

Psychometrische eigenschappen van de ANTAT-CU Real Time

Op weg naar de klinische implementatie van een nieuwe scoringswijze

Nijmegen, 24 maart 2022
Radboud Universiteit Nijmegen
Faculteit der Letteren, Masteropleiding Taal- en Spraakpathologie

Auteur: Julie E.M.C. Theunissen, s4843894
Begeleider: Dr. Marina B. Ruiter
Tweede lezer: Caroline E. van Slooten – Wietsma, MA
Datum eindgesprek: 13-04-2022

© Maart, 2022, Radboud Universiteit, Nijmegen.

Niets uit deze uitgave mag verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt worden, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier zonder voorafgaande schriftelijke toestemming van de auteur en de Radboud Universiteit Nijmegen.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronics, mechanical, photocopying, recording or otherwise, without the written permission of the author and Radboud University.

Voorwoord

Deze scriptie is geschreven in het kader van mijn afstuderen voor de materspecialisatie Taal- en Spraakpathologie aan de Radboud Universiteit te Nijmegen. Voor personen met afasie (PMA) is het uiteindelijke doel van afasietherapie om de functionele communicatie te verbeteren. De traditionele Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid (ANTAT; Blomert, Koster, & Kean, 1995) is een test waarmee de verbaal-mondelinge communicatie bij PMA kwalitatief in kaart wordt gebracht. De laatste jaren heeft de ANTAT – met name ten behoeve van het scoringssysteem – verschillende ontwikkelingen doorgemaakt. Zo hebben Ruiters, Kolk, Rietveld, Dijkstra en Lotgering (2011) een kwantitatief scoringssysteem – op basis van Content Units (CU) – ontwikkeld, genaamd de ANTAT-CU. Echter, gezien de vereiste orthografische transcripties was het niet realistisch om dit scoringssysteem in de klinische praktijk in te zetten. Deze masterscriptie is onder andere een vervolg op de masterscriptie van Otters (2019), waarin de ANTAT-CU Real Time (RT) scoringswijze is geïntroduceerd. Ten behoeve van de klinische toepasbaarheid zijn in deze scriptie enkele psychometrische eigenschappen van de ANTAT-CU RT nader onderzocht en is een verbetervoorstel gegeven voor de scoringshandleiding.

Begin februari 2020 ben ik afgestudeerd als logopedist aan de Hogeschool van Arnhem en Nijmegen. Gedurende deze studie en mijn eerste werkervaringen ben ik erachter gekomen dat mijn interesses liggen in het behandelen van patiënten met niet aangeboren hersenletsel (NAH). Parallel aan mijn scriptie was ik werkzaam als logopedist op de afdeling neurorevalidatie, in de Sint Maartenskliniek. Deze scriptie maakte het mogelijk om een unieke bijdrage te leveren aan de ontwikkeling van de ANTAT, een waardevol meetinstrument dat ook in deze klinische setting regelmatig wordt ingezet om een beeld te krijgen van de alledaagse communicatie bij PMA. Het is mooi om bij te kunnen dragen aan de verdere ontwikkeling van het ANTAT-CU scoringssysteem en de handleiding, die ik later wellicht zelf in de klinische praktijk zal gaan inzetten.

Gedurende het schrijven van deze scriptie heb ik veel geleerd en heb ik me, zowel op persoonlijk als op wetenschappelijk vlak, verder kunnen ontwikkelen. Dit proces was zeker niet altijd even makkelijk, maar ik kan met trots zeggen dat ik blij ben met het eindresultaat. Graag wil ik een aantal mensen bedanken die het mede mogelijk hebben gemaakt om tot dit mooie eindresultaat te komen. Allereerst dank ik dr. Marina Ruiters voor de fijne begeleiding en waardevolle feedback tijdens dit traject. Tevens wil ik dr. Marina Ruiters en dr. V. Piai bedanken voor het beschikbaar stellen van de data voor mijn scriptie. Tot slot wil ik mijn lieve ouders, broer en vriend bedanken voor de onvoorwaardelijke steun en liefde die ze mij hebben gegeven tijdens dit proces. De wijze woorden die ik telkens heb mogen ontvangen en het vertrouwen dat ze altijd in me hebben gehad, hebben ervoor gezorgd dat ik deze scriptie met een fijn gevoel heb kunnen afronden.

Ik wens u veel leesplezier toe.

Julie Theunissen
Nijmegen, maart 2022

Samenvatting

Inleiding: De Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid (ANTAT; Blomert, Koster, & Kean, 1995) beoordeelt de verbaal-mondelinge communicatie bij personen met afasie (PMA). Ruiters et al. (2011) hebben het kwalitatieve scoringsstelsel vervangen door een kwantitatief scoringsstelsel, gebaseerd op Content Units (ANTAT-CU). Echter, door de vereiste orthografische transcriptie (OT) is deze scoringswijze niet haalbaar voor de klinische praktijk. Otters (2019) heeft de ANTAT-CU Real Time (RT) scoringswijze – waarbij de CUs direct tijdens testafname gescoord kunnen worden – geïntroduceerd en concludeert dat deze scoringswijze een volwaardig alternatief is voor de ANTAT-CU OT scoringswijze. Otters doet daarbij de aanbeveling om de door haar ontwikkelde scoringshandleiding te verbeteren. Met het oog op de klinische toepasbaarheid is het van belang om enkele psychometrische eigenschappen van de ANTAT-CU RT nader te onderzoeken en een verbetervoorstel voor de scoringshandleiding te formuleren.

Methode: In totaal namen 79 Nederlandstalige sprekers deel: 54 personen zonder afasie (PZA) en 25 PMA. Voor iedere participant werden de uitingen van de ANTAT-testafname zowel middels de ANTAT-CU OT (op basis van transcripten) als middels de ANTAT-CU RT (op basis van audio-opnamen) gescoord. De resultaten werden vergeleken op diverse psychometrische eigenschappen: known-group validiteit, convergente validiteit, interne consistentie, test-hertestbetrouwbaarheid en interbeoordelaarsbetrouwbaarheid. Om te komen tot een verbetervoorstel voor de handleiding zijn twijfelgevallen, die zich tijdens het scoren hebben voorgedaan, geïntroduceerd.

Resultaten: De psychometrische eigenschappen bleken tussen beide scoringswijzen vergelijkbaar te zijn. Ze voldeden beide aan het pathologiecriterium (known-group validiteit). Er was sprake van een uitzonderlijk sterke, positieve correlatie tussen de verbaal-mondelinge effectiviteit gescoord middels beide scoringswijzen (convergente validiteit). De interne consistentie voor parallelversie 1 bij PMA was voor beide scoringswijzen goed, hoewel deze voor PZA een slechte betrouwbaarheid liet zien. Voor de totaalscore geldt dat de interbeoordelaarsbetrouwbaarheid voor beide scoringswijzen goed tot excellent was, al werden op de afzonderlijke scenario's aanzienlijk brede betrouwbaarheidsintervallen gevonden. Voor beide scoringswijzen werd een goede test-hertestbetrouwbaarheid bij PZA gevonden. Ten behoeve van de verbetering van de handleiding zijn de geïntroduceerde twijfelgevallen afhankelijk van de aard ingedeeld in categorieën.

Conclusie: Concluderend kan gesteld worden dat, hoewel toekomstig onderzoek vereist is om de klinische toepasbaarheid te verbeteren, de resultaten uit de huidige studie suggereren dat de ANTAT-CU RT een betrouwbare en valide methode is om de verbaal-mondelinge effectiviteit bij PMA kwantitatief en op een snelle, minder arbeidsintensieve wijze in kaart te brengen. Hiermee biedt de ANTAT-CU RT een geschikt alternatief voor de ANTAT-CU OT. Dit betekent – samen met de suggesties voor de handleiding en het scoringsstelsel – een belangrijke stap voor de klinische praktijk.

