonen, op de 20 scenario's die in de definitieve versie van de ANTAT-CU4 zullen komen, een gemiddelde CU-score van 68.42 met een standaarddeviatie van 11.89. Met een Cronbachs alfawaarde van .754 is de betrouwbaarheid van de definitieve versie van op het eerste afnamemoment adequaat (Van Wijk, 2000). Uit de analyse kwam naar voren dat drie scenario's minder goed pasten bij de overige scenario's. Het verwijderen van elk van deze scenario's, zou de betrouwbaarheid van de ANTAT-CU4 verhogen. Het betrof scenario 1 "stomerij", scenario 12 "verjaardag" en scenario 20.2 "gemeente". De Cronbachs alfawaarde zou na verwijdering van deze scenario's verhogen naar respectievelijk .762 (scenario 1), .756 (scenario 12) of .759 (scenario 20.2). In Tabel IX staan de Cronbachs alfawaarden weergegeven, indien het scenario uit de test verwijderd zou worden. Wanneer alle drie de scenario's uit de test verwijderd zouden worden, zou de Cronbachs alfawaarde verhogen naar .770.

*Tabel IX Item-totaal correlatie wanneer het scenario in de test wordt opgenomen en Cronbachs alfawaarde wanneer het scenario uit de test verwijderd zou worden.*

<b>Cronbachs alfa-waarde ANTAT-CU4; 20 scenario's</b>	<b>Scenario</b>	<b>Item-totaal correlatie</b>	<b>Cronbachs alfa na verwijderen scenario</b>
<b>T1 = .754</b>	1	.075	.762
	12	.180	.756
	20.2	.101	.759
<b>T2 = .687</b>	6	.109	.695
	11.2	-.188	.721
	12	.090	.695
	17	-.010	.704
	20.2	.051	.693

Op de tweede testafname behaalden de 31 proefpersonen op deze 20 scenario's een gemiddelde CU-score van 71.48 met een standaarddeviatie van 10.71. Met een Cronbachs alfawaarde van .687 is de betrouwbaarheid op het tweede afnamemoment matig (Van Wijk, 2000). Uit de analyse kwam naar voren dat het verwijderen van vijf scenario's de betrouwbaarheid van de test zou verhogen. Het ging om scenario 6 "slager", 11.2 "bril", 12 "verjaardag", 17 "station" en 20.2 "gemeente". De Cronbachs alfawaarde zou na verwijdering uitkomen op respectievelijk .695 (scenario 6), .721 (scenario 11.2), .695 (scenario 12), .704 (scenario 17) en .693 (scenario 20.2). Wanneer alle vijf scenario's uit de test verwijderd zouden worden, zou de Cronbachs alfawaarde verhogen naar .765, wat adequaat is volgens Van Wijk. Tevens was bij de tweede afname te zien

dat scenario 11.2 en 17 een negatieve waarde hadden, wat betreft de item-totaal correlatie. Zie Tabel IX voor een overzicht van de item-totaal correlaties en Cronbachs alfawaarden indien het scenario uit de test verwijderd zou worden.

#### 4.2 Betrouwbaarheid van de ANTAT-CU4; 23 scenario's

Op de eerste testafname behaalden de 31 proefpersonen op de 23 scenario's een gemiddelde CU-score van 76.00 met een standaarddeviatie van 13.48. Met een Cronbachs alfawaarde van .789 is de betrouwbaarheid van de ANTAT-CU4 op het eerste toetsmoment adequaat (Van Wijk, 2000). Uit de analyse kwam naar voren dat het verwijderen van twee scenario's de betrouwbaarheid van de test zouden verhogen. Het gaat om scenario 1 "stomerij" en scenario 12 "verjaardag". Wanneer scenario 1 uit de ANTAT-CU4 verwijderd zou worden, zou de waarde van de Cronbachs alfa .796 zijn. Wanneer scenario 12 niet meegenomen zou worden in het geheel, zou de waarde van de Cronbachs alfa op .792 uitkomen. In Tabel X staan de Cronbachs alfawaarden na verwijdering van het scenario uit de test weergegeven. Als beide scenario's uit de test verwijderd zouden worden, zou de Cronbachs alfawaarde verhogen naar .799.

*Tabel X Item-totaal correlatie wanneer het scenario in de test wordt opgenomen en Cronbachs alfawaarde wanneer het scenario uit de test verwijderd zou worden.*

<b>Cronbachs alfa-waarde ANTAT-CU4; 23 scenario's</b>	<b>Scenario</b>	<b>Item-totaal correlatie</b>	<b>Cronbachs alfa na verwijderen scenario</b>
<b>T1 = .789</b>	1	.073	.796
	12	.168	.792
<b>T2 = .632</b>	6	.078	.642
	11.1	-.110	.652
	11.2	-.197	.670
	12	.078	.639
	13	.119	.633
	16.1	-.136	.643
	17	.017	.645
	20.1	-.149	.655

Op de tweede testafname behaalden de 31 proefpersonen op dezelfde 23 scenario's een gemiddelde CU-score van 80.06 met een standaarddeviatie van 10.44. Met een Cronbachs alfawaarde van .632 is de betrouwbaarheid van de ANTAT-CU4 op het tweede toetsmoment matig (Van Wijk, 2000). Daarnaast kwam uit de analyse naar voren dat het verwijderen van acht scenario's de betrouwbaarheid van de ANTAT-CU4 bij de tweede afname zou verhogen. Het ging om scenario 6 "slager", 11.1 "bril", 11.2 "bril", 12 "verjaardag", 13 "rekening", 16.1 "gips", 17 "station" en 20.1 "gemeente". De Cronbachs alfawaarde zou na verwijdering uitkomen op respectievelijk .642 (scenario 6), .652 (scenario 11.1), .670 (scenario 11.2), .639 (scenario 12), .633 (scenario 13), .643 (scenario 16.1), .645 (scenario 17) en .655 (scenario 20.1)

Wanneer al deze scenario's uit de test verwijderd zouden worden, dan zou de Cronbachs alfawaarde verhogen naar .764, wat volgens Van Wijk adequaat is. Op de tweede afname was tevens te zien dat scenario 11.1, 11.2, 16.1 en 20.1 een negatieve waarde hadden, wat betreft de

item-totaal correlatie. Zie Tabel X voor een overzicht van de item-totaal correlaties en Cronbachs alfawaarden indien het scenario uit de test verwijderd zou worden.

### **4.3 Test-hertestbetrouwbaarheid ANTAT-CU4; 20 scenario's**

#### **4.3.1 Effectiviteit**

Er werd geen significant verschil gevonden tussen het percentage geproduceerde Content Units op het eerste afnamemoment ( $M=58.34$ ,  $SD=10.00$ ) en het tweede afnamemoment ( $M=60.73$ ,  $SD=9.00$ ),  $t(30) = -1.91$ ,  $p = .066$ ;  $r = .71$ ,  $p < .001$ .

In totaal was voor 20 van de 31 proefpersonen het percentage geproduceerde Content Units op het tweede afnamemoment groter dan op het eerste afnamemoment en 11 proefpersonen produceerden tijdens het tweede toetsmoment een lager percentage Content Units. In Bijlage N is een schematisch overzicht te zien van het percentage geproduceerde Content Units per proefpersoon per toetsmoment.

#### **4.3.2 Spreeksnelheid**

Er werd geen significant verschil gevonden tussen het aantal geproduceerde woorden op de eerste testafname ( $M=405.16$ ,  $SD=207.08$ ) en de tweede testafname ( $M=381.84$ ,  $SD=193.60$ ),  $t(30) = 1.91$ ,  $p = .066$ ;  $r = .945$ ,  $p < .001$ .

Er werd wel een significant verschil gevonden in de totale spreekduur in seconden, waarbij de spreekduur bij de tweede testafname gemiddeld significant korter was ( $M=146.29$ ,  $SD=88.54$ ) dan bij de eerste testafname ( $M=158.74$ ,  $SD=89.05$ ),  $t(30) = 2.06$ ,  $p < .05$ ;  $r = .928$ ,  $p < .001$ .

Er werd geen significant verschil gevonden tussen de gemiddelde spreesnelheid, oftewel het aantal geproduceerde woorden per minuut, op de eerste en de tweede afname. Op de eerste afname produceerden de proefpersonen gemiddeld 158.97 woorden per minuut ( $SD=24.29$ ) en op de tweede afname gemiddeld 163.17 woorden per minuut ( $SD=23.94$ ),  $t(30) = -1.44$ ,  $p = .161$ ;  $r = .762$ ,  $p < .001$ .

Uit een *oneway* ANOVA voor het eerste testafnamemoment, bleek dat het aantal geproduceerde woorden niet significant correleerde met de spreekduur. Dat wil zeggen dat het aantal geproduceerde woorden niet significant hoger werd naarmate een spreker een langere totale spreekduur had,  $F(1,29) = 55.42$ ,  $p = .106$ ,  $\eta^2_p = .999$ . Ditzelfde bleek voor het tweede testafnamemoment,  $F(1,29) = 23.83$ ,  $p = .161$ ,  $\eta^2_p = .999$ .

#### **4.3.3 Efficiëntie**

Er werd een significant verschil gevonden tussen het gemiddelde aantal geproduceerde Content Units per minuut op het eerste testmoment ( $M=31.16$ ,  $SD=11.98$ ) en het tweede testmoment ( $M=35.64$ ,  $SD=12.81$ ),  $t(30) = -4.32$ ,  $p < .001$ ;  $r = .893$ ,  $p < .001$ . Dit betekent dat de participanten tijdens de tweede afname per minuut gemiddeld meer Content Units produceerden dan tijdens de eerste testafname. In totaal produceerden 24 proefpersonen tijdens het tweede afnamemoment meer Content Units per minuut dan tijdens de eerste afname. Zeven proefpersonen produceerden juist tijdens de eerste testafname meer Content Units per minuut. In Bijlage O staat een schematisch overzicht van de geproduceerde Content Units per minuut door alle 31 proefpersonen op de twee toetsmomenten weergegeven.



## 4.4 Test-hertestbetrouwbaarheid van de ANTAT-CU4; 23 scenario's

### 4.4.1 Effectiviteit

Er werd een significant verschil gevonden tussen het percentage geproduceerde Content Units op het eerste afnamemoment ( $M=59.15$ ,  $SD=10.39$ ) en het tweede afnamemoment ( $M=62.21$ ,  $SD=8.01$ ),  $t(30) = -2.32$ ,  $p = .027$ . Dit betekent dat tijdens de tweede testafname het percentage geproduceerde Content Units significant hoger was dan tijdens de eerste testafname. De correlatie tussen de effectiviteit op de ANTAT-CU4 op meetmoment 1 en meetmoment 2 was met een waarde van .711 vrij goed (Hutcheson & Sofroniou, 1999) en had een significantie van  $p < .001$ .

In totaal was voor 20 van de 31 proefpersonen het percentage geproduceerde Content Units op het tweede afnamemoment groter dan op het eerste afnamemoment en 11 proefpersonen produceerden tijdens het tweede toetsmoment een lager percentage Content Units. In Bijlage P is een schematisch overzicht te zien van het percentage geproduceerde Content Units per proefpersoon per toetsmoment.

### 4.4.2 Spreeksnelheid

Bij de tweede testafname werden door de 31 participanten gemiddeld significant minder woorden geproduceerd ( $M=413.97$ ,  $SD=213.66$ ) dan bij de eerste testafname ( $M=458.48$ ,  $SD=240.95$ ),  $t(30) = 2.99$ ,  $p = .006$ ;  $r = .940$ ,  $p < .001$ .

Eveneens was de totale spreekduur in seconden voor de 31 proefpersonen bij de tweede testafname gemiddeld significant korter ( $M=165.65$ ,  $SD=103.30$ ) dan bij de eerste testafname ( $M=181.16$ ,  $SD=104.32$ ),  $t(30) = 2.21$ ,  $p = .035$ ;  $r = .929$ ,  $p < .001$ .

Wat betreft de spreeksnelheid, het gemiddeld aantal geproduceerde woorden per minuut, werd voor de 31 deelnemers op de eerste en de tweede testafname geen significant verschil gevonden. Bij de eerste testafname werden gemiddeld ongeveer evenveel woorden per minuut geproduceerd ( $M=157.35$ ,  $SD=23.81$ ) als tijdens de tweede testafname ( $M=157.06$ ,  $SD=23.36$ ),  $t(30) = .10$ ,  $p = .92$ ;  $r = .743$ ,  $p < .001$ .

Uit een *oneway* ANOVA voor het eerste testafnamemoment, bleek dat het aantal geproduceerde woorden significant correleerde met de spreekduur, dat wil zeggen, hoe langer de totale spreekduur hoe meer woorden er in totaal werden geproduceerd,  $F(3,27) = 32.49$ ,  $p = .007$ ,  $\eta^2_p = .997$ . Uit een *oneway* ANOVA voor het tweede testafnamemoment, bleek dat wederom het aantal geproduceerde woorden significant correleerde met de spreekduur,  $F(3,27) = 28.63$ ,  $p = .009$ ,  $\eta^2_p = .996$ .

### 4.4.3 Efficiëntie

Er werd een significant verschil gevonden tussen het gemiddelde aantal geproduceerde Content Units per minuut op het eerste testmoment ( $M=30.54$ ,  $SD=11.91$ ) en het tweede testmoment ( $M=35.92$ ,  $SD=13.63$ ),  $t(30) = -5.39$ ,  $p < .001$ ;  $r = .914$ ,  $p < .001$ . Dit betekent dat de participanten tijdens de tweede afname per minuut gemiddeld meer Content Units produceerden dan tijdens de eerste testafname. In totaal produceerden 24 proefpersonen tijdens het tweede afnamemoment meer Content Units per minuut dan tijdens de eerste afname. Zeven proefpersonen produceerden juist tijdens de eerste testafname meer Content Units per minuut. In Bijlage Q staat een schematisch overzicht van de geproduceerde Content Units per minuut door alle 31 proefpersonen op de twee toetsmomenten weergegeven.

#### 4.5 Begrip van testinstructie

In Tabel XI staat schematisch weergegeven hoeveel proefpersonen moeite hadden met de verschillende versies van de dubbele scenario's. Hierbij is onderscheid gemaakt tussen het begrijpen van de instructie op de eerste en tweede testafname, omdat hier soms verschillen in waren.

*Tabel XI Aantal proefpersonen die op de eerste of tweede testafname een scenario niet goed begrepen (n=31)*

<b>Scenario</b>	<b>Testafname 1</b>	<b>Testafname 2</b>
11.1 "bril" scheef pootje	1	1
11.2 "bril" kapotte neusbrug <sup>1</sup>	0	0
16.1 gips <sup>1</sup>	0	0
16.2 gips	0	0
20.1 "gemeente"	6	2
20.2 "gemeente" <sup>1</sup>	7	2

<sup>1</sup> De versie van het dubbele scenario dat in de ANTAT-CU4 is opgenomen.

In Tabel XI is af te lezen dat scenario 11.1 door een proefpersoon minder goed werd begrepen. Met de afbeelding die werd gebruikt bij scenario 11.2 had geen van de deelnemers moeite om deze te begrijpen. In Tabel XI is ook te zien dat geen van de proefpersonen moeite had met scenario 16.1 en 16.2. De meeste onduidelijkheden waren bij de testinstructies van scenario 20.1 en 20.2.

Om na te gaan of de proefpersonen ook daadwerkelijk anders scoorden op de verschillende versies van de dubbele scenario's, werd voor beide testafnamemomenten de gemiddelde behaalde CU-score voor deze scenario's berekend. In Tabel XII staan de behaalde CU-scores per scenario, per testafnamemoment weergegeven.

*Tabel XII Gemiddelde CU-scores met bijbehorende standaarddeviaties (SD) op het eerste en tweede testafnamemoment op scenario 11.1, 11.2, 16.1, 16.2, 20.1 en 20.2*

<b>Scenario</b>	<b>Gemiddelde CU-score T1</b>	<b>Gemiddelde CU-score T2</b>
<b>11.1</b>	3.35 (SD 1.58)	3.90 (SD 1.19)
<b>11.2</b>	3.48 (SD 1.46)	3.29 (SD 1.51)
<b>16.1</b>	1.35 (SD .71)	1.58 (SD .50)
<b>16.2</b>	1.55 (SD .62)	1.55 (SD .62)
<b>20.1</b>	2.87 (SD 1.23)	3.03 (SD 1.17)
<b>20.2</b>	2.84 (SD 1.32)	3.16 (SD 1.13)

In Tabel XII is af te lezen dat de verschillen in de behaalde CU-scores op de verschillende versies van scenario 11, 16 en 20 minimaal zijn. Uit gepaarde t-toetsen (tweezijdig) bleek tevens dat deze verschillen niet significant zijn.

Behalve de hierboven besproken scenario's, bleek dat ook andere scenario's niet altijd even goed werden begrepen door de proefpersonen. Het ging om scenario 9 "nieuwe buurman", 13

“afschrift”, 15 “huisarts” en 18 “kat”. In Tabel XIII staat schematisch weergegeven hoeveel personen op de eerste en tweede testafname moeite hadden met de scenario’s.

*Tabel XIII Scenario’s die door een of meer proefpersonen niet goed begrepen werden tijdens de eerste of tweede testafname*

Scenario	Testafname 1	Testafname 2
9 “nieuwe buurman”	1	5
13 “afschrift”	4	0
15 “huisarts”	1	0
18 “kat”	1	1

Bij scenario 9 “nieuwe buurman” had tijdens de eerste testafname een persoon moeite met het scenario. Bij de tweede testafname hadden vijf deelnemers moeite met dit item. Scenario 13 “afschrift” werd alleen op het eerste testafnamemoment door een vier proefpersonen minder goed begrepen. Scenario 15 “huisarts” werd door een deelnemer minder goed begrepen op de eerste testafname. Scenario 18 “kat” werd op beide testafnamemomenten door een proefpersoon mogelijk niet goed begrepen.

#### **4.6 Interbeoordelaarsbetrouwbaarheid ANTAT-CU4; 20 scenario’s**

##### **4.6.1 Vergelijking tussen de CU-scores van de auteur van deze scriptie en Dassek (2016)**

Voor het eerste testafnamemoment was de ICC-waarde .09 (zwak; Koo & Li, 2016),  $F(5,5) = 2.31$ ,  $p = .190$ . Voor het tweede testafnamemoment was de ICC-waarde .17 (zwak; Koo & Li),  $F(5,5) = 6.86$ ,  $p = .027$ .

Op de eerste testafname werd door de auteur van deze scriptie een significant lager gemiddelde CU-score van 67.17 ( $SD=11.67$ ) gegeven dan door Dassek (2016), die een gemiddelde CU-score gaf van 103.17 ( $SD=5.35$ ),  $t(5) = -8.84$ ,  $p < .001$ . Op de tweede testafname werd door de auteur van deze scriptie wederom gemiddeld een significant lagere CU-score gegeven dan door Dassek. Zo werd door de auteur van deze scriptie een gemiddelde CU-score gegeven van 70.50 ( $SD=13.02$ ), terwijl Dassek een gemiddelde CU-score gaf van 109.50 ( $SD=6.54$ ),  $t(5) = -13.00$ ,  $p < .001$ .

##### **4.6.2 Vergelijking tussen de CU-scores van Dassek (2016) en Van den Berg**

Voor het eerste toetsmoment was de ICC-waarde .11 (zwak; Koo & Li, 2016),  $F(5,5) = 2.35$ ,  $p = .185$ . Voor het tweede toetsmoment was de ICC-waarde .21 (zwak; Koo & Li),  $F(5,5) = 7.44$ ,  $p = .023$ .

Op de eerste testafname werd door Dassek een significant hogere gemiddelde CU-score gegeven, van 103.71 ( $SD=5.34$ ), dan door Van den Berg, die een gemiddelde CU-score gaf van 72.00 ( $SD=11.24$ ),  $t(5) = 7.94$ ,  $p = .001$ . Op de tweede afname werd door Dassek eveneens een significant hogere gemiddelde CU-score gegeven, van 109.50 ( $SD = 6.54$ ), dan door Van den Berg, die een gemiddelde CU-score gaf van 76.00 ( $SD=12.70$ ),  $t(5) = 11.81$ ,  $p < .001$ .

#### 4.6.3 Vergelijking tussen de CU-scores van de auteur van deze scriptie en Van den Berg

Voor het eerste testafnamemoment was de ICC-waarde .96 (uitmuntend; Koo & Li, 2016),  $F(5,5) = 542.24$ ,  $p < .001$ . Voor het tweede testafnamemoment was de ICC-waarde .95 (uitmuntend; Koo & Li),  $F(5,5) = 119.26$ ,  $p < .001$ .

Uit gepaarde t-toetsen, waarmee de gemiddelde CU-scores van Van den Berg en de auteur van deze scriptie werden vergeleken, kwam naar voren dat bij de eerste testafname Van den Berg een significant hogere gemiddelde CU-score gaf ( $M=72.00$ ,  $SD=11.24$ ) dan de auteur van deze scriptie ( $M=67.17$ ,  $SD=11.70$ ),  $t(5) = -12.04$ ,  $p < .001$ . Ook op de tweede testafname werd door Van den Berg een significant gemiddeld hogere CU-score gegeven ( $M=76.00$ ,  $SD=12.70$ ) dan door de auteur van deze scriptie ( $M=70.50$ ,  $SD=13.02$ ),  $t(5) = -5.75$ ,  $p = .002$ .

In Tabel XIV staan de gemiddelde CU-scores, gegeven door de beoordelaars op de eerste en de tweede testafname, schematisch weergegeven.

Tabel XIV Gemiddelde CU-scores interbeoordelaarsbetrouwbaarheid; 20 scenario's

Beoordelaar	Gemiddelde CU-score op eerste testafname	Gemiddelde CU-score op tweede testafname
Dassek	103.17	109.50
Van den Berg	72.00	76.00
Zwartjens	67.17	70.50

#### 4.7 Interbeoordelaarsbetrouwbaarheid ANTAT-CU4; 23 scenario's

##### 4.7.1 Vergelijking tussen de CU-scores van de auteur van deze scriptie en Dassek (2016)

Voor het eerste testafnamemoment was de ICC-waarde .08 (zwak; Koo & Li, 2016),  $F(5,5) = 2.01$ ,  $p = .232$ . Voor het tweede testafnamemoment was de ICC-waarde .16 (zwak; Koo & Li),  $F(5,5) = 7.22$ ,  $p = .024$ .

Op de eerste testafname werd door de auteur van deze scriptie een significant lager gemiddelde CU-score gegeven van 75.5 ( $SD=13.25$ ) dan door Dassek (2016), die een gemiddelde CU-score gaf van 113.67 ( $SD=5.28$ ),  $t(5) = -8.04$ ,  $p < .001$ . Op de tweede testafname werd door de auteur van deze scriptie wederom gemiddeld een significant lagere CU-score gegeven dan door Dassek. Zo werd door de auteur van deze scriptie een gemiddelde CU-score gegeven van 80.33 ( $SD=12.83$ ), terwijl Dassek een gemiddelde CU-score gaf van 120.33 ( $SD=7.01$ ),  $t(5) = -13.59$ ,  $p < .001$ .

##### 4.7.2 Vergelijking tussen de CU-scores van Dassek (2016) en Van den Berg

Voor het eerste toetsmoment was de ICC-waarde .09 (zwak; Koo & Li, 2016),  $F(5,5) = 2.04$ ,  $p = .227$ . Voor het tweede toetsmoment was de ICC-waarde .22 (zwak; Koo & Li),  $F(5,5) = 6.47$ ,  $p = .031$ .

Op de eerste testafname werd door Dassek een significant hogere gemiddelde CU-score gegeven, van 113.67 ( $SD=5.28$ ), dan door Van den Berg, die een gemiddelde CU-score gaf van 79.83 ( $SD=12.38$ ),  $t(5) = 7.59$ ,  $p = .001$ . Op de tweede afname werd door Dassek eveneens een significant hogere gemiddelde CU-score gegeven, van 120.33 ( $SD=7.01$ ), dan door Van den Berg, die een gemiddelde CU-score gaf van 85.00 ( $SD=13.93$ ),  $t(5) = 10.73$ ,  $p < .001$ .

#### 4.7.3 Vergelijking tussen de CU-scores van de auteur van deze scriptie en Van den Berg

Voor het eerste testafnamemoment was de ICC-waarde .97 (uitmuntend; Koo & Li, 2016),  $F(5,5) = 213.48$ ,  $p < .001$ . Voor het tweede testafnamemoment was de ICC-waarde .96 (uitmuntend; Koo & Li),  $F(5,5) = 213.48$ ,  $p < .001$ .

Uit gepaarde t-toetsen, waarmee de gemiddelde CU-scores van Van den Berg en de auteur van deze scriptie werden vergeleken, kwam naar voren dat bij de eerste testafname Van den Berg een significant hogere gemiddelde CU-score gaf ( $M=79.83$ ,  $SD=12.38$ ) dan de auteur van deze scriptie ( $M=75.5$ ,  $SD=13.25$ ),  $t(5) = 6.06$ ,  $p = .002$ . Ook op de tweede testafname werd door Van den Berg een significant hogere gemiddelde CU-score gegeven ( $M=85.00$ ,  $SD=13.93$ ) dan door de auteur van deze scriptie ( $M=80.33$ ,  $SD=12.83$ ),  $t(5) = 3.44$ ,  $p = .019$ .