Abstract

Introduction: The Amsterdam-Nijmegen Test for Everyday Language (ANELT; Blomert, Koster, & Kean, 1995) assesses verbal communication of persons with aphasia (PWA). Ruiter et al. (2011) replaced the original qualitative scoring system with a quantitative scoring system based on Content Units (ANELT-CU). However, due to the required orthographic transcription (OT), the ANELT-CU OT is not feasible for clinical practice. Otters (2019) introduced the ANELT-CU Real Time (RT) scoring method - where the CUs can be scored directly during test taking - and concludes that this scoring method is a suitable alternative to the ANELT-CU OT scoring method. In doing so, Otters recommends an improvement to the scoring manual she has developed. With a view to clinical applicability, it is important to further investigate some psychometric properties of the ANELT-CU RT and to formulate an improvement to the scoring manual.

Methods: A total of 79 Dutch speakers participated: 54 neurologically healthy speakers (NHS) and 25 PWA. For each participant, the utterances of the ANELT test were scored by the ANELT-CU OT (based on transcripts) and the ANELT-CU RT (based on audio recordings). Resulting data were systematically compared on the following psychometric properties: known-group validity, convergent validity, internal consistency, test-retest reliability and inter-rater reliability. Doubtful cases, which occurred during scoring, were inventoried in order to create an improvement to the manual.

Results: The psychometric properties were found to be similar between the two scoring methods. They both met the pathology criterion (known-group validity). There was an exceptionally strong, positive correlation between verbal effectiveness scored using both scoring methods (convergent validity). Internal consistency for parallel version I for PWA was good for both scoring methods, although it showed poor reliability for NHS. Inter-rater reliability was good to excellent for both scoring methods on the overall score, although wide confidence intervals were found on the individual scenarios. Good test-retest reliability was found for both scoring methods for NHS. For the purpose of improving the scoring manual, the inventoried doubtful cases were divided into categories depending on their nature.

Conclusion: Although future research is required to improve its clinical applicability, the results from the present study suggest that the ANELT-CU RT is a reliable and valid method for quantitatively mapping verbal effectiveness in PWA in a rapid, less labor-intensive manner. The ANELT-CU RT offers a suitable alternative to the ANELT-CU OT, which is – together with the suggestions for the scoring manual and the scoring scheme – an important step for clinical practice.

Inhoudsopgave

Voorwoord	3
Samenvatting	4
Abstract	5
1. Inleiding	8
1.1 Aanleiding	8
1.2 Hoofd- en deelvragen	14
1.3 Hypothesen	15
1.3.1 Hoofdvraag 1	15
1.3.1.1 Deelvraag 1 (known-group validiteit)	15
1.3.1.2 Deelvraag 2 (convergente validiteit)	15
1.3.1.3 Deelvraag 3 (interne consistentie)	16
1.3.1.4 Deelvraag 4 (test-hertestbetrouwbaarheid)	16
1.3.1.5 Deelvraag 5 (interbeoordelaarsbetrouwbaarheid)	17
1.3.2 Hoofdvraag 2	17
2. Methode	18
2.1 Participanten en materiaal	18
2.1.1 ANTAT-afname	18
2.1.2 Participanten dataset 1	18
2.1.3 Audio-opnames dataset 1	18
2.1.4 Participanten dataset 2	19
2.1.5 Audio-opnames dataset 2	19
2.1.6 Participanten dataset 3	19
2.1.7 Audio-opnames dataset 3	20
2.1.8 Orthografische transcripties	20
2.2 Procedure	20
2.2.1 Opbouw scoringsschema	21
2.2.2 Scholing	22
2.2.3 Werkwijze scoring	23
2.2.4 Dataverwerking	23
2.3 Statistische-analyse hoofdvraag 1	24
2.3.1 Known-group validiteit (deelvraag 1)	24
2.3.2 Convergente validiteit (deelvraag 2)	24
2.3.3 Interne consistentie (deelvraag 3)	25
2.3.4 Test-hertestbetrouwbaarheid (deelvraag 4)	25
2.3.5 Interbeoordelaarsbetrouwbaarheid (deelvraag 5)	26
2.4 Kwalitatieve analyse hoofdvraag 2	26

3. Resultaten	28
4. Discussie.....	29
5. Conclusie.....	29
Referenties.....	30
Bijlagen	35

1. Inleiding

1.1 Aanleiding

Ongeveer een derde van de mensen die een beroerte heeft doorgemaakt, ontwikkelt afasie (Brady, Kelly, Godwin, Enderby, & Campbell, 2016). Afasie is een verzamelnaam voor taalstoornissen die ten gevolge van een hersenbeschadiging – bij personen met een normale taalontwikkeling – zijn ontstaan. In Nederland wordt de prevalentie van afasie geschat op ongeveer 30.000 en de incidentie op 10.000 (Berns et al., 2015). In de meeste gevallen (ongeveer 85%) wordt afasie veroorzaakt door een cerebrovasculair accident (CVA) (Bastiaanse, 2010). Een tweede belangrijke oorzaak is een (schedel) trauma, dat bijvoorbeeld ontstaat ten gevolge van een verkeersongeluk. Afasie manifesteert zich vrijwel altijd in zowel het productieve als het receptieve taalgebruik. Onder het productieve taalgebruik vallen de taalmodaliteiten spreken en schrijven en onder het receptieve taalgebruik worden de taalmodaliteiten auditief taalbegrip en lezen geschaard. Ondanks dat bij afasie in de meeste gevallen zowel het productieve als het receptieve taalgebruik is aangedaan, is de afasiepopulatie heterogeen. Dit houdt in dat de beschadiging van de taalmodaliteiten – afhankelijk van de ernst en locatie van het hersenletsel en de mate van aanpassing – per patiënt kunnen variëren (Prins & Bastiaanse, 1997).

Problemen in een of meerdere taalmodaliteiten leiden tot beperkingen in de communicatie. Uit onderzoek blijkt dat veel personen met afasie (PMA) naast hun linguïstische stoornis ook cognitieve stoornissen vertonen (Paemeleire, 2010). Aangezien communicatie in het dagelijks leven centraal staat, is het niet verrassend dat afasie (en mogelijke cognitieve stoornissen) een negatieve invloed heeft op verschillende aspecten van het dagelijks functioneren. Logopedisten zijn verantwoordelijk voor de beoordeling, diagnose en revalidatie van afasie. Vroegtijdige herkenning van afasie is van cruciaal belang, zodat adequate verwijzing plaatsvindt en therapie zo snel mogelijk van start kan gaan. Het uiteindelijke doel binnen de afasietherapie is erop gericht om de functionele communicatie van PMA te verbeteren (Thompson & Worrall, 2008; Wallace et al., 2016). Met andere woorden, het is belangrijk dat PMA zo optimaal mogelijk leren communiceren en dit – bij voorkeur zelfstandig – kunnen toepassen in hun dagelijkse omgeving.