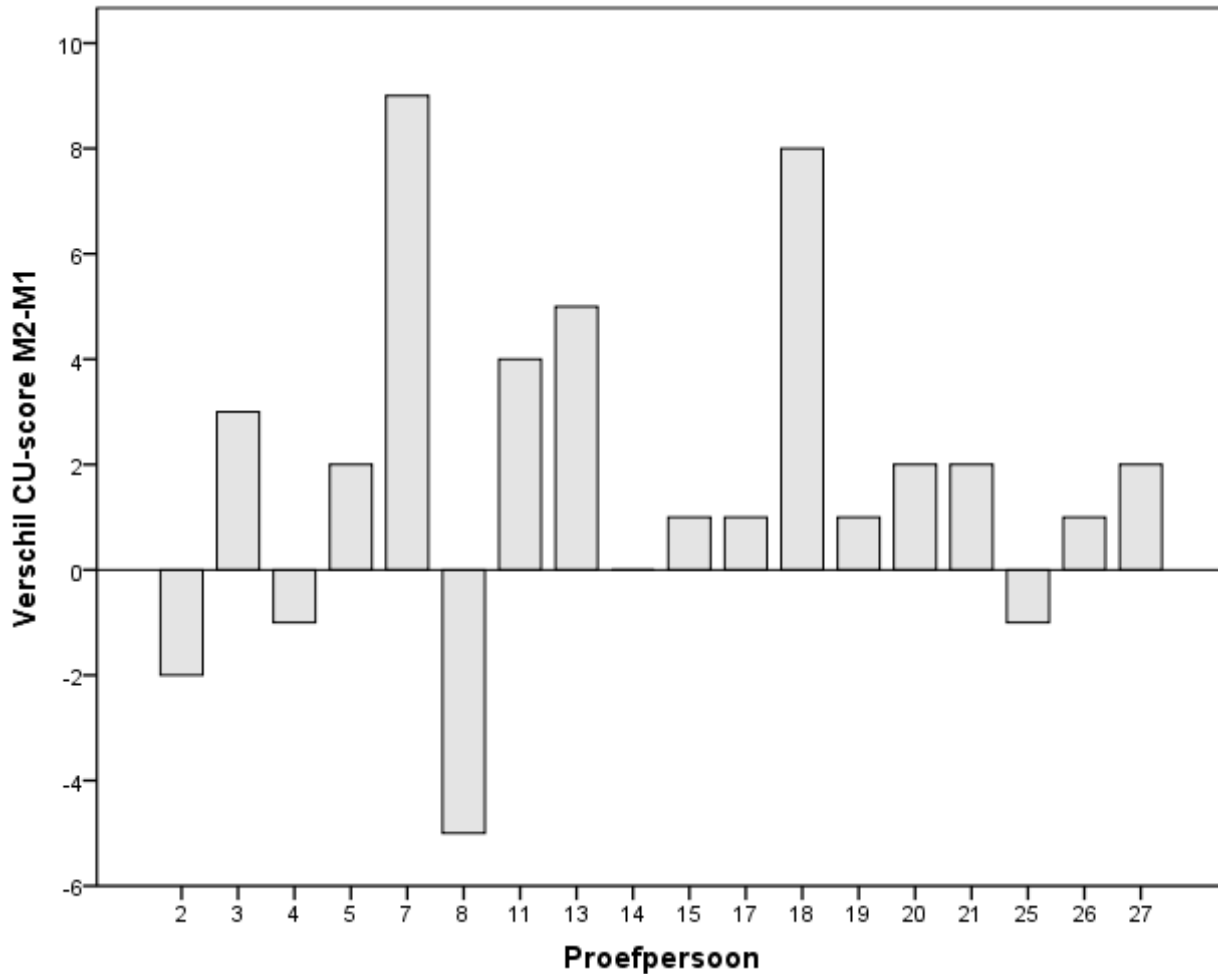
In Tabel XV staan de gemiddelde CU-scores, die gegeven werden door de beoordelaars op de eerste en de tweede testafname, voor 23 scenario's, schematisch weergegeven.

Tabel XV Gemiddelde CU-scores interbeoordelaarsbetrouwbaarheid; 23 scenario's

Beoordelaar	Gemiddelde CU-score op eerste testafname	Gemiddelde CU-score op tweede testafname
Dassek	113.67	120.33
Van den Berg	79.83	85.00
Zwartjens	75.50	80.33

#### 4.8 Intrabeoordelaarsbetrouwbaarheid; 20 scenario's

Uit de analyse bleek dat door de auteur van deze scriptie op het tweede meetmoment aan de 18 proefpersonen gemiddeld een hogere CU-score was gegeven ( $M=70.28$ ,  $SD=11.33$ ) dan op het eerste meetmoment ( $M=68.50$ ,  $SD=13.16$ ),  $t(17) = -2.26$ ,  $p = .037$ . De correlatie tussen de ANTAT-CU4 score op het eerste en tweede meetmoment was bijzonder goed (Hutcheson & Sofroniou, 1999), namelijk .977 met een significantie van  $p < .001$ . In totaal werd aan 13 deelnemers op het tweede meetmoment een hogere totale CU-score toegekend dan op het eerste meetmoment. Vier proefpersonen kregen een lagere totale CU-score en bij een proefpersoon was de gegeven totale CU-score op beide meetmomenten gelijk. In Figuur 3 worden de verschillen in CU-scores op het eerste en tweede meetmoment weergegeven. In het figuur is te zien hoe de proefpersonen op het tweede meetmoment scoorden ten opzichte van het eerste meetmoment (zie Tabel XVI in Bijlage R met ruwe CU-scores op het eerste en tweede meetmoment).



Figuur 3. Verschil in behaalde CU-scores op 20 scenario's op het tweede meetmoment (M2) ten opzichte van het eerste meetmoment (M1)

Uit een gepaarde t-toets met de gemiddelde CU-scores van de 18 proefpersonen op alle afzonderlijke scenario's op het eerste en tweede meetmoment, kwam naar voren dat alleen op scenario 2 "voetballen" op het tweede meetmoment gemiddeld significant hogere CU-scores waren gegeven ( $M=3.67$ ,  $SD=1.82$ ) dan op het eerste meetmoment ( $M=3.44$ ,  $SD=1.92$ ),  $t(17) = -2.20$ ,  $p < .05$ . De gegeven gemiddelde CU-scores op het eerste en het tweede toetsmoment van alle andere scenario's waren niet significant verschillend. In Tabel XVII staan de gegeven CU-scores op scenario 2 "voetballen" van de 18 proefpersonen op het eerste en het tweede meetmoment weergegeven. In de tabel is te zien dat bij vier van de 18 proefpersonen andere punten waren toegekend op scenario 2. Deze vier proefpersonen behaalden bij het tweede meetmoment hogere CU-scores dan bij het eerste meetmoment.

Tabel XVII Verschil in gegeven CU-scores ( $n=18$ ) voor scenario 2 “voetballen” op het eerste (M1) en tweede meetmoment (M2).

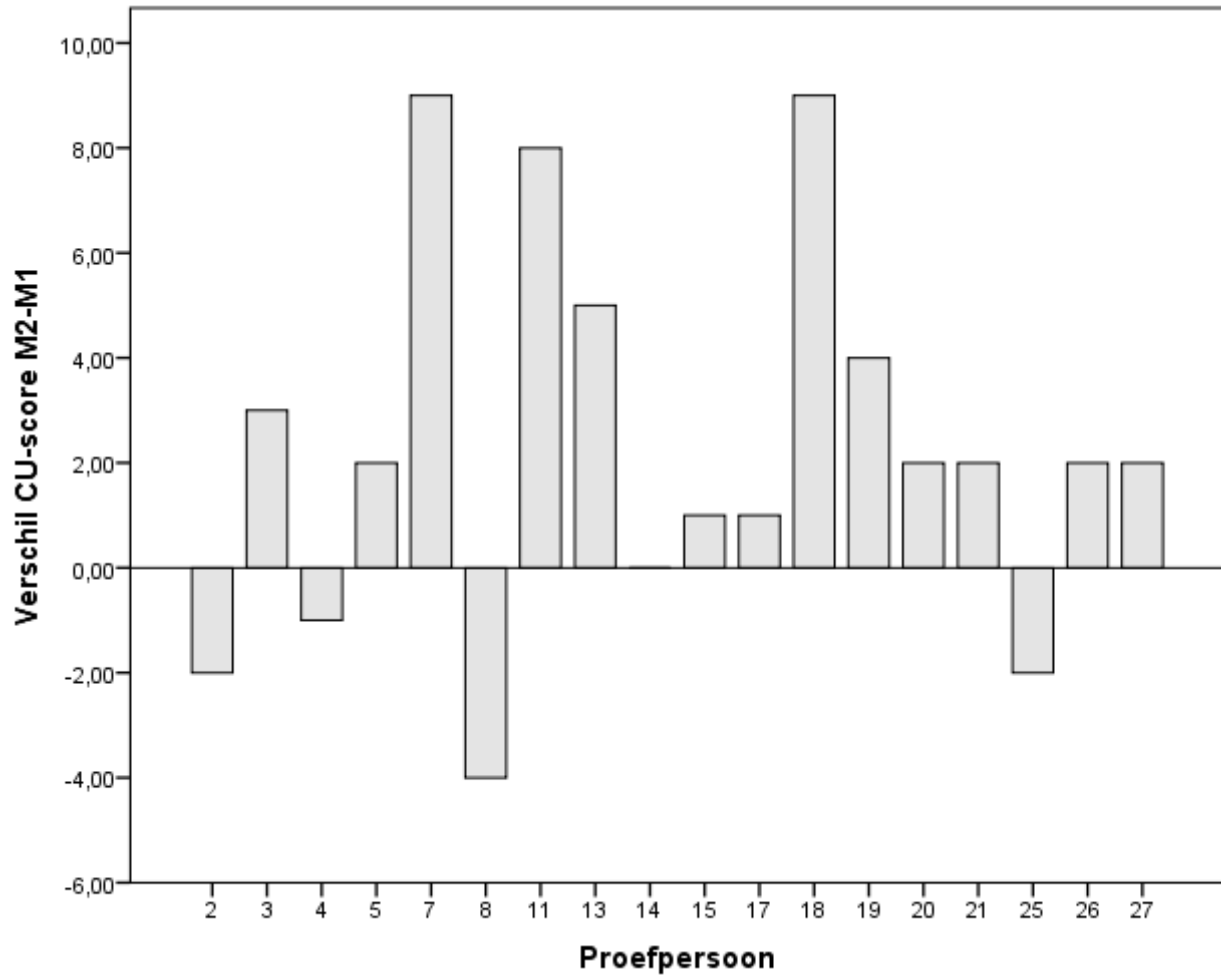
Proefpersoon	Scenario 2 – M1	Scenario 2 – M2
2	4	4
3	4	4
4	6	6
5	1	2
7	0	0
8	7	7
11	3	3
13	2	2
14	4	4
15	6	6
17	3	3
18	3	3
19	4	4
20	3	4
21	4	4
25	0	1
26	5	6
27	3	3

Na verwijdering van scenario 2, bleek uit een gepaarde t-test dat er geen significant verschil was in de gegeven gemiddelde CU-scores door de auteur van deze scriptie op de verschillende meetmomenten. Op het eerste meetmoment was de gemiddelde gegeven CU-score 65.06 ( $SD=12.29$ ) en op het tweede meetmoment was de gemiddelde gegeven CU-score 66.61 ( $SD=10.65$ ),  $t(17) = -1.93$ ,  $p = .070$ ;  $r = .97$ ,  $p < .001$ .

#### 4.9 Intrabeoordelaarsbetrouwbaarheid; 23 scenario's

Uit de analyse bleek dat door de auteur van deze scriptie op het tweede meetmoment aan de 18 proefpersonen gemiddeld een hogere CU-score was gegeven ( $M=78.83$ ,  $SD=15.14$ ) dan op het eerste meetmoment ( $M=76.56$ ,  $SD=13.13$ ),  $t(17) = -2.63$ ,  $p = .018$ . De correlatie tussen de ANTAT-CU4 score op het eerste en tweede meetmoment was bijzonder goed (Hutcheson & Sofroniou, 1999), namelijk .976 met een significantie van  $p < .001$ .

In totaal werd aan 13 deelnemers op het tweede meetmoment een hogere totale CU-score toegekend dan op het eerste meetmoment. Vier proefpersonen kregen een lagere totale CU-score en bij een proefpersoon was de gegeven totale CU-score op beide meetmomenten gelijk. In Figuur 4 worden de verschillen in CU-scores op het eerste en tweede meetmoment weergegeven. In het figuur is te zien hoe de proefpersonen op het tweede meetmoment scoorden ten opzichte van het eerste meetmoment (zie Tabel XVIII in Bijlage R met ruwe CU-scores op het eerste en tweede meetmoment).



Figuur 4. Verschil in behaalde CU-scores op 23 scenario's op het tweede meetmoment (M2) ten opzichte van het eerste meetmoment (M1).



## 5. Discussie

Het doel van het onderzoek was het toetsen van verschillende betrouwbaarheidsmaten van de ANTAT-CU4, zodat het testinstrument verder ontwikkeld kan worden en zodoende in de toekomst gebruikt kan worden om de verbale functionele communicatievaardigheden van cliënten met afasie te toetsen.

In dit hoofdstuk worden de gevonden resultaten omtrent de betrouwbaarheid van de ANTAT-CU4 met 20 scenario's geïnterpreteerd en bediscussieerd. Dit wordt gedaan met behulp van een terugblik op de onderzoeksvragen, die in de inleiding zijn opgesteld. De reden dat is gekozen om de ANTAT-CU4 met 20 scenario's te bespreken, is dat het testinstrument uit 20 scenario's bestaat en dat de versie met 23 scenario's alleen voor het huidige onderzoek werd gebruikt. Eerst worden de betrouwbaarheidsmaten van de ANTAT-CU4 besproken. Daarna wordt er kritisch gekeken naar de onderzoeksopzet en de scenario's.

### 5.1 Wat is de mate van de interne consistentie van de ANTAT-CU4?

De verwachting was dat uit *reliability analyses* naar voren zou komen dat de waarde van Cronbachs alfa op beide testafnamemomenten .80 of hoger zou zijn (goed; Van Wijk, 2000). Op zowel het eerste als het tweede testafnamemoment kwam de waarde van Cronbachs alfa niet boven de verwachte .80. Op het eerste testafnamemoment was de Cronbachs alfawaarde voor de interne consistentie .754, wat volgens Sijsma (2009) te laag is om van een goede betrouwbaarheid te spreken. Tevens kwam naar voren dat drie scenario's minder goed leken te passen bij de overige scenario's. Het betrof scenario 1 "stomerij", scenario 12 "verjaardag" en scenario 20.2 "gemeente". Wanneer elk van deze scenario's uit de test verwijderd zou worden, zou de waarde van de Cronbachs alfa licht stijgen (zie Tabel IX voor een overzicht). De Heus, Van der Leeden en Gazedam (1995) geven als vuistregel aan om een item uit een test te verwijderen, indien de alfawaarde in dat geval met meer dan .05 zou stijgen. In het geval van scenario 1, 12 en 20.2 is de stijging van de alfawaarde na het verwijderen van deze items lager dan .05. Als alle drie de scenario's uit de test verwijderd zouden worden, stijgt de Cronbachs alfa met .016. De waardeverschillen zijn minimaal. Om deze reden is het advies om de scenario's niet te verwijderen.

Scenario 1 "stomerij" paste mogelijk minder goed bij de overige scenario's, omdat deze in de helft van de gevallen als eerste item werd aangeboden. Deze hypothese lijkt echter niet helemaal op te gaan, omdat voorafgaand aan de testafname eerst oefenopgaven werden afgenomen. Een andere mogelijkheid is dat bij scenario 1 gebruik werd gemaakt van een afbeelding, maar andere scenario's met ondersteunende afbeeldingen zijn wel passend bij de rest van de test. Als laatste zou de formulering van de instructie bij scenario 1 een rol kunnen spelen. In de instructie wordt namelijk niet expliciet benoemd wat er precies aan de hand is, namelijk dat een vlek niet schoongemaakt is.

Op het tweede testafnamemoment was de waarde van Cronbachs alfa .687 (matig; Van Wijk, 2000). Dit houdt in dat de scenario's niet allemaal hetzelfde leken te meten, waardoor de test mogelijk minder valide is. Daarnaast leken scenario 6 "slager", 11.2 "bril", 12 "verjaardag", 17 "station" en 20.2 "gemeente" minder goed te passen bij de overige scenario's. Wanneer scenario 6, 12 of 20.2 verwijderd zou worden, stijgt de Cronbachs alfawaarde minder dan .05 (zie Tabel IX). Zelfs wanneer ze alle drie uit de test verwijderd zouden worden, stijgt de waarde enkel van .687 naar .706. Volgens de vuistregel van De Heus et al. (1995) zouden deze scenario's dus niet

uit de test verwijderd hoeven te worden. Wat opvalt is dat scenario 12 en 20.2 op beide testafnamen minder passend lijken te zijn bij de rest van de test. Scenario 12 “verjaardag” zou minder passend kunnen zijn, omdat het een sociaal gebeuren betreft, hoewel er ook andere scenario’s met een sociaal thema zijn (o.a. scenario 7 “vriend afspreken” en scenario 9 “nieuwe buurman”). Scenario 20.2 past mogelijk minder goed bij de rest, omdat dit scenario, met een licht afwijkende instructie, tweemaal voorkwam in de testafname. Scenario 6 heeft een soortgelijke instructie als scenario 1, waarbij het probleem niet expliciet benoemd wordt. Mogelijk is dit de reden waarom dit item minder goed bij de overige scenario’s lijkt te passen.

Wanneer scenario 11.2 of scenario 17 verwijderd zou worden, stijgt de Cronbachs alfawaarde ook minder dan .05, maar bij deze scenario’s is sprake van een negatieve item-totaalcorrelatie (zie Tabel IX). Ondanks dat niet aan de voorwaarde van De Heus et al. wordt voldaan, zou de test wellicht toch meer valide worden wanneer deze items worden verwijderd, omdat deze scenario’s een negatief effect lijken hebben op de onderlinge samenhang. Tevens zou de Cronbachs alfawaarde van .687 stijgen naar .739 indien beide scenario’s uit de test verwijderd worden. Dit zou neerkomen op een verbetering van matig naar adequaat (Van Wijk, 2000). Een reden waarom scenario 11.2 wellicht minder goed past bij overige scenario’s, is dat bij dit item gebruik werd gemaakt van een ondersteunende afbeelding. Een mogelijke verklaring waarom scenario 17 minder goed paste bij de rest van de test kan zijn dat de testinstructie bij dit item langer was dan bij de overige scenario’s.

Opvallend is, dat scenario 2 “voetballen”, 15 “huisarts” en 18 “kat”, waarbij de maximale haalbare CU-score per persoon kan verschillen, geen negatief effect hebben op de interne consistentie van de ANTAT-CU4. Variabele mogelijke CU-scores lijken dus niet van invloed te zijn op de validiteit van het testinstrument.

Kort samengevat is op beide testafnamemomenten de interne consistentie van de ANTAT-CU4 bij gezonde sprekers te laag om van een goede betrouwbaarheid te kunnen spreken (Van Wijk, 2000; Sijtsma, 2009). Mogelijk hebben de plaats van een item in de test, instructies en ondersteunende afbeeldingen bij de scenario’s invloed op de interne consistentie. In vervolgonderzoek zal kritisch naar de scenario’s van de ANTAT-CU4 gekeken moeten worden om na te gaan waarom bepaalde scenario’s minder goed bij de overige scenario’s lijken te passen.

## **5.2 Wat is de mate van de test-hertestbetrouwbaarheid van de ANTAT-CU4 bij gezonde sprekers?**

Naast de interne consistentie werd de test-hertestbetrouwbaarheid van de ANTAT-CU4 gemeten. Dit werd gedaan door na te gaan of de scores van de participanten, wat betreft verbale effectiviteit en efficiëntie, stabiel waren over de tijd heen. Er werd verwacht dat de verbale effectiviteit stabiel zou blijven over de tijd heen, terwijl de verbale efficiëntie zou toenemen. Hoewel uit het onderzoek naar voren kwam dat de verbale effectiviteit over de tijd heen gemiddeld toenam, was dit verschil niet significant<sup>3</sup>. Wat betreft de verbale efficiëntie, bleek dat de proefpersonen tijdens de tweede testafname gemiddeld significant hoger scoorden dan tijdens de eerste testafname.

---

<sup>3</sup> Wanneer dezelfde analyses werden toegepast om de effectiviteit te onderzoeken op de ANTAT-CU4-versie met 23 scenario’s, bleek het hogere percentage geproduceerde Content Units op de tweede afname wel significant verschillend te zijn ten opzichte van de eerste afname.

### 5.2.1 Effectiviteit

Uit de analyse kwam naar voren dat de proefpersonen op de tweede testafname gemiddeld genomen een hoger percentage Contents Units benoemden dan op de eerste afname, hoewel dit verschil niet significant bleek te zijn. Dit komt onder andere overeen met de resultaten uit het onderzoek van Dassek (2016). Uit haar onderzoek bleek ook geen significant verschil te zijn in de gegeven CU-scores op het eerste en het tweede testafnamemoment. Uit onderzoek waarbij werd gekeken naar herhalingseffecten bij het navertellen van verhalen (Linebarger, McGall, Virata & Berndt, 2007) kwam naar voren dat gezonde sprekers bij herhaalde metingen geen significante verschillen lieten zien in het percentage geproduceerde relevante informatie-eenheden (%CIUs). Er leek geen sprake te zijn van een herhalingseffect. Linebarger et al. geven aan dat zelfs wanneer een spreker dus bekend is met een verhaal, dit niet betekent dat er ook voldoende verwerkingsmechanismen in gang worden gezet om aspecten wat betreft de productie van structuur en inhoud te beïnvloeden. Brookshire en Nicholas (1994a; 1994b) hebben de test-hertestbetrouwbaarheid voor “*connected speech*” onderzocht bij gezonde sprekers en bij sprekers met afasie, waarbij ze ook onder andere keken naar het percentage correct geproduceerde relevante informatie-eenheden (%CIUs). Met *connected speech* wordt een reeks aaneengeschakelde verbale uitingen bedoeld, ofwel gesproken taal zoals in een conversatie voorkomt. Uit beide onderzoeken van Brookshire en Nicholas bleek dat de gezonde proefpersonen gemiddeld genomen op de tweede testafname een hoger percentage correct geproduceerde relevante informatie-eenheden lieten zien dan op de eerste testafname. Volgens Brookshire en Nicholas is dit verschil te wijten aan het aantal gegeven stimuli (1994a) en de lengte van het taalsample (1994b). Ze benoemden echter niet of er sprake was van een significant verschil.

### 5.2.2 Efficiëntie

Zoals werd verwacht, bleek uit analyse dat er sprake was van een significant verschil in het gemiddelde aantal geproduceerde Content Units per minuut op het eerste en tweede toetsmoment. Op het tweede toetsmoment werden door de proefpersonen gemiddeld significant meer Contents Units per minuut geproduceerd dan op het eerste toetsmoment. Onderzoek van Linebarger et al. (2007) toonde aan dat gezonde sprekers bij een vierde testafname gemiddeld significant meer woorden per minuut produceerden dan bij een eerdere testafname. Of tegelijkertijd ook meer relevante informatie-eenheden (CIUs) per minuut worden geproduceerd, wanneer meer woorden worden gebruikt, wordt in het onderzoek echter niet benoemd. Bij het onderzoek van Linebarger et al. werd dezelfde meting viermaal uitgevoerd. Mogelijk kan dit onderzoek dus niet direct met het huidige onderzoek worden vergeleken. Onderzoek van Brookshire en Nicholas (1994b) heeft aangetoond dat bij een test-hertestprocedure gezonde sprekers op de tweede testafname gemiddeld genomen meer relevante informatie-eenheden (CIUs) produceerden dan op de eerste testafname. Echter, wordt in hun onderzoek wederom niet aangegeven of dit verschil al dan niet significant is.

Concluderend bleek dat de verbale effectiviteit over de tijd heen niet significant toenam. Dit was in lijn met de verwachting en met Linebarger et al. (2007). Wat betreft de verbale efficiëntie, bleek dat tijdens het tweede testafnamemoment de proefpersonen gemiddeld significant hoger scoorden dan tijdens het eerste testafnamemoment. Dit was eveneens in lijn met de verwachting en kwam overeen met onderzoek van Brookshire en Nicholas (1994b), hoewel zij in hun

onderzoek niet aantonen of de toename van relevante informatie-eenheden significant is. De test-hertestbetrouwbaarheid bij gezonde sprekers lijkt dus voldoende te zijn.