Om de beperkingen en mogelijkheden van PMA zowel voorafgaand als tijdens de therapie in kaart te brengen, wordt binnen de zorg steeds vaker gebruik gemaakt van het International Classification of Functioning, Disability and Health model (ICF; RIVM, 2002; WHO, 2001). Binnen dit model worden drie perspectieven onderscheiden: het lichaam (functies en anatomische eigenschappen, ook wel functieniveau genoemd), het menselijk handelen (activiteitsniveau) en het perspectief van de mens als deelnemer aan het maatschappelijk leven (participatieniveau) (NVLF, 2017). Naast deze drie perspectieven beschrijft het ICF-model twee factoren die (zowel positief als negatief) van invloed kunnen zijn op het functioneren. Het gaat hierbij om externe factoren, zoals familie en/of werk, en persoonlijke factoren, zoals ziekte-inzicht en communicatiebehoeften. Om deze termen te verduidelijken is in Bijlage 1 een voorbeeld van een ingevuld ICF-model bij een persoon met afasie opgenomen.

De in Nederland bekende en beschikbare meetinstrumenten kunnen zowel in de acute, revalidatie- als chronische fase¹ worden ingezet. Ten behoeve van een gerichte afasiediagnostiek en -therapie is het van belang dat de logopedist inzicht krijgt in de beperkingen en mogelijkheden op alle drie de perspectieven van het ICF-model. Daarbij is het van belang zich te realiseren dat er een discrepantie kan bestaan tussen de prestaties van PMA op taaltests (functieniveau) en op functionele communicatietaken (activiteiten-/participatieniveau) (Irwin, Wertz, & Avent, 2002; Marini, Caltagirone, Pasqualetti, &

¹ Acute fase = eerste 2 weken na het ontstaan, revalidatiefase = tot 6 maanden na het ontstaan, chronische fase = vanaf 6 maanden na het ontstaan (NVAT, 2015).

Carlomagno, 2007; Worrall, McCooey, Davidson, & Larkins, 2002). Zo kan het voorkomen dat een patiënt beter presteert op functieniveau, maar dat de functionele communicatievaardigheid achterblijft. De patiënt presteert in dit geval bijvoorbeeld goed op een benoemtaak, maar loopt vast in de functionele communicatie doordat de belasting op overige cognitieve functies, zoals (sociaal-)pragmatische en narratieve vaardigheden, groter is. Aan de andere kant kan het ook voorkomen dat een patiënt juist beter presteert op functionele communicatietaken en slechter bij taaltaken op functieniveau, waarbij minder ruimte is voor aanpassingsgedragingen (Hüsers & Jonkers, 2017).

In Nederland zijn met name meetinstrumenten in gebruik ter bepaling van stoornissen op functieniveau (Berns et al., 2015). Bij dit soort meetinstrumenten worden taalproductie- en taalbegripstaken item per item aan de patiënt voorgelegd. Hier kan bijvoorbeeld gedacht worden aan taken waarbij de patiënt in één woord moet aangeven wat er op de afbeelding staat of taken waarbij de patiënt een zin krijgt aangeboden en de bijpassende afbeelding moet aanwijzen. Deze taken vinden doorgaans plaats in een één-op-één setting waarbij alle mogelijke afleiding wordt weggehaald en de patiënt vaak ruim de tijd krijgt om te reageren.

In de afasiologie is men het er over eens dat meetinstrumenten op functieniveau door onder andere de gecreëerde setting niet volstaan om functionele communicatie (die zich bevindt op activiteiten-/participatieniveau) te beschrijven en/of te voorspellen (Beeke, Maxim, Best, & Cooper, 2011; Holland, 1980). Om alledaagse communicatie zo valide en realistisch mogelijk in kaart te brengen, zouden PMA bij voorkeur gedurende langere tijd geobserveerd moeten worden in alledaagse situaties. Echter, in de praktijk is een dergelijke observatie veelal subjectief en te tijdrovend (Prins & Bastiaanse, 2001).

Ondanks deze inzichten is het verrassend om vast te stellen dat er nog altijd een fundamenteel gebrek is aan overeenstemming over wat functionele communicatie in de praktijk überhaupt inhoudt en hoe deze gedefinieerd moet worden, terwijl dit binnen de afasietherapie wel centraal staat (Doedens, 2020). De term functionele communicatie wordt in de afasiologie op zeer uiteenlopende manieren gebruikt (Elman & Bernstein-Ellis, 1995). Men is het er hooguit over eens dat met functionele communicatie bedoeld wordt op iemands vermogen tot communiceren in het dagelijks leven, maar er bestaat een gebrek aan consensus over welke vaardigheden daadwerkelijk nodig zijn voor functionele communicatie. Het ontbreken aan consensus hierover wordt duidelijk wanneer gekeken wordt naar de verschillende meetinstrumenten die dit beogen te meten (Brady et al., 2016; Barnes & Bloch, 2018).

Doedens (2020) heeft naar aanleiding van deze kwestie in een kritisch review onderzoek gedaan naar een aantal instrumenten die momenteel gebruikt worden om functionele communicatie bij PMA te kwantificeren. De beschrijvingen van de meetinstrumenten en de operationalisering vertonen overlap, aangezien ze allemaal als doel hebben de communicatie in het dagelijks leven vast te stellen. Echter, in deze meetinstrumenten worden verschillende methodieken gebruikt (Irwin et al., 2002; Linnik, Bastiaanse, & Hohle, 2016). Zo zijn er gestandaardiseerde instrumenten die functionele communicatie kwantificeren als de mate van succes in het overbrengen van boodschappen in nagebootste alledaagse situaties, uitgelokt door foto's of in rollenspellen met de testleider (Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid (ANTAT), Blomert, Koster, & Kean, 1995; Communicative Abilities in Daily Living 2, Holland, Frattali, & Fromm, 1999; Scenario Test, Van der Meulen, Van de Sandt-Koenderman, Van Gelder-Houthuizen, Wiegers, & Wielart, 2008). Daarnaast zijn er instrumenten, zoals de Conversatie Analyse (Beeke, Maxim, & Wilkinson, 2007), die zich richten op direct geobserveerde face-to-face communicatie, waarbij de focus ligt op patronen die de conversatie structureren. Bij weer andere meetinstrumenten geldt juist de impact van de communicatiestoornis op het leven van de patiënt als onderdeel van een maat van communicatie. Hierbij worden waarnemingen door de patiënt zelf of een familielid beoordeeld (Communication Outcome after Stroke, client and carer version, Long, Hesketh, & Bowen,

2009; Long, Hesketh, Paszek, Booth, & Bowen, 2008). De variatie van deze instrumenten weerspiegelt de moeilijkheid om alledaagse communicatie, die complex, veelzijdig en zeer dynamisch is, op een uniforme manier te definiëren en in kaart te brengen.

Doedens (2020) is in haar review op zoek gegaan naar een theoretisch onderbouwde maat voor functionele communicatie en komt uit op drie kernprincipes, ontleend aan Clark (1996). Deze kernprincipes stellen dat functionele communicatie interactief (1), multimodaal (2) en afhankelijk is van gemeenschappelijkheid (3). Met interactief wordt bedoeld dat communicatie een gezamenlijke activiteit is (Clark, 1996; Schegloff, 1982). Taalgebruik komt tot stand door twee of meer mensen, waarbij elke beslissing die tijdens een gesprek wordt genomen, afhankelijk is van het gedrag van de ander. De term multimodaal houdt in dat mensen tijdens het communiceren in het dagelijks leven gebruik maken van verschillende modaliteiten, zoals gesproken taal, gezichtsuitdrukkingen en gebaren. Alle kanalen worden in één samengestelde boodschap geïntegreerd om betekenis uit te drukken, maar ook om te controleren en te begrijpen wat de gesprekspartner communiceert (Clark & Krych, 2004; Kendon, 2004; McNeill, 1992). Gemeenschappelijke afhankelijkheid verwijst volgens Clark naar een set van gedeelde kennis, overtuigingen en veronderstellingen die bestaat tussen twee sprekers. Dit betekent dat hoe meer gemeenschappelijke kennis twee sprekers delen, hoe makkelijker de communicatie verloopt. Doedens stelt dat deze drie perspectieven kunnen fungeren als uitgangspunt voor de discussie over functionele communicatie in de afasiologie.

Ook heeft Doedens (2020) in haar review verschillende instrumenten – die tot doel hebben om functionele communicatie te meten – getoetst aan de drie genoemde perspectieven. Er lijkt op dit moment geen test beschikbaar te zijn die alle aspecten van functionele communicatie in zijn geheel en realistisch meeneemt en tevens praktisch toepasbaar is in de klinische setting (Doedens, 2020). Doedens komt tot de conclusie dat de Scenario Test nog het meest in de buurt komt om functionele communicatie bij PMA op een objectieve, gestandaardiseerde manier in een klinische setting vast te leggen. Deze test meet de communicatieve vaardigheden van PMA, waarbij wordt onderzocht in hoeverre iemand een bepaalde vorm van (alternatieve) communicatie beheerst, en in welke mate dit zelfstandig, of met hulp van een communicatiepartner wordt ingezet. Het interactieve aspect tussen de patiënt en de testleider wordt ook meegenomen (Van der Meulen et al., 2008).

De ANTAT – die in het huidige onderzoek centraal staat – beoogt de vaardigheid te meten om een boodschap effectief en zelfstandig over te brengen door middel van gesproken taal, in gesitueerde scenario's (Blomert et al., 1995). Hierover zegt Doedens (2020) terecht dat – aangezien er geen 'echte' interactie of feedback plaatsvindt tussen de testleider en de patiënt – de ANTAT niet geschikt is om een uitspraak te doen over het interactieve aspect van functionele communicatie. De ANTAT ontlokt op basis van scenario's een monoloog van de patiënt (Blomert & Koster, 2008). Met de inzet van deze scenario's doet de ANTAT bovendien een beroep op extra cognitieve aspecten, zoals het 'doen alsof je ergens bent' (Ramsberger, 1994; Wirz, Skinner, & Dean, 1990). Deze extra cognitieve aspecten stroken niet (altijd) met situaties in het dagelijks leven en kunnen tevens – zoals eerder vermeld – aangetast zijn bij PMA. De taal- en communicatievaardigheden worden hier mogelijk door vertekend. Daarnaast wordt ook het multimodale aspect van functionele communicatie niet meegenomen. De ANTAT richt zich namelijk enkel op de verbaal communicatieve vaardigheid. Door taal geïsoleerd te bestuderen wordt een schat aan informatie gemist die wel degelijk relevant is voor communicatie (Vigliocco, Perniss, & Vinson, 2014). Tot slot wordt binnen de ANTAT slechts voor een gedeelte rekening gehouden met de gemeenschappelijke afhankelijkheid. Verschillende situaties – die in het dagelijks leven kunnen voorkomen – worden geëvalueerd, maar de ANTAT beoordeelt niet het vermogen van de patiënt om met verschillende gesprekspartners te communiceren. Bovendien wordt weinig communicatieve context opgebouwd, aangezien per scenario slechts één vraag aan de patiënt wordt gesteld. Het

voornaamste doel van Doedens was om te evalueren of een instrument de drie hierboven genoemde kernprincipes van functionele communicatie omvat. Aan deze kernprincipes voldoet de ANTAT – zoals hierboven beschreven – slechts voor een deel. De vraag is dan ook of de term functionele communicatie als uitkomstmaat bij de ANTAT wel terecht is.

Ondanks de genoemde tekortkomingen van de ANTAT (in relatie tot de functionele communicatie) wordt in de klinische praktijk in Nederland echter toch vaak de ANTAT ingezet. Waar de Scenario Test met name gericht is op patiënten met een matige tot ernstige afasie, is de ANTAT juist meer geschikt voor patiënten bij wie het taalbegrip relatief behouden is gebleven en de taalproductie slechts mild tot matig gestoord is (Blomert et al., 1995; Van der Meulen et al., 2008). De ANTAT brengt in verschillende alledaagse scenario's de vaardigheid om een verbale boodschap effectief en zelfstandig over te brengen in kaart en is daarmee alsnog van groot belang in de afasietherapie en -diagnostiek op het niveau van activiteiten en participatie. Een belangrijke vraag hierbij is in hoeverre de ANTAT in staat is om de verbaal-mondelinge effectiviteit op een valide en betrouwbare manier te meten.

De validiteit en betrouwbaarheid worden gezien als belangrijke psychometrische eigenschappen van een meetinstrument. De validiteit staat voor de geldigheid van een test en beoordeelt of een test meet wat hij daadwerkelijk zou moeten meten (Drenth & Sijtsma, 2001). Binnen de validiteit van een test kunnen een aantal vormen worden onderscheiden. Momenteel heerst de discussie over de mate waarin meetinstrumenten in staat (moeten) zijn om een voorspelling te doen over het functioneren in het dagelijks leven, ook wel de ecologische validiteit genoemd (Bouwens, 2010; Kessels, Eling, Ponds, Spikman, Van Zandvoort, 2018). Men moet zich voortdurend realiseren dat tests en een testsituatie niet een op een vergelijkbaar zijn met iemands vaardigheden in het dagelijks leven. Het is belangrijk om dit te betrekken bij de interpretatie van resultaten (Kessels et al., 2018). Hoewel het noodzakelijk is dat een meetinstrument valide is, is een hoge mate van validiteit alleen niet voldoende (Field, 2009). Om te kunnen spreken van een valide meetinstrument, moet het instrument volgens Field (2009) op de eerste plaats betrouwbaar zijn. De betrouwbaarheid van een test verwijst naar de mate waarin de test scores vrij zijn van toevallige meetfouten (Drenth & Sijtsma, 2001). Om misverstanden te vermijden, moet vermeld worden dat dé betrouwbaarheid van een test niet bestaat (Evers, Lucassen, Meijer, & Sijtsma, 2010). Verschillende aspecten zijn hierbij van belang, zoals de consistentie van de items in het instrument (interne consistentie), de overeenstemming in scores tussen beoordelaars (interbeoordelaarsbetrouwbaarheid), en de stabiliteit van test scores voor dezelfde persoon in de tijd (test-hertest betrouwbaarheid) (Doedens, 2020).

Het oorspronkelijke scoringssysteem van de ANTAT (Blomert et al., 1995) meet kwalitatief, aan de hand van een 5-puntsschaal, hoe begrijpelijk en verstaanbaar PMA hun boodschap in (gesproken) woorden overbrengen. De testleider scoort direct, tijdens de testafname, de uitingen van de PMA. Dit scoringssysteem is echter erg subjectief aangezien hierbij geen extern criterium² gehanteerd wordt (Ruiter, Kolk, Rietveld, Dijkstra & Lotgering, 2011). Daarnaast kan aan de hand van het oorspronkelijke scoringssysteem geen uitspraak worden gedaan over de efficiëntie van de informatieoverdracht. Dit houdt in dat niet wordt meegenomen hoe snel iemand de relevante informatie communiceert. Wanneer men uitspraak wil doen over de verbale communicatievaardigheid en de voortgang hiervan is de beoordeling van de mate van efficiëntie net zo belangrijk als de mate van effectiviteit. Compensatietherapie kan beide deelaspecten namelijk verschillend beïnvloeden (Ruiter et al., 2011). Wanneer de afasietherapie – bijvoorbeeld bij iemand met woordvindingsproblemen – gericht is op het geven van doeltreffende omschrijvingen, kan dit namelijk zorgen voor een toename van de effectiviteit en een afname van de efficiëntie. De boodschap wordt door het geven van een

² Met het extern criterium wordt in deze context bedoeld op het soort en de hoeveelheid informatie die essentieel wordt geacht om het communicatieve doel in ieder scenario te bereiken (Ruiter et al., 2011).

doeltreffende omschrijving namelijk begrijpelijker, maar het overbrengen van de boodschap kost in dat geval meer tijd. Omgekeerd, wanneer de afasietherapie gericht is op de productie van ellipsen (grammaticaal gereduceerde uitingen) kan dit juist zorgen voor een toename van de efficiëntie en een gelijke mate van effectiviteit.