### 5.2.3 Begrip van testinstructie

Soms kwam het voor dat een proefpersoon een testinstructie niet goed had begrepen. De proefpersoon gaf dan zelf aan dat hij de vraag niet goed had begrepen, vroeg om herhaling van de instructie of gaf een foutieve respons waarna de testleider de instructie nogmaals herhaalde.

#### 5.2.3.1 Begrip van testinstructie dubbele scenario's

Op basis van het al dan niet goed begrijpen van een scenario, is het vrij lastig om te bepalen welke versie van de dubbele scenario's het beste in de ANTAT-CU4 kan worden opgenomen. Hoewel bij scenario 11.1 maar een proefpersoon was, die niet doorhad dat het pootje van de bril zodanig scheef stond dat de bril niet meer gedragen kon worden, zou ik adviseren om toch voor scenario 11.2 te kiezen, omdat deze door alle proefpersonen goed begrepen werd. Dassek (2016) gaf ook aan de voorkeur te hebben voor scenario 11.2, omdat enkele proefpersonen niet goed zagen dat het pootje verbogen was. Tevens gaf Dassek aan dat een aantal proefpersonen benoemden dat ze voor een scheef pootje niet naar de opticien zouden gaan.

Bij scenario 16 maakte de gegeven testinstructie niet uit voor het begrijpen van het scenario. Toch gaat, op basis van de reacties van de proefpersonen, de voorkeur uit naar versie 16.2. De formulering van deze instructie zorgt ervoor dat de sprekers meer geneigd zijn een rollenspel te spelen in plaats gebruik te maken van indirecte rede. Voorbeelden (8) en (9) met het laatste deel van de instructies van 16.1 en 16.2 en de reacties van een van de proefpersonen zullen dit mogelijk verduidelijken.

- (8) 16.1 *“Hij vraagt wat er gebeurd is. Wat zegt u?”*  
*“Dat ik van de fiets gevallen ben en dat ik m'n arm in het gips heb”*
- (9) 16.2 *“Leg aan hem uit wat er met u gebeurd is”*  
*“Ik ben van m'n fiets gevallen en helaas m'n arm is gebroken”*

Hoewel een reactie, zoals in voorbeeld (8) maar bij enkele proefpersonen voorkwam, zouden meer personen mogelijk op eenzelfde manier kunnen reageren. Om zoveel mogelijk te zorgen voor situaties die in het dagelijks leven voor zouden kunnen komen, zou ik aanraden om formulering 16.2 te gebruiken. Dit in tegenstelling tot Dassek (2016), die aangaf dat de formulering van 16.1 mogelijk beter zou zijn, omdat deze minder belastend zou zijn voor het werkgeheugen. Personen met afasie zouden deze formulering dan mogelijk gemakkelijker kunnen begrijpen. Onderzoek met Nederlandstalige sprekers met afasie, zou hier meer verduidelijking in kunnen brengen.

Wat betreft scenario 20, hadden meerdere participanten moeite met het juist interpreteren van de testinstructie. Op basis van de reacties van de proefpersonen kan geen eenduidige voorkeur uitgesproken worden voor versie 20.1 of 20.2. Bij zeven proefpersonen is op advies van Dassek (2016) een derde versie van scenario 20 afgenomen, namelijk 20.3. De instructie van scenario 20.3 was als volgt: *“U belt naar de gemeente. U wilt iemand van de gemeente spreken, maar blijkbaar klopt het nummer niet. Per ongeluk krijgt u de tandarts aan de lijn. Wat zegt u?”*. Hoewel

het bij deze instructie voor de proefpersonen duidelijk was dat ze niet met iemand van de gemeente spraken, werd niet door de proefpersonen gereageerd zoals werd verwacht. Soms legden ze uit wat ze in die situatie zouden doen, in plaats van direct te reageren. Dit is vergelijkbaar met scenario 16.1, te zien in voorbeeld (8). Mijn advies is dan ook om de volgende testinstructie te gebruiken: *“U belt naar de tandarts (om een afspraak te maken), maar blijkbaar klopt het nummer niet. Per ongeluk krijgt u mij aan de lijn. Wat zegt u?”*. Deze instructie wijkt maar licht af van de versie 20.2, zodat de Content Units waarop wordt gescoord, zoveel mogelijk gelijk blijven. Daarnaast is het met deze instructie voor de persoon mogelijk duidelijker dat hij een verkeerd nummer heeft gebeld en niet dat hij nog verder doorverbonden moet worden. Ook is deze zinconstructie niet te lang, waardoor het werkgeheugen zo min mogelijk wordt belast. In vervolgonderzoek kan mogelijk nagegaan worden of deze constructie door de deelnemers beter wordt begrepen.

### **5.2.3.2 Begrip van testinstructie overige scenario's**

In Tabel XIII is te zien dat enkele deelnemers moeite hadden met het begrijpen van scenario 9 “nieuwe buurman”, 13 “afschrift” en 15 “huisarts”. Na het herhalen van de instructie werden deze scenario's wel door alle proefpersonen begrepen.

Scenario 18 “kat” werd op beide testafnamemomenten door een proefpersoon mogelijk niet goed begrepen. De proefpersoon reageerde op het eerste testafnamemoment anders dan verwacht op het betreffende scenario, maar niet per se onjuist. De testinstructie bij scenario 18 is als volgt: *“Er komt bij u thuis een kat aangelopen. Om zijn nek zit een bandje met een telefoonnummer. U belt op en zegt?”*. De instructie zou zodanig geïnterpreteerd moeten worden dat de proefpersoon doet alsof hij de eigenaar van de kat heeft opgebeld. Deze proefpersoon reageerde op het eerste testafnamemoment echter als volgt: *“Spreek ik met de dierenbescherming? Er is een zojuist een kat bij mij komen binnenlopen en er zit een nummer op zijn nek. Hij is niet van mij, maar ik zou hem graag bij de rechtmatige eigenaar terugbrengen.”*. Tijdens de tweede testafname begon het antwoord van de proefpersoon als volgt: *“Goedemorgen, spreek ik met de politie?”*. De testleider heeft beide keren het scenario niet voor de proefpersoon herhaald. Hierdoor is het onduidelijk of de proefpersoon het scenario daadwerkelijk niet goed begrepen heeft.

Ondanks dat scenario's 15 en 18 mogelijk niet door alle proefpersonen direct werden begrepen, is het niet noodzakelijk de scenario's of de instructie aan te passen. Na herhaling van de instructie, werd scenario 15 wel begrepen door de proefpersoon en bij scenario 18 is onduidelijk of de proefpersoon de instructie niet goed had begrepen of dat de instructie op een andere manier door de proefpersoon werd geïnterpreteerd. Scenario 13 werd door meerdere deelnemers niet direct goed begrepen. Mogelijk kwam dit doordat de testinstructie langer is dan de instructies van de overige scenario's en doordat er twee afbeeldingen zijn ter ondersteuning van het scenario. Door de testinstructie licht aan te passen, wordt het mogelijk al gemakkelijker voor de deelnemers om het scenario direct te begrijpen. Zo kan bijvoorbeeld door de testleider worden benoemd dat de deelnemer “een boek” heeft gekocht in plaats van “iets”. Deze kleine wijziging kan ervoor zorgen dat de afbeelding al gemakkelijker wordt gelezen. Daarnaast kan eventueel iets meer nadruk worden gelegd op het woord “belt” in de zin: *“U belt naar de winkel”*. Op deze manier is het voor de participant mogelijk duidelijker dat hij niet fysiek in de winkel aanwezig is. Een ander advies omtrent dit scenario is om deze langzaam uit te spreken en de deelnemers de tijd te geven om de afbeeldingen om de beurt te bekijken.

Bij scenario 9 was het opvallend dat deze juist bij de tweede testafname vaker verwarrend was voor de participanten. Mogelijk letten de proefpersonen bij de tweede testafname minder goed op de instructies, omdat ze deze al bij de eerdere testafname hadden gehoord. Om deze verwarring bij andere sprekers te voorkomen, zijn er twee mogelijkheden. Ten eerste zou de testinstructie kunnen worden uitgesproken met meer nadruk op het woord “u”. De nadruk komt dan op het onderstreepte woord te liggen: “U bent net verhuisd en komt uw nieuwe burens tegen. U wilt met ze kennis maken. Wat zegt u?”. Eventueel kan de testleider tijdens het uitspreken van het woord “u” gebaren naar de participant. Ten tweede zou de testinstructie licht aangepast kunnen worden, zodat duidelijker wordt benadrukt dat de spreker degene is die is verhuisd. Een voorstel voor de instructie zou zijn: “U bent net verhuisd naar een andere stad en u komt uw nieuwe burens tegen. U wilt met ze kennis maken. Wat zegt u?”. Mogelijk zorgt het stukje “naar een andere stad” ervoor dat de deelnemers beter begrijpen dat zij de verhuisde personen zijn, maar het kan er ook voor zorgen dat het werkgeheugen van personen met afasie zwaarder wordt belast. In vervolgonderzoek zou ik dan ook adviseren om eerst na te gaan of de extra nadruk op het woord “u” met eventueel een gebaar al helpend is voor de participanten en dat het scenario zodoende al beter wordt begrepen voordat de testinstructie wordt aangepast.

### **5.3 Wat is de mate van interbeoordelaarsbetrouwbaarheid van de ANTAT-CU4?**

Om de interbeoordelaarsbetrouwbaarheid te onderzoeken hebben verschillende beoordelaars punten toegekend aan de prestaties van zes proefpersonen zonder afasie op de eerste en tweede testafname. De verwachting was dat de beoordelaars gelijke puntenaantallen zouden toekennen en dat er geen significant verschil zou zijn in de gegeven CU-scores. Tijdens het vergelijken van de gegeven scores door Dassek (2016) en de auteur van deze scriptie, bleek de ICC-waarde zeer zwak te zijn. Ook was er een significant verschil in de gegeven scores. Vanwege deze grote discrepantie werd een derde onafhankelijke beoordelaar, Van den Berg, ingeschakeld. De ICC-waarde tussen Dassek en Van den Berg was eveneens zeer zwak. De ICC-waarde tussen de auteur van deze scriptie en Van den Berg was daarentegen uitmuntend. Dat wil zeggen dat er een grote overeenstemming was in de manier van scoren. Desondanks waren de gegeven scores significant verschillend. Dit was niet in lijn met de verwachting.

#### **5.3.1 Interbeoordelaarsbetrouwbaarheid auteur van deze scriptie en Dassek (2016)**

De gegeven CU-scores door Dassek (2016) en de auteur van deze scriptie waren niet alleen significant verschillend, tevens was de correlatie tussen de beoordelaars zeer laag. Door de grote discrepantie in de beoordelingen konden geen betrouwbare conclusies worden getrokken aangaande het verschil in de gegeven CU-scores. Hoewel beide beoordelaars voor het scoren gebruik hebben gemaakt van het CU-schema dat door Dassek (2016) werd opgesteld, kan een verschil in interpretatie voor verschillende manieren van scoren hebben gezorgd. Dassek heeft in haar onderzoek namelijk het ANTAT-CU4-schema met de bijbehorende Content Units vastgesteld. Zij had hierdoor meer achtergrondinformatie ter beschikking over het opstellen van het ANTAT-CU4-schema, waardoor zij mogelijk in de geproduceerde informatie van de proefpersonen sneller een Content Unit herkende dan een beoordelaar die zich minder had verdiept in het opstellen van Content Units. De auteur van de huidige scriptie heeft onder andere in onderzoek van Ruiters et al. (2011) Filipinski (2014), Aan de Stegge (2015), Giessen (2015) en Dassek (2016) kunnen lezen hoe de ontwikkeling van de ANTAT-CU4 is verlopen. Hierdoor werd

kennis opgedaan omtrent het opstellen van het CU-schema met de bijbehorende Content Units. Desondanks was het voor de auteur van deze scriptie soms moeilijk om in te schatten of een bepaald genoemd Content Unit al dan niet goedgekeurd mocht worden volgens de voorschriften. Zo mocht bijvoorbeeld bij scenario 11 “bril” het Content Unit “dit” goedgekeurd worden, mits door de spreker naar de bril op de bijbehorende afbeelding werd gewezen. Zie Figuur 5 voor een voorbeeld.

Scenario 11: Bril						Score	5
A	66	67	EN V	68	69	70	
A1	pootje, bril	scheef, verbogen, kapot	V1	u	bril, <b>die, dit,</b>	repareren, maken	
OF A2	bril	correctie, aanpassen	OF V2	u	mij	helpen	
			OF V3	ik	nieuwe bril	willen	
Opmerkingen: V1: “die/dit” mag gebruikt worden, mits er in de aanleiding “bril” geïntroduceerd werd, of als er duidelijk naar de bril op de afbeelding gewezen wordt.							