Naar aanleiding van bovengenoemde tekortkomingen heeft de ANTAT de laatste jaren verschillende ontwikkelingen doorgemaakt om te komen tot een meer valide en betrouwbaar meetinstrument. Zo hebben Ruiter et al. (2011) een kwantitatief scoringsstelsel voor de ANTAT ontwikkeld, genaamd de ANTAT-CU. Hierbij is gebruik gemaakt van essentiële informatie-eenheden, ook wel content units genoemd (CUs). Een CU wordt gezien als gegroepeerde informatie die als één eenheid wordt geproduceerd (Yorkston & Beukelman, 1980). In het ANTAT-CU scoringsstelsel is gekwantificeerd welke informatie personen zonder afasie (PZA) noemen om het communicatieve doel in een bepaald ANTAT-scenario te behalen. Door de communicatieve productie van PZA als uitgangspunt te nemen, is een extern criterium toegevoegd. Met het toevoegen van dit criterium wordt de subjectiviteit van het scoringsstelsel gereduceerd. Ruiter et al. hebben voor elk scenario een lijst met relevante CUs samengesteld, waarbij een 30% criterium³ is gehanteerd als extern referentiekader. Dit criterium houdt in dat elke CU, die door minimaal 30% van de gezonde geïnccludeerde participanten geproduceerd werd, beschouwd werd als relevant voor het communicatieve doel dat in een specifiek scenario moet worden bereikt. Hoewel Ruiter et al. in hun onderzoek aangeven dat de bevindingen gerepliceerd moeten worden in een grotere groep, suggereren de belangrijkste resultaten dat de nieuwe kwantitatieve maat van de verbale effectiviteit gevoeliger is om de verandering van de verbaal-mondelinge effectiviteit bij PMA over de tijd heen te meten. Daarnaast kan aan de hand van het nieuwe scoringsstelsel ook de verbaal-mondelinge efficiëntie gemeten worden, die zoals eerder beschreven naast de effectiviteit ook een deelaspect van de verbale communicatievaardigheid betreft. Kortom, het nieuwe kwantitatieve ANTAT-CU scoringsstelsel geeft een vollediger en gedifferentieerder beeld van de verbaal-mondelinge communicatievaardigheid (Ruiter et al., 2011). Sinds de publicatie van Ruiter et al. in 2011 heeft de ANTAT-CU zich nog verder ontwikkeld. De scenario's zijn allemaal getest op de actualiteit en inleefbaarheid en op basis van deze bevindingen deels aangepast en vervangen. Dassek (2016) en Zwartjens (2017) hebben vervolgens voor alle scenario's CU-scoringschema's vormgegeven (zie Bijlage 2).

Ondanks deze positieve ontwikkelingen van het ANTAT-CU scoringsstelsel, is dit scoringsstelsel weliswaar haalbaar in de wetenschappelijke setting, maar niet toepasbaar in de klinische setting. Om de verbaal-mondelinge effectiviteit en efficiëntie aan de hand van de CUs te kunnen scoren, moeten alle uitingen van PMA namelijk eerst door de testleider orthografisch getranscribeerd worden. Dit is een arbeidsintensief proces. Om het scoren minder arbeidsintensief te maken en daarmee de klinische toepasbaarheid te verbeteren, heeft Otters (2019) een nieuwe scoringswijze geïntroduceerd, genaamd de ANTAT-CU Real Time. Hierbij worden de CUs gelijktijdig met het produceren van de communicatieve uitingen tijdens testafname in het scoringschema gescoord. Daarnaast heeft Otters de CU-scoringschema's van Dassek (2016) en Zwartjens (2017) aangepast en voorzien van een handleiding voor de testleider (zie Bijlage 3 respectievelijk Bijlage 4). Per aanleiding en verzoek is het communicatieve doel beschreven, gevolgd door een aantal kolommen met conceptuele termen (nader toegelicht in paragraaf 2.2.1). Deze overzichtelijkere structuur maakt het voor logopedisten gemakkelijker om de CUs direct tijdens de testafname te scoren. Een orthografische transcriptie is in dat geval niet (meer) noodzakelijk. Otters heeft voor haar studie de uitingen van 31 PZA en 8 PMA middels beide scoringswijzen – zowel orthografisch (OT) als Real Time (RT) – gescoord. De resultaten van beide scoringswijzen zijn naderhand met

³ Gezien de aanzienlijke variabiliteit in de hoeveelheid overgebrachte informatie die gezonde sprekers in de studie van Blomert (1990) vertoonden, is hier niet voor een 100% criterium gekozen (Ruiter et al., 2011).

elkaar vergeleken. De interne consistentie, test-hertestbetrouwbaarheid en de interbeoordelaarsbetrouwbaarheid zijn gerapporteerd en bleken voor de ANTAT-CU RT goed te zijn en op de meeste vlakken zelfs iets beter dan de ANTAT-CU OT (zie toelichting bij de afzonderlijke deelvragen in paragraaf 1.3.1). Het feit dat de geproduceerde uitingen voor de ANTAT-CU RT niet eerst getranscribeerd hoeven te worden, maakt de ANTAT-CU RT minder arbeidsintensief en klinisch toepasbaar (Otters, 2019). Op basis van haar onderzoeksresultaten concludeert Otters dat de ANTAT-CU RT een volwaardig alternatief vormt voor de ANTAT-CU OT. De voor- en nadelen van de verschillende ANTAT-scoringsystemen en -scoringwijzen zijn in Tabel 1 samengevat.

Tabel 1. De voor- en nadelen van de ANTAT-scoringsystemen en -scoringwijzen.

Scoringssysteem	Scoringwijze	Voordelen	Nadelen
Kwalitatief scoringssysteem: ANTAT 5-puntsschaal (Blomert et al., 1995)	Direct (tijdens testafname)	<ul style="list-style-type: none"> - Geen orthografische transcriptie vereist - Minimaal arbeidsintensief 	<ul style="list-style-type: none"> - Subjectief: geen extern criterium waarmee vergeleken wordt - Slechts één deelaspect van de verbale communicatie wordt gescoord, namelijk de effectiviteit; de efficiëntie ontbreekt
Kwantitatief scoringssysteem: ANTAT Content Unit analyse	Achteraf (na testafname), middels orthografisch transcript (OT) (Ruiter et al., 2011)	<ul style="list-style-type: none"> - Beide deelaspecten van verbale communicatie worden gescoord: effectiviteit én efficiëntie - Toename objectiviteit (t.o.v. het kwalitatieve scoringssysteem) - Veranderingen in de communicatievaardigheid worden sensitiever over de tijd heen gemeten (t.o.v. het kwalitatieve scoringssysteem) 	<ul style="list-style-type: none"> - Orthografische transcriptie is noodzakelijk - Arbeidsintensief en tijdrovend voor de klinische setting
	Direct (tijdens testafname), Real Time (RT) (Otters, 2019)	<ul style="list-style-type: none"> - Geen orthografische transcriptie vereist - Minimaal arbeidsintensief en daarmee klinisch toepasbaar - Toename objectiviteit (t.o.v. het kwalitatieve scoringssysteem) - Veranderingen in de communicatievaardigheid worden waarschijnlijk sensitiever over de tijd heen gemeten (t.o.v. het kwalitatieve scoringssysteem) 	<ul style="list-style-type: none"> - Slechts één deelaspect van de verbale communicatie wordt gescoord, namelijk de effectiviteit; de efficiëntie ontbreekt