Figuur 5. Scenario 11 “bril” met voorbeeld verwijzwoord “die/dit”

Omdat er alleen audio-opnamen beschikbaar waren, kon niet worden achterhaald of er daadwerkelijk naar de afbeelding werd gewezen. Om deze reden is door de auteur van deze scriptie besloten om in dit geval geen score toe te kennen, hoewel de testleider hiervoor mogelijk wel een punt heeft toegekend. Ook kan een geproduceerde Content Unit door verschillende beoordelaars mogelijk anders opgevat worden, waardoor deze niet altijd als een correcte informatie-eenheid wordt gezien. De mate van ervaring en het hierdoor anders beoordelen van de prestaties van de proefpersonen kan een mogelijke verklaring zijn voor de grote discrepantie in het scoren.

Hieronder worden eerst de gegeven scores door Dassek met de gegeven scores door Van den Berg met elkaar vergeleken. Daarna worden de overeenkomsten en verschillen in de gegeven scores gegeven door de auteur van de scriptie en Van den Berg besproken.

### 5.3.2 Interbeoordelaarsbetrouwbaarheid Dassek (2016) en Van den Berg

Tussen Dassek en Van den Berg was, net als tussen Dassek en de auteur van deze scriptie, een zeer lage correlatie en een significant verschil in de gegeven CU-scores. Mogelijk heeft hierbij de mate van ervaring in het gebruik van het CU-schema een rol gespeeld. En wellicht speelt wederom mee dat Dassek de onderzoeken zelf bij de proefpersonen heeft afgenomen en mogelijk scores heeft toegekend aan geproduceerde Content Units waarbij bijvoorbeeld gewezen moest worden. Van den Berg heeft, in de gevallen waarbij een Content Unit alleen correct gescoord mocht worden indien naar de afbeelding werd gewezen, geen punten toegekend, omdat zij niet kon nagaan of aan die voorwaarde werd voldaan.

### 5.3.3 Interbeoordelaarsbetrouwbaarheid auteur van de scriptie en Van den Berg

Van den Berg kende op beide meetmomenten gemiddeld significant hogere CU-scores toe dan de auteur van deze scriptie. Desondanks was er wel sprake van een uitstekende ICC-waarde op

beide testafnamemomenten. Dit betekent dat door beide beoordelaars op eenzelfde manier is gescoord, waarbij Van den Berg wel gemiddeld iets hogere scores toekende aan de prestaties van de proefpersonen dan de auteur van deze scriptie.

Het verschil in scores kan op verschillende manieren verklaard worden. Zo had de auteur van de scriptie zich meer ingelezen over het opstellen van het ANTAT-CU4-schema en over de terminologie rondom het begrip *Content Units*, omdat dit van belang was voor het onderzoek. Hierdoor ging de auteur van de scriptie mogelijk op een andere, wellicht zelfs kritischere, manier om met het toekennen van CU-scores. Daarnaast had de auteur van deze scriptie meer ervaring met het ANTAT-CU4-schema en het beoordelen van de prestaties van de sprekers. Dit kwam doordat ze niet alleen punten heeft gegeven aan de zes proefpersonen van wie de data werd gebruikt voor het onderzoeken van de interbeoordelaarsbetrouwbaarheid. Ze heeft namelijk de responsen van de gehele onderzoeksgroep, 31 proefpersonen, beoordeeld. Hierdoor had de auteur van deze scriptie meer ervaring in het gebruik van het ANTAT-CU4-schema en met het toekennen van CU-scores. Verder werden de instructies voor het gebruik van het ANTAT-CU4-schema mogelijk letterlijker toegepast door de auteur van deze scriptie. Een voorbeeld van scenario 10 “boeket bezorgen” zal dit verduidelijken. In Figuur 6 staat het CU-schema met mogelijke Content Units voor de aanleiding en het verzoek van scenario 10 weergegeven.

Scenario 10: Boeket bezorgen										Score	8
<b>A</b>	<b>58</b>	<b>59</b>	<b>60</b>	<b>61</b>	<b>EN V</b>	<b>62</b>	<b>63</b>	<b>64</b>	<b>65</b>		
<b>A1</b>	ik	bloemen, boeket	bezorgen, uitzoeken, bestellen	voor een kennis	<b>V1</b>	u, ik	<b>bloemen, boeket</b>	bezorgen	adres, prijs, naam		

Figuur 6. Scenario 10 “boeket bezorgen”

In het CU-schema is te zien dat Content Units voor aanleiding en verzoek soms uit dezelfde bewoordingen kunnen bestaan. Hierdoor was soms twijfel over het scoren van de Content Units, wanneer een Content Unit maar een keer werd benoemd. Een respons van een van de proefpersonen op de instructie van scenario 10 was: “Ik zou graag een boeketje willen laten bezorgen bij mevrouw Jansen voor ongeveer 20 euro in de kleur roze.” In deze productie zijn de aanleiding en het verzoek als het ware samengevoegd. Van den Berg had in dit voorbeeld alle Content Units (58-65, zie Figuur 6) als goed gescoord, waardoor de persoon op dit scenario in totaal acht punten kreeg. De auteur van deze scriptie had alleen Content Units 58-61 en 65 goed gerekend (zie Figuur 6), waardoor de persoon in totaal vijf punten kreeg. Mogelijk kan een andere beoordelaar ervoor kiezen om alleen de Content Units van het verzoek (62-65, zie Figuur 6) goed te rekenen. Voor zowel de auteur van deze scriptie als Van den Berg was het onduidelijk of de Content Units wel of niet dubbel geteld mochten worden. De auteur van deze scriptie heeft besloten minder Content Units goed te tellen, omdat deze niet twee keer apart werden benoemd. Van den Berg koos ervoor om de Content Units wel goed te tellen, omdat ze vond dat de uiting van de spreker zowel uit een aanleiding als een verzoek bestond.

Samenvattend kan gezegd worden dat de interbeoordelaarsbetrouwbaarheid momenteel nog onvoldoende is, hoewel de hoge ICC-waarden tussen de auteur van deze scriptie en Van den Berg suggereren dat er mogelijk wel een hoge interbeoordelaarsovereenstemming bereikt kan worden. Om dit te realiseren is het van belang om in vervolgonderzoek een



interbeoordelaarsbetrouwbaarheidsanalyse uit te voeren met beoordelaars die dezelfde mate van ervaring hebben met het gebruik van het ANTAT-CU4-schema. Daarnaast is het aan te raden om voor een duidelijke handleiding te zorgen, die bij twijfel handvaten biedt voor een eenduidige manier van scoren. Het advies is om na te gaan op welke manier relatief onervaren gebruikers van het ANTAT-CU4-schema CU-scores toekennen, zodat een inschatting gemaakt kan worden hoe het scoren in de praktijk zal gaan.

#### **5.4 Intrabeoordelaarsbetrouwbaarheid**

De intrabeoordelaarsbetrouwbaarheid werd onderzocht door de prestaties van 18 proefpersonen op de eerste testafname op twee momenten, met een tussenperiode van acht weken, te beoordelen. De prestaties werden door de auteur van deze scriptie beoordeeld. Uit de analyse bleek dat er een significant verschil was in de gegeven CU-scores op het eerste en het tweede meetmoment. Dit was niet in lijn met de verwachting. Op het tweede meetmoment werden gemiddeld hogere CU-scores gegeven dan op het eerste meetmoment. Echter, was de correlatie tussen deze meetmomenten wel zeer hoog.

In de handleiding van de ANTAT (Blomert et al., 1995) staat beschreven dat wanneer twee metingen sterk met elkaar samenhangen, er soms significante verschillen voor kunnen komen, omdat een t-toets dan gevoelig is voor minieme verschillen. Als de gemiddelde scores (bijna) gelijk zijn, moet deze significantie volgens Blomert et al. dan ook niet gezien worden als een daadwerkelijk verschil, maar als een statistisch artefact. Als onderbouwing geven zij aan dat een zeer hoge correlatie deze interpretatie aannemelijk maakt. Mogelijk is hetzelfde bij het huidige onderzoek aan de hand. De behaalde CU-scores op beide meetmomenten zijn nagenoeg gelijk en er was sprake van een zeer hoge correlatie (.974). Daarnaast bleek uit verdere analyses dat het significante verschil in de gegeven CU-scores toe te kennen was aan de scores op een scenario, namelijk scenario 2 "voetballen". Op dit scenario hadden vier van de 18 proefpersonen op het tweede meetmoment een andere waardering gekregen dan op het eerste meetmoment. Op alle andere scenario's werden op beide meetmomenten de prestaties van de deelnemers niet significant verschillend beoordeeld.

Een andere mogelijkheid voor het verschil in scoren op de twee meetmomenten is dat er sprake was van leereffect bij de beoordelaar. De prestaties van de proefpersonen werden met een tussentijd van acht weken beoordeeld. In deze periode werd gewerkt aan het onderzoek rondom de ANTAT-CU4 en het bijbehorende CU-schema. Mogelijk zouden de beoordelingen meer gelijk aan elkaar zijn geweest, indien de auteur van deze scriptie in de tussenliggende periode niet bezig was geweest met het onderwerp. Een goede handleiding zou wederom een uitkomst kunnen bieden, zodat meer eenduidig wordt gescoord.

#### **5.5 Onderzoeksopzet**

In het huidige onderzoek werd bij de proefpersonen tweemaal, met een tussenperiode van acht weken, de volledige ANTAT-CU4 afgenomen om de interne consistentie van de test te onderzoeken. Het nadeel van deze test-hertestprocedure is dat er bij de deelnemers een leereffect op kan treden. Mogelijk is de uitslag daardoor minder betrouwbaar (Carmines & Zeller, 1979). Het voorkomen van een leereffect was ook een van de redenen voor Blomert et al. (1995) om twee parallelle versies van de ANTAT te maken. Overigens zijn er ook testinstrumenten, zoals de Comprehensive Aphasia Test Nederlandstalige bewerking; CAT-NL, Swinburn, Porter & Howard,

2014) waarbij een kortere periode werd aangehouden (gemiddeld 28 dagen) tussen de eerste en tweede testafname voor het onderzoeken van de test-hertestbetrouwbaarheid (Kerkmeer & Heethuis, 2014).

Een andere manier om de betrouwbaarheid te onderzoeken is de *split-half*-methode. Hierbij wordt een test in twee parallelle versies verdeeld. Er is dan sprake van een test-hertestprocedure, waarbij de kans op een leereffect kleiner is. Wellicht had deze methode ook voor het huidige onderzoek gebruikt kunnen worden, omdat de ANTAT-CU4 net als de ANTAT (Blomert et al., 1995) uit twee parallelle versies met tien scenario's zal bestaan. Voorwaarde voor de *split-half*-methode is wel dat de twee parallelle versies gelijke ware scores en gelijke varianties hebben (Carmines & Zeller). Een onderzoeksinstrument kan namelijk op veel verschillende manieren gesplitst worden, hierdoor is exacte testparallelliteit moeilijk te realiseren (Carmines & Zeller). Toch is het in vervolgonderzoek noodzakelijk om de testparallelliteit van de ANTAT-CU4 te onderzoeken, wanneer het onderzoeksinstrument evenals de ANTAT uit 1995 uit twee parallelle versies zal bestaan.

## 5.6 Handleiding

Om verschillen in het toekennen van CU-scores te voorkomen en de overeenstemming tussen beoordelaars te vergroten, is het van belang dat er een duidelijke handleiding komt waarin staat aangegeven wanneer een Content Unit al dan niet goed gerekend mag worden. Het genereren van een handleiding was ook het advies van Dassek (2016). Volgens haar zou in de handleiding duidelijk moeten worden uitgelegd wat de betekenis is van 'propositie' en 'CU', zodat testleiders scores geven op basis van Content Units en niet worden verleid tot het scoren van 'woorden'.

De volgende aanbeveling heeft niet zozeer te maken met een handleiding voor de ANTAT-CU4, als wel met het scoreformulier. Wanneer een scoreformulier wordt ontwikkeld voor de ANTAT-CU4 is het wenselijk om per scenario ruimte over te laten voor opmerkingen of aantekeningen van de testleider. Hierin kan dan bijvoorbeeld beschreven worden of naar afbeeldingen werd gewezen. Eerder werd namelijk aangegeven dat er soms twijfel bestond of een Content Unit al dan niet goed gerekend mocht worden, omdat het voor de beoordelaars, die de test niet zelf hebben afgenomen, onduidelijk was of door de proefpersoon wel of niet naar een afbeelding werd gewezen. Wanneer de testleider deze handeling noteert, kan eventuele twijfel bij een derde beoordelaar voor het al dan niet correct scoren van een Content Unit worden weggenomen.