De conclusies van Otters (2019) dienen echter met enige voorzichtigheid geïnterpreteerd te worden. Wanneer kritisch gekeken wordt naar de onderzoeksopzet, wordt allereerst opgemerkt dat de PMA-onderzoeksgroep in de studie van Otters zeer klein was; deze bestond uit slechts 8 participanten. Bij een te kleine steekproef bestaat het risico op een breder betrouwbaarheidsinterval. Hierdoor is het resultaat minder precies en geen representatieve

afspiegeling van de oorspronkelijke populatie (Vander Stichele, Van Maele, De Paepe, Petrovic, & Van Bortel, 2009). Ten tweede moet vermeld worden dat het in de vergelijking tussen de ANTAT-CU OT en de ANTAT-CU RT niet enkel ging om een andere scoringswijze (namelijk CUs scoren aan de hand van orthografische transcripties vs. direct scoren), maar dat voor beide scoringswijzen ook een ander scoringsschema is gebruikt. Voor de ANTAT-CU OT heeft Otters het door Dassek (2016) en Zwartjens (2017) opgestelde scoringsschema gehanteerd (Bijlage 2). Voor de ANTAT-CU RT heeft Otters het eerder genoemde overzichtelijke scoringsschema ontwikkeld en gehanteerd (Otters, 2019; zie Bijlage 3). Het feit dat voor beide scoringswijzen verschillende scoringsschema's zijn gehanteerd, kan in de vergelijking geleid hebben tot vertekende resultaten. Tot slot geeft Otters zelf ook al aan dat – door de grote sprekersvariatie, de wisselende respons en de wijze waarop deze geïnterpreteerd en gescoord kunnen worden – het altijd lastig zal blijven om alle taaluitingen binnen één scoringsschema te vangen. Otters geeft mede om deze reden de aanbeveling om de instructies in de oorspronkelijke handleiding (Otters, 2019) te expliciteren en concrete voorbeelden op te nemen.

De ANTAT-CU RT scoringswijze en het nieuwe scoringsschema (Otters, 2019) worden op dit moment nog niet ingezet in de klinische praktijk. Ten behoeve van de inzetbaarheid in de klinische praktijk is het van belang om enkele psychometrische eigenschappen nader te onderzoeken. Het huidige onderzoek is daarmee deels een replicatie van het onderzoek dat in 2019 door Otters is uitgevoerd. De beoogde meerwaarde van het huidige onderzoek kan in vier aspecten worden beschreven. Allereerst worden de statistische analyses in het huidige onderzoek uitgevoerd aan de hand van grotere steekproefgroottes. Daarnaast wordt voor beide scoringswijzen hetzelfde scoringsschema gehanteerd. Tevens worden de twee parallelversies van de ANTAT-CU in de huidige studie ook afzonderlijk van elkaar onderzocht. Tot slot zal – met het oog op de klinische toepassing – een voorstel geformuleerd worden voor een verbeterde versie van de handleiding die door Otters is ontwikkeld.⁴

Wanneer aan de hand van het huidige onderzoek kan worden aangetoond dat de betrouwbaarheid en validiteit van de nieuw ontwikkelde ANTAT-CU RT net zo goed óf zelfs beter is dan ANTAT-CU OT én de handleiding is uitgebreid met een aantal concrete scoringsregels en voorbeelden, kan voor het in kaart brengen van de verbaal-mondelinge effectiviteit definitief afstand genomen worden van de orthografische transcripten. Dit betekent een belangrijke stap richting de klinische toepasbaarheid, aangezien de ANTAT-CU RT minder arbeidsintensief is en op een snelle en zo objectief mogelijke wijze inzicht geeft in de verbaal-mondelinge effectiviteit.

1.2 Hoofd- en deelvragen

De eerste hoofdvraag die in dit onderzoek centraal staat is als volgt: “Is de ANTAT-CU RT scoringswijze (ten minste) net zo betrouwbaar en valide voor het in kaart brengen van de verbaal-mondelinge effectiviteit als de ANTAT-CU OT scoringswijze, gemeten middels hetzelfde scoringsschema?”

Om deze hoofdvraag te beantwoorden, zijn vijf deelvragen geformuleerd.

Kwantitatief:

1. Known-group validiteit: Stelt de kwantitatieve score van de verbaal-mondelinge effectiviteit logopedisten in staat om PMA op betrouwbare wijze te onderscheiden van PZA, zowel middels de ANTAT-CU RT als middels de ANTAT-CU OT?
2. Convergente validiteit: Wat is de samenhang tussen de ANTAT-CU RT en de ANTAT-CU OT betreffende de scoring van de verbaal-mondelinge effectiviteit, gemeten bij zowel PMA als PZA?

⁴ In tegenstelling tot het onderzoek van Otters (2019) richt het huidige onderzoek zich enkel op het meten van de verbale effectiviteit; de efficiëntie wordt buiten beschouwing gelaten.

3. Interne consistentie (als aspect van betrouwbaarheid): Wat is de interne consistentie van de ANTAT-CU RT betreffende de verbaal-mondelinge effectiviteit en verschilt deze van de ANTAT-CU OT, gemeten bij zowel PMA als PZA?
4. Test-hertestbetrouwbaarheid: Wat is de test-hertestbetrouwbaarheid van de ANTAT-CU RT betreffende de verbaal-mondelinge effectiviteit en verschilt deze van de ANTAT-CU OT, gemeten bij PZA?
5. Interbeoordelaarsbetrouwbaarheid: Wat is de interbeoordelaarsbetrouwbaarheid van logopedisten betreffende de scoring van de verbaal-mondelinge effectiviteit middels de ANTAT-CU RT en verschilt deze van de ANTAT-CU OT, gemeten bij zowel PMA als PZA?

De tweede hoofdvraag heeft betrekking op de handleiding. De vraag hierbij luidt: “Hoe kan de handleiding verbeterd worden om interpretatieverschillen bij het gebruik van het scoringsschema tot een minimum te reduceren?”.

1.3 Hypothesen

In deze paragraaf worden allereerst de hypothesen van de deelvragen – behorend bij de eerste hoofdvraag – besproken. Daarnaast wordt ingegaan op de hypothese met betrekking tot de tweede hoofdvraag.

1.3.1 Hoofdvraag 1

1.3.1.1 Deelvraag 1 (known-group validiteit)

“Stelt de kwantitatieve score van de verbaal-mondelinge effectiviteit logopedisten in staat om PMA op betrouwbare wijze te onderscheiden van PZA, zowel middels de ANTAT-CU RT als middels de ANTAT-CU OT?”

Known-group validiteit gaat over het vermogen van een meetinstrument om te discrimineren tussen subgroepen met verschillende karakteristieken (Hattie & Cooksey, 1984). Het expliciete doel van de oorspronkelijke ANTAT is het meten van de verbale communicatieve vaardigheden van personen met een taalstoornis (Blomert et al., 1995). Blomert en Koster (2008) stellen dat de ANTAT personen zonder taalstoornis zou moeten onderscheiden van personen met een taalstoornis.

Aangezien het expliciete doel van de oorspronkelijke ANTAT in de ANTAT-CU RT en ANTAT-CU OT scoringswijzen onveranderd is gebleven en het nieuwe scoringsschema is opgesteld aan de hand van het eerder genoemde 30% criterium, is de verwachting dat er aan het pathologiecriterium wordt voldaan en er dus sprake is van een goede known-group validiteit. Met andere woorden, er wordt verwacht dat PZA een (nagenoeg) maximale score op de test behalen en daarmee te onderscheiden zijn van PMA.

1.3.1.2 Deelvraag 2 (convergente validiteit)

“Wat is de samenhang tussen de ANTAT-CU RT en de ANTAT-CU OT betreffende de scoring van de verbaal-mondelinge effectiviteit, gemeten bij zowel PMA als PZA?”

Convergente validiteit betreft de samenhang tussen de resultaten van twee meetinstrumenten, die beiden hetzelfde begrip beogen te meten (Cronbach & Meehl, 1955). De ANTAT-CU OT en de ANTAT-CU RT beogen beide de verbaal-mondelinge effectiviteit in kaart te brengen, echter de scoringswijze is verschillend. Middels deze deelvraag zal worden onderzocht in hoeverre de testcores verkregen middels de ANTAT-CU OT en de ANTAT-CU RT met elkaar samenhangen.

Otters (2019) heeft de samenhang tussen de ANTAT-CU OT en ANTAT-CU RT (weliswaar met een verschillend scoringsschema) eerder onderzocht en vond een significante, positieve, sterke tot uitzonderlijk sterke correlatie tussen beide scoringswijzen ($r > .70$, $p < .001$), gemeten bij zowel PZA (parallelversie I + II) als PMA (parallelversie I). In het huidige

onderzoek wordt de samenhang tussen beide scoringswijzen opnieuw bekeken, maar ditmaal met gebruikmaking van hetzelfde nieuwe scoringsschema. In lijn met de bevindingen van Otters wordt voor de totale testscore een sterke tot uitzonderlijk sterke, positieve correlatie tussen beide scoringswijzen verwacht, gemeten bij zowel PZA als PMA ($r > .70$; Doorn & Rhebergen, 2006). Voor de samenhang tussen de testcores van de twee scoringswijzen op de afzonderlijke scenario's vond Otters correlatiecoëfficiënten die varieerden van .47 tot .99. Aangezien in het huidige onderzoek hetzelfde scoringsschema wordt gehanteerd, wordt minder variatie van deze correlatiecoëfficiënten verwacht en voor elk afzonderlijk scenario een sterke, positieve correlatie tussen de testcores van beide scoringswijzen ($r > .70$; Doorn & Rhebergen, 2006).

1.3.1.3 Deelvraag 3 (interne consistentie)

“Wat is de interne consistentie van de ANTAT-CU RT betreffende de verbaal-mondelinge effectiviteit en verschilt deze van de ANTAT-CU OT, gemeten bij zowel PMA als PZA?”

Door de interne consistentie te onderzoeken, kan een uitspraak worden gedaan over de mate waarin de verschillende scenario's bijdragen aan het concept verbaal-mondelinge effectiviteit, dat beide scoringswijzen beogen te meten. Otters (2019) heeft de interne consistentie van 20 scenario's eerder onderzocht bij 31 gezonde participanten. Beide scoringswijzen lieten een redelijke betrouwbaarheid zien (α tussen .60 en .80; De Heus, Van der Leeden & Gazendam, 2002). In lijn met haar verwachting was de Cronbachs alfawaarde voor de ANTAT-CU RT iets hoger ($\alpha = .75$) dan voor de ANTAT-CU OT ($\alpha = .71$). Als verklaring hiervoor geeft Otters het gebruik van het nieuwe, verbeterde scoringsschema bij de ANTAT-CU RT.

De verwachting voor de huidige studie is dat de Cronbachs alfawaarde – zowel voor de ANTAT-CU RT als voor de ANTAT-CU OT – de waarde gevonden door Otters (2019) zal benaderen en daarmee een redelijke interne consistentie aan zal nemen ($\alpha > .60$; De Heus et al., 2002), zowel gemeten bij PZA als bij PMA. Aangezien voor beide scoringswijzen hetzelfde nieuwe scoringsschema gehanteerd wordt, is tevens de verwachting dat de interne consistentie van beide scoringswijzen niet van elkaar zal verschillen.

1.3.1.4 Deelvraag 4 (test-hertestbetrouwbaarheid)

“Wat is de test-hertestbetrouwbaarheid van de ANTAT-CU RT betreffende de verbaal-mondelinge effectiviteit en verschilt deze van de ANTAT-CU OT, gemeten bij PZA?”

De test-hertestbetrouwbaarheid geeft aan in hoeverre de test tot dezelfde resultaten komt als hij op verschillende momenten bij eenzelfde persoon wordt afgenomen (Van der Meulen et al., 2008). De test-hertestbetrouwbaarheid wordt in de huidige studie gemeten bij PZA, van wie verondersteld wordt dat de communicatieve taalvaardigheden relatief stabiel zijn.

In een eerdere studie vond Otters (2019) een significante, positieve, sterke correlatie voor de verbaal-mondelinge effectiviteit op meetmoment 1 en meetmoment 2, zowel voor de ANTAT-CU OT als voor de ANTAT-CU RT ($r = .76$ respectievelijk $r = .80$, $p < .001$). Ook hier geeft Otters aan dat de hogere waarde voor de ANTAT-CU RT mogelijk te wijten is aan het gebruik van het verbeterde scoringsschema.

Tijdens het huidige onderzoek zal voor beide scoringswijzen opnieuw de test-hertestbetrouwbaarheid gemeten worden. In lijn met de bevindingen van Otters (2019) wordt verwacht dat de testcores zowel bij de ANTAT-CU RT als bij de ANTAT-CU OT stabiel zullen zijn over de tijd heen en daarmee een goede test-hertestbetrouwbaarheid laten zien. Dit houdt in dat de scores op meetmoment 1 niet significant ($p > .05$) verschillen van de scores op meetmoment 2. Aangezien voor beide scoringswijzen hetzelfde nieuwe scoringsschema gehanteerd wordt, is de verwachting dat de test-hertestbetrouwbaarheid van de ANTAT-CU RT en de ANTAT-CU OT niet van elkaar zullen verschillen.

1.3.1.5 Deelvraag 5 (interbeoordelaarsbetrouwbaarheid)

“Wat is de interbeoordelaarsbetrouwbaarheid van logopedisten betreffende de scoring van de verbaal-mondelinge effectiviteit middels de ANTAT-CU RT en verschilt deze van de ANTAT-CU OT, gemeten bij zowel PMA als PZA?”

Aan de hand van de interbeoordelaarsbetrouwbaarheid wordt onderzocht in hoeverre de testscore afhankelijk is van degene die de test scoort. Bij een betrouwbare test bestaat er geen verschil tussen de score van verschillende beoordelaars (Van der Meulen et al., 2008).