## 5.7 Aanbevelingen ten aanzien van het ANTAT-CU4-schema

Tijdens het scoren van de Content Units vielen bij enkele scenario's een aantal dingen op, die mogelijk nog binnen het ANTAT-CU4-schema verbeterd kunnen worden (zie bijlage S voor aangepaste ANTAT-CU4-schema's). Zo worden in scenario 3 "televisie" twee verschillende verzoeken beschreven. In verzoek 1 lijkt echter een inconsistentie te zitten. Dit verzoek kan bestaan uit drie Content Units (18, 19, 20). In Tabel XIX staan de Content Units weergegeven.

Tabel XIX Verzoek 1 scenario 3 "televisie"

Content Unit 18	Content Unit 19	Content Unit 20
u	advies / informatie / uitleg	geven
u	-	uitleggen
u	-	helpen

Wanneer de spreker een uiting produceert als: “Kunt u advies geven?”, dan worden alle drie de Content Units als correct gezien en scoort de spreker drie punten. Wanneer een spreker een zin produceert als: “Kunt u mij helpen?”, dan scoort de spreker twee punten, omdat twee van de drie Content Units worden genoemd. In Tabel XIX is te zien dat Content Unit 18 het subject is, Content Unit 19 het object en Content Unit 20 het verbum. De zin “Kunt u mij helpen?” bevat ook een subject “u”, object “mij” en verbum “helpen” en zou daarom, naar mijn mening, ook drie punten moeten krijgen. De Content Units “uitleggen” en “helpen” kwamen namelijk niet voor zonder het object “mij”. Bovendien bestaat verzoek 2 van scenario 11 “bril” ook uit de drie Content Units “u – mij – helpen” en wordt er wel een punt toegekend wanneer het Content Unit “mij” wordt geproduceerd. In Tabel XX staat weergegeven hoe de Content Units voor scenario 3 “televisie” zouden kunnen worden weergegeven.

Een soortgelijke situatie komt voor bij scenario 17 “station”. Bij dit scenario worden de Content Units “u – mij – helpen” op verzoek 2 alle drie goedgekeurd, wanneer de zin “Kunt u mij helpen?” wordt geproduceerd. Echter, kunnen bij dit verzoek maximaal vier punten behaald worden (zie Bijlage M). Het vierde Content Unit dat nodig is voor het laatste punt, past alleen niet helemaal bij de overige Content Units. Mogelijk zou bij dit scenario, dus net als bij scenario 2 “voetballen”, 15 ‘huisarts’ en 18 “kat”, een variabel aantal Content Units gescoord kunnen worden.

*Tabel XX Aanbeveling verzoek 1 scenario 3 “televisie”*

<b>Content Unit 18</b>	<b>Content Unit 19</b>	<b>Content Unit 20</b>
u	advies / informatie / uitleg	geven
u	<b>mij</b>	uitleggen
u	<b>mij</b>	helpen

Een ander opvallend punt is dat bij enkele scenario’s verwijswaarden, zoals “die” en “dat”, goed gerekend mogen worden wanneer het subject of object al eerder is genoemd, of wanneer naar het argument op de afbeelding wordt gewezen. Bij andere scenario’s krijgt men alleen een punt wanneer de naam van het subject of object wordt benoemd en niet wanneer een verwijswaard wordt gebruikt. In Figuur 5 is te zien dat bij scenario 11 “bril” het verwijswaard “die/dit” goed gerekend mag worden wanneer eerder al het Content Unit “bril” is genoemd. Bij scenario 10 “boekje bezorgen” (zie Figuur 6) moet het object “bloemen/boekje” echter twee keer genoemd worden. Een verwijswaard als “die/deze” zou in dit geval niet als correct Content Unit gescoord mogen worden. Dit komt bij verschillende scenario’s voor (o.a. scenario 6 “slager” en 14 “wekker”). Per scenario zou bekeken moeten worden, wanneer een verwijswaard goed gerekend mag worden en wanneer het noodzakelijk is om het argument nogmaals te benoemen.

Dassek (2016) had ook enkele aanbevelingen ten behoeve van het ANTAT-CU4-schema. Zo beschrijft ze dat de Content Units voor scenario 11 “bril”, 16 “gips” en 20 “gemeente” aangepast moeten worden, omdat het CU-schema uitgaat van de scenarioversies zoals beschreven in Giessen (2015) en Aan de Stegge (2015). Naar mijn mening hoeven de Content Units voor scenario 16 en 20 niet aangepast te worden. Deze scenario’s zijn wellicht een beetje anders geformuleerd dan voorheen, maar de strekking blijft gelijk evenals de mogelijke reacties die op de scenario’s gegeven kunnen worden. Deze CU-schema’s kunnen daarom blijven zoals ze zijn.

Het CU-schema van scenario 11 “bril” behoeft echter wel een aanpassing van de Content Units. Dit schema is namelijk gebaseerd op de afbeelding met de bril met een verbogen pootje. Hierop

kunnen sprekers anders reageren dan op een afbeelding met een gebroken bril. Op basis van de responsen van de 31 proefpersonen uit het huidige onderzoek is een aanbeveling voor een nieuw CU-schema gedaan (zie Bijlage S).

### 5.8 Onderzoeksgroep

Het is van belang om de ANTAT-CU4 te toetsen bij een normgroep die vergelijkbaar is met de populatie sprekers met afasie (Filipinski, 2014; Dassek, 2016). Een CVA is de meest voorkomende oorzaak van afasie. Hoewel ook jonge personen door een CVA getroffen kunnen worden, komt dit toch voornamelijk bij ouderen voor (Dharmaperwira-Prins & Maas, 1998). Om deze reden dient bij het samenstellen van een onderzoeksgroep rekening gehouden te worden met leeftijd. Uit onderzoek van Kemper, Greiner, Marquis, Prenovost en Mitzner (2001), waarin gesproken en geschreven taal werd geanalyseerd, bleek dat het taalgebruik, zowel op grammaticaal als semantisch gebied, verandert wanneer men ouder wordt. Ook uit onderzoek naar de vertelvaardigheden van volwassenen (Wright et al., 2005) bleek dat er een significant verschil was tussen de geproduceerde informatie door jongeren en ouderen, waarbij jongeren onder andere meer belangrijke gebeurtenissen van een verhaal benoemden. Filipinski (2014) beschrijft dan ook dat een onderzoeksgroep met sprekers, die gemiddeld jonger is dan de gemiddelde spreker met afasie, mogelijk niet representatief is voor de populatie sprekers met afasie. Wanneer een gemiddelde oudere onderzoeksgroep wordt meegenomen in het onderzoek naar de verbale communicatieve effectiviteit en efficiëntie, kan worden voorkomen dat getoetst wordt op leeftijd in plaats van dat onderzocht wordt of een persoon al dan geen afasie heeft. Dassek (2016) oppert om de scores op de ANTAT-CU te corrigeren op leeftijd. Afhankelijk van de leeftijdscategorie waarin de cliënt valt, kunnen dan punten worden opgeteld of afgetrokken. Naast de factor leeftijd, dient bij het samenstellen van de onderzoeksgroep mogelijk rekening gehouden te worden met handvoorkeur. Bij rechtshandigen is de kans dat in de aangedane zijde eveneens de taal wordt gerepresenteerd namelijk groter dan bij linkshandigen. Dit kan invloed hebben op de vorm en ernst van afasie. Zie Tabel XXI voor een schematische uitleg.

*Tabel XXI Schematische weergave van taalrepresentaties en zijde waarin het letsel zich bevindt (Bastiaanse, 2011; Dharmaperwira-Prins & Maas, 1998)*

	<b>Rechtshandigen</b>	<b>Linkshandigen</b>
Taalrepresentatie in linkerhemisfeer	95-98%	70%
Taalrepresentatie in rechterhemisfeer	2-5%	30%
Letsel in linkhemisfeer	95%	50%
Letsel in rechterhemisfeer	5%	40%
Letsel bilateraal	-	10%

In het huidige onderzoek werd de ANTAT-CU4 getoetst op interne consistentie, test-hertestbetrouwbaarheid en inter- en intrabeoordelaarsbetrouwbaarheid. Kanttekening hierbij is dat deze test werd afgenomen bij gezonde proefpersonen en niet bij sprekers met afasie. Brookshire en Nicholas (1994b) beschrijven dat gezonde sprekers tijdens testafnamen minder stabiele scores laten zien dan sprekers met afasie, zoals bijvoorbeeld bij de test-hertestprocedure in het huidige onderzoek naar voren kwam. Wanneer de ANTAT-CU4 wordt getest bij personen

met afasie, kan het mogelijk zijn dat de scores stabiel zijn en dat de betrouwbaarheidswaarden hoger uitkomen dan bij gezonde proefpersonen is gemeten.

### **5.9 Conclusie**

In dit onderzoek zijn verschillende betrouwbaarheidsmaten van de ANTAT-CU4 getoetst. Hieruit bleek dat de interne consistentie van de ANTAT-CU4 bij gezonde sprekers te laag is en dus onvoldoende betrouwbaar is. De test-hertestbetrouwbaarheid bij gezonde sprekers lijkt wel voldoende te zijn. De inter- en intrabeoordelaarsbetrouwbaarheid is momenteel nog onvoldoende, hoewel dit mogelijk voorkomen kan worden door voor een scoringshandleiding te zorgen. Wanneer de ANTAT-CU4 en het ANTAT-CU4-schema aangepast en verbeterd worden, zou het testinstrument in de toekomst gebruikt kunnen worden om de verbale functionele communicatievaardigheden van cliënten met afasie te toetsen.

## Referenties

- Aan de Stegge, S. (2015). *Aanpassing van ANTAT-scenarios, een onderzoek naar de mogelijkheid tot aanpassing van scenario's uit de ANTAT ter verbetering van actualiteit, inleefbaarheid en verbale eenduidigheid*. Afstudeerproject Taalwetenschap, Radboud Universiteit Nijmegen.
- Afasie Vereniging Nederland. (2017). *Afasie, feiten en getallen*. Verkregen via <https://www.afasie.nl/new/?cat=afasie&nr=13>
- Albin Davis, G. (2014). *Aphasia and related cognitive communicative disorders*. Pearson: Boston.
- Appels, B., & Ponds, R. (2012). De neurologische praktijk. In: *Klinische neuropsychologie*, (R. Kessels, P. Eling, R. Ponds, J. Spikman, & M. van Zandvoort, red.), 51-76, Amsterdam: Boom.
- Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, 14(9), 875-892.
- Baarda, D.B., & De Goede, M.P.M. (2001). *Basisboek methoden en technieken: Handleiding voor het opzetten en uitvoeren van onderzoek*. Groningen: Wolters-Noordhoff.
- Bastiaanse, R. (2011). *Afasie*. Houten: Bohn Stafleu van Loghum.
- Berndt, R.S. (2007). Language production in Aphasia. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 565-578). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780198568971.013.0034
- Berns, P.E.G., Jünger, N., Boxum, E., Nouwens, F., Van der Staaij, M. G., Van Wessel, S., Van Dun, W., Van Lonkhuijzen, J.G., CBO & TNO. (2015). *Logopedische richtlijn 'Diagnostiek en behandeling van afasie bij volwassenen'*. Woerden: Nederlandse Vereniging voor Logopedie en Foniatrie.
- Blomert, L. (1990). What functional assessment can contribute to setting goals for aphasia therapy. *Aphasiology* 4(4), 307-320.
- Blomert, L. (1993). *Aphasie Partner Vragenlijst (APV)*. Eindrapport project Stichting Afasie Nederland.
- Blomert, L., Kean, M.L., Koster, Ch., & Schokker, J. (1994). Amsterdam-Nijmegen Everyday Language Test: construction, reliability and validity. *Aphasiology*, 8(4), 381-407.
- Blomert, M., Koster, Ch. & Kean, M.-L. (1995). *Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden: Handleiding*. Lisse: Swets & Zeitlinger.
- Blomert, L., Koster, C., Van Mier, H., & Kean, M.-L. (1987). Verbal communication abilities of aphasic patients: the everyday language test. *Aphasiology*, 1(6), 463-474.
- Boeije, H. (2014). *Analyseren in kwalitatief onderzoek: Denken en doen*. Den Haag: Boom Lemma uitgevers.
- Boxum, E., Van der Scheer, F., & Zwaga, M. (2013). *ASTA: Analyse voor Spontane Taal bij Afasie*. Vereniging voor Klinische Linguïstiek.
- Brookshire, R.H., & Nicholas, L.E. (1994a). Speech sample size and test-retest stability of connected speech: Measures for adults with aphasia. *Journal of Speech and Hearing Research*, 37, 399-407.
- Brookshire, R.H., & Nicholas, L.E. (1994b). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology*, 22, 119-133.
- Carmines, E.G., & Zeller, R.A. (1979). *Reliability and validity assessment*. London: Sage.
- Centraal Bureau voor de Statistiek. (2004). *Standaard Onderwijsindeling*. Heerlen: uitgave CBS.