In eerder onderzoek heeft Otters (2019) de interbeoordelaarsbetrouwbaarheid van beide scoringswijzen bij 8 PMA onderzocht. De Interclass Correlation Coefficient (ICC; type average measures en absolute agreement) had een waarde van .93 voor de ANTAT-CU OT en een waarde van .97 voor de ANTAT-CU RT, wat in beide gevallen duidt op een excellente interbeoordelaarsbetrouwbaarheid (Koo & Li, 2016). Tijdens het huidige onderzoek zal voor beide scoringswijzen opnieuw de interbeoordelaarsbetrouwbaarheid worden uitgevoerd, waarbij beide (getrainde) beoordelaars de door Otters ontwikkelde handleiding zullen hanteren. In lijn met de bevindingen van Otters wordt verwacht dat zowel de ANTAT-CU RT als de ANTAT-CU OT een excellente interbeoordelaarsbetrouwbaarheid laten zien voor de totaalscore (ICC > .90; Koo & Li, 2016). Aangezien voor beide scoringswijzen hetzelfde scoringsschema gehanteerd wordt, is de verwachting dat de interbeoordelaarsbetrouwbaarheid van de ANTAT-CU RT en de ANTAT-CU OT niet van elkaar zullen verschillen.

Ondanks de goede interbeoordelaarsbetrouwbaarheid op de gehele test gaf Otters (2019) echter aan dat – door de grote sprekersvariatie, de wisselende respons en de wijze waarop deze geïnterpreteerd en gescoord kunnen worden – enige mate van subjectiviteit binnen en tussen testleiders blijft bestaan. In sommige gevallen staat het al dan niet toekennen van punten aan een bepaalde uiting namelijk ter discussie. De verwachting is dat de impact van deze subjectiviteit – voor het al dan niet toekennen van CUs aan een bepaalde uiting – mogelijk meer zichtbaar is op het niveau van de afzonderlijke scenario's, dan op het niveau van de totaalscore. Dit leidt bij de een-op-een vergelijking tussen sommige scenario's mogelijk tot een iets lagere interbeoordelaarsbetrouwbaarheid.

1.3.2 Hoofdvraag 2

“Hoe kan de handleiding aangepast worden om interpretatieverschillen bij het gebruik van het scoringsschema tot een minimum te reduceren?”

Om alle testleiders voorafgaand aan het scoren eenduidig te instrueren heeft Otters (2019) een handleiding opgesteld. Zoals hierboven aangegeven heeft Otters (2019) geconstateerd dat tijdens het scoren enige mate van subjectiviteit binnen en tussen testleiders blijft bestaan. Om de scoring op dit punt te optimaliseren is het uitbreiden van het scoringsschema naar alle mogelijk te produceren CUs helaas geen optie. Dit zou het scoringsschema namelijk erg uitgebreid maken, waardoor het voor testleiders lastiger is om de geproduceerde CUs direct tijdens de testafname te scoren. Otters geeft aan dat de door haar opgestelde handleiding mogelijk niet volledig genoeg is om testleiders eenduidig te laten scoren. In de huidige studie wordt onderzocht op welke wijze de handleiding het best aangepast kan worden om de subjectiviteit en de daaruit volgende interpretatieverschillen tot een minimum te reduceren.

2. Methode

In dit hoofdstuk wordt de methodische verantwoording rondom de onderwerpen participanten, materiaal, scholing, scoringsprocedure, dataverwerking en statistische analyses achtereenvolgens uitvoerig beschreven.

2.1 Participanten en materiaal

In deze paragraaf wordt allereerst de ANTAT-afname kort toegelicht. Om de psychometrische eigenschappen van de twee scoringswijzen te onderzoeken, is gebruikgemaakt van audio-opnames en orthografische transcripten, afkomstig uit drie verschillende datasets. Alle data zijn verzameld na goedkeuring van de participanten. De twee datasets waarbij PMA zijn betrokken, zijn goedgekeurd door de Medisch Ethische Toetsingscommissie. De drie datasets worden na de beschrijving van de ANTAT-afname elk afzonderlijk besproken.

2.1.1 ANTAT-afname

De afname van de ANTAT-CU (Ruiter, Lotgering, & Rietveld, 2016) is aan de hand van een audio-opname vastgelegd. Tijdens de afname zijn verschillende scenario's aan de participanten voorgelegd (zie Bijlage 5). Gedurende de afname is een strikte procedure gevolgd. Bij ieder scenario is door de testleider een testinstructie gegeven, waarna aan de participant is gevraagd hoe hij/zij in deze bepaalde situatie zou reageren. Bij een aantal scenario's is de mondelinge instructie visueel ondersteund middels standaardfoto's. Deze standaardfoto's zijn opgenomen in Bijlage 6. Voor een volledige beschrijving van de instructie en procedure van de ANTAT-CU afname wordt verwezen naar het afnameprotocol in Bijlage 7 (Ruiter et al., 2016).

2.1.2 Participanten dataset 1

Het eerste deel van de participanten was afkomstig uit de scripties van Dassek (2016) en Zwartjens (2017), waarbij de onderzoekers hebben gekeken naar de betrouwbaarheid van de ANTAT-CU bij gezonde sprekers. Bijlage 8 geeft een overzicht van de in- en exclusiecriteria voor deelname aan deze studies. De totale groep bestond uit 31 gezonde Nederlandstalige participanten⁵, van wie 16 man waren. De leeftijd van de participanten varieerde tussen de 25 en 86 jaar ($M = 49.35$, $SD = 17.10$). Het opleidingsniveau varieerde tussen lagere school en wetenschappelijk onderwijs. In Bijlage 9 is een uitgebreid overzicht van de participantkarakteristieken van deze groep opgenomen.

2.1.3 Audio-opnames dataset 1

In de studies van Dassek (2016) en Zwartjens (2017) zijn alle 20 scenario's (zie Bijlage 10) bij iedere participant op 2 meetmomenten afgenomen. De tussenperiode van deze twee meetmomenten bedroeg 8 weken, waarin de participanten geen taalinterventie kregen. Voorafgaand aan de afname zijn voor de 20 scenario's twee verschillende versies gemaakt (zie Bijlage 11). Om volgorde-effecten uit te sluiten verschilde de keuze om versie 1 óf versie 2 op meetmoment 1 aan te bieden per participant. Bovendien, indien een participant op meetmoment 1 versie 1 kreeg aangeboden, werd op meetmoment 2 versie 2 aangeboden en vice versa. Bij één participant (participant 34) is om onbekende reden (waarschijnlijk door een testerfout) zowel op meetmoment 1 als op meetmoment 2 versie 1 afgenomen. Bij de scenario's 1, 4, 6, 11, 13, en 14 werd naast de mondelinge instructie een foto gepresenteerd.

Dassek (2016) en Zwartjens (2017) hebben van drie scenario's ('bril', 'arm' en 'gemeente') twee verschillende versies afgenomen. Deze scenario's verschilden van elkaar wat betreft de instructie of gepresenteerde foto. Op deze manier kon vastgesteld worden welke versie van het scenario het duidelijkst was voor de participant. Zwartjens heeft naar aanleiding

⁵ Van de 31 participanten waren 24 participanten afkomstig uit de studie van Dassek (2016) en 7 uit de studie van Zwartjens (2017).

van haar onderzoek de 20 scenario's verdeeld over twee parallelversies, van ieder 10 scenario's. Deze nieuwe indeling is weergegeven in Bijlage 12. In de huidige studie zijn enkel de scenario's meegenomen die uiteindelijk in deze twee parallelversies zijn opgenomen.

2.1.4 Participanten dataset 2

Voor het tweede deel van de participanten is gebruikt gemaakt van een selectie van de data verzameld in de lopende studie 'Taal in het brein, van links naar rechts', onder leiding van dr. V. Piai (Veni 451-17-003). Dit onderzoek heeft tot doel te onderzoeken wat de rol is van het taalsysteem in de rechterhemisfeer bij personen die schade hebben opgelopen aan hun taaldominante linkerhemisfeer. De onderzoeksgroep van Piai bestond uit 47 Nederlandstalige participanten, waaronder 23 PZA. De overige 24 participanten hebben in de (sub