- Corpus gesproken Nederlands, (2000). *Protocol voor Orthografische Transcriptie*. Verkregen op 02.02.2017 via [http://lands.let.ru.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/orthography/ort\\_prot.pdf](http://lands.let.ru.nl/cgn/doc_Dutch/topics/version_1.0/annot/orthography/ort_prot.pdf).
- Dassek, S. (2016). *Eerste stappen richting validiteitsonderzoek en betrouwbaarheidsonderzoek van de ANTAT-CU4: Onderzoek naar de test-herstest sensibiliteit en betrouwbaarheid van de ANTAT-CU4*. Afstudeerproject Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- De Heus, P., Van der Leeden, R., & Gazedam, B. (1995). *Toegepaste data-analyse. Technieken voor niet-experimenteel onderzoek in de sociale wetenschappen*. Utrecht: Lemma.
- De Renzi, E., & Vignolo, L.A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, 85, 665-678. <https://doi-org.ru.idm.oclc.org/10.1093/brain/85.4.665>.
- Dharmaperwira-Prins, R., & Maas, W. (1998). *Afasie: beschrijving, onderzoek, behandeling*. Amsterdam: Harcourt Assessment BV.
- Dijkstra, N. (2006). *Begrijpelijkheid of propositiemodel? Het beoordelen van verbale zelfredzaamheid met de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid*. Afstudeerproject Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- Dollaghan, C.A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brookes Publishing Co.
- Eithward. (2012). *Objectief beoordelen van mensen? Dat kan helemaal niet!* Verkregen via <http://www.eithward.nl/page12.html>
- Field, A. (2013). *Discovering Statistics Using SPSS (4th ed.)*. London: Sage Publications.
- Filipinski, F. (2014). *Het scoringsmodel ANTATcu voor de kwantitatieve meting van de functionele communicatievaardigheid bij afasie Laveren tussen een model met standaardreacties en het behoud van de variatie in spontaan taalgebruik*. Afstudeerproject Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- Frattali, C. M. (1992). Functional assessment of communication: Merging public policy with clinical views. *Aphasiology*, 6, 63–83.
- Giessen, M. (2015). *Aanpassing van scenario's van de Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden, een verbetering van de scenario's met als doel een duidelijkere rolverdeling en verbeterde actualiteit en inleefbaarheid*. Afstudeerproject Taalwetenschap, Radboud Universiteit Nijmegen.
- Goodglass, H., & Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination*. Philadelphia: Lea and Febiger.
- Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders, 2e ed.* Philadelphia: Lea & Febiger.
- Graetz, P., de Bleser, R., & Willmes, K. (1992). *Akense Afasietest, Nederlandse versie*. Lisse: Swets & Zeitlinger.
- Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M., Willmess, K., & Huber, W. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language & Communication disorders*, 43, 408-426.
- Hagoort, P. (1992). Vertraagde lexicale integratie bij afatisch taalverstaan. *Stem-, Spraak- en Taalpathologie*, 1(1), 5-22.

- Hallowell, B., & Chapey, R. (2008). Introduction in language intervention strategies in adult aphasia. In: *Language Intervention Strategies in Aphasia and Related Neurogenetic Communication Disorders*, (R. Chapey, Ed.), 3-19, Lippincott: Williams & Wilkins.
- Hanks, P. (2015). *Propositional content*. Oxford: Oxford University Press.
- Hesketh, A., & Bishop, D.V.M. (1996). Agrammatism and adaptation theory. *Aphasiology*, 10(1), 49-80.
- Holland, A.L. (1980). *Communication Abilities in Daily Living*. Baltimore: University Park Press.
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: introductory statistics using generalized linear models*. London: Sage.
- IBM (2016). *IBM SPSS Statistics Version 24*. IBM Corporation.
- Kaplan, E.H., Goodglass, S., & Weintraub, S. (1978). *The Boston Naming Test*. Philadelphia: Lea & Febiger.
- Kemmerer, D. (2015). *Cognitive neuroscience of language*. New York: Psychology Press.
- Kemper, S., Greiner, L.H., Marquis, J.G., Prenovost, K., & Mitzner, T.L. (2002). Language decline across the life span: Results from the nun study. *Psychology and Aging*, 16(2), 227-239.
- Kerkmeer, M., & Heethuis, A.K. (2014). *CAT-NL Psychometrische eigenschappen White Paper 2*. Amsterdam: Pearson Assessment and Information B.V.
- Kirchman, L.-L. (2003). *Anatomie en fysiologie van de mens (15e herziene druk)*. Maarsse: Elsevier gezondheidszorg.
- Kolk, H., & Heeschen, H. (1990). Adaptation symptoms and impairment symptoms in Broca's aphasia. *Aphasiology*, 4(3), 221-231.
- Koo, T.K., & Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15, 155-163.
- Linebarger, M., McGall, D., Virata, T., & Berndt, R.S. (2007). Widening the temporal window: Processing support in the treatment of aphasic language production. *Brain and Language*, 100, 53-68.
- Luria, A.R. (1970). *Traumatic aphasia: Its syndromes, psychology and treatment*. The Hague: Mouton.
- Maso, I., & Smaling, A. (1998). *Kwalitatief onderzoek: praktijk en theorie*. Amsterdam: Boom.
- McNett, M., Omato, S., & Olson, D.M. (2017). Sensitivity, specificity, and receiver operating characteristics: A primer for neuroscience nurses. *American Association of Neuroscience Nurses*, 49(2), 99-101.
- Nicholas, L.E., & Brookshire, R.H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, 338-350.
- Nicholas, L.E., & Brookshire, R.H. (1995). Presence, completeness and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, 38, 145- 156.
- Niewold, C. (2006). *Spontaan herstel van afasie in en na de acute fase* (academisch proefschrift Universiteit van Amsterdam). Utrecht: LOT.
- Oelschlaeger, M. L., & Thorne, J. C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech, Language, and Hearing Research*, 42(3), 636-648.



- Papathanasiou, I., Coppens, P., & Potagas, C. (2013). *Aphasia and related neurogenic communication disorders*. Burlington: Jones & Bartlett Learning.
- Patterson, J.P., & Chapey, R. (2008). Assessment of Language Disorders in Adults. In: *Language Intervention Strategies in Aphasia and Related Neurogenetic Communication Disorders*, (R. Chapey, Ed.), 64-160, Lippincott: Williams & Wilkins.
- Potkamp-van den Belt, J. (2011). *Test-hertestbetrouwbaarheid van spontane-taalonderzoek bij gezonde volwassenen*. Masteronderzoek Taalwetenschap, Rijksuniversiteit Groningen.
- Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075-1091.
- Rousson, V., Gasser, Th., & Seifert, B. (2002). Assessing intrarater, interrater and test–retest reliability of continuous measurements. *Statistics in Medicine*, 21, 3431-3446.
- Ruiter, M., Desain, P., Van Heesch, J., Van Hengel, J., Kraaij, W., Lindhout, M., & Verschaeve, M. (2015). Afasie Gespreksboek, van papier naar app. *Stem-, Spraak- en Taalpathologie*, 20, 151-174.
- Ruiter, M.B., Kolk, H.H.J., Rietveld, T., Dijkstra, N., & Lotgering, E. (2011). Towards a quantitative measure of verbal effectiveness and efficiency in the Amsterdam-Nijmegen Everyday Language Test (ANELT). *Aphasiology* 25(8), 961-975.
- Ruiter, M.B., Rietveld, T.C.M., Hoskam, V., & Van Beers, M.M.A. (2016). An exploratory investigation of e-REST: Teletherapy for chronically aphasic speakers. *International Journal of Telerehabilitation*, 8(1), 21-28.
- Sacks, H. (1972). An initial investigation of the usability of conversational data for doing sociology. In: *Studies in Social Interaction*, (D. Sudnow, Ed.), 31-74, New York: Free Press.
- Shewan, C.M. (1988). The Shewan Spontaneous Language Analysis (SSLA) system for aphasic adults: Description, reliability, and validity. *Journal of Communication Disorders*, 21, 103-138.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin*, 86, 420-428.
- SHV Hersenletsel. (2017). *Afasie*. Verkregen via <https://www.hersenletsel.nl/alles-over-hersenletsel/wat-is-hersenletsel/?/Afasie/>
- Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog*, 44, 561-567.
- Simmons-Mackie, N.N., & Damico, J.S. (1997). Reformulation the definition of compensatory strategies in aphasia. *Aphasiology*, 11(8), 761-781.
- Smolík, F., Stepankova, H., Vyhnálek, M., Nikolai, T., Horáková, K., & Matějka, S. (2016). Propositional density in spoken and written language of Czech-speaking patients with mild cognitive impairment. *Journal of Speech, Language, and Hearing Research*, 59, 1461-1470.
- Stark, J., & Viola, M.S. (2006). *ANELT scale for understandability under scrutiny: What is „a little”, “medium”, “reasonable”, “good” in terms communicative effectiveness?* Clinical Aphasiology Conference 2006, 36. Verkregen via <http://eprints-prod-05.library.pitt.edu/1771/1/215.pdf>, geraadpleegd op 9 maart 2017.
- Swinburn, K., Porter, G., & Howard, D. (2014). *Comprehensive Aphasia Test Nederlandstalige bewerking (CAT-NL)*. Amsterdam: Pearson Assessment and Information B.V. (Nederlandstalige bewerking door Visch-Brink, E., Vandenborre, D., Smet, de. H.J., & Mariën, P.)

- Tilburg University. (2017). *Betrouwbaarheid en validiteit*. Verkregen via <https://www.tilburguniversity.edu/nl/studenten/studie/colleges/spsshelpdesk/eddesk/betrouwba.htm>, geraadpleegd op 20 april 2017.
- Tompkins, C.A., Gibbs-Scott, A., & Sharp, V.L. (2008). Research principles for the clinician. In: *Language Intervention Strategies in Aphasia and Related Neurogenetic Communication Disorders*, (R. Chapey, Ed.), 163-185, Lippincott: Williams & Wilkins.
- Turner, A., & Greene, E. (1977). The construction and use of a propositional text base (No. Tech Report 87-02). Boulder, CO: University of Colorado, Institute of Cognitive Neuroscience.
- Van den Brink, W.P. (1998). Modellen voor testcores. In: *Testleer en testconstructie*, (W.P. van den Brink, & G.J. Mellenbergh (Red.)), Amsterdam: Boom.
- Van der Meulen, I., Van Gelder-Houthuizen, J., Wieggers, J., Wielaert, S., & Van de Sandt-Koenderman, M. (2008). *Scenario Test: Verbale en non-verbale communicatie bij afasie*. Houten: Bohn Stafleu van Loghum.
- Van der Meulen, I., Van de Sandt-Koenderman, W.M.E., Duivenvoorden, H.J., & Ribbers, G.M. (2010). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language and Communication Disorders*, 45(4), 424-435.
- Van Wijk, C. (2000). *Toetsende statistiek: Basistechnieken*. Bussum: Coutinho.
- Van Zwieten, M., & Willems, D. (2004). Waardering van Kwalitatief onderzoek. *Huisarts en Wetenschap*, 47, 38-43. doi: 10.1007/BF03083653.
- Verhage, F. (1964). *Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zeventenzeventig jaar* [Intelligence and age: Study with Dutch people from age 12 to 77]. Assen: Van Gorcum.
- Wolthuis, N, Mendez Orellana, C., Nouwens, F., Jonkers, R., Visch-Brink, E., & Bastiaanse, R. (2014). Stabiliteit spontane taal bij chronische milde afasie. *Stem-, Spraak- en Taalpathologie*, 19, 103-120.
- Wright, H.H., Capilouto, G., Wagovich, S., Cranfill, T., & Davis, J. (2005). Development and reliability of quantitative measure of adults' narratives. *Aphasiology*, 19(3/4/5), 263-273.
- Yorkston, K.M., & Beukelman, D.R. (1977). *A system for quantifying verbal output of high-level aphasic patients*. In R.H. Brookshire (ed.). *Clinical aphasiology: Conference proceedings, 1977*. Minneapolis: BRK publishers.
- Yorkston, K.M., & Beukelman, D.R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, 45(1), 27-36.

## **Bijlagen**

De bijlagen zijn in deze uitgave niet beschikbaar ter inzage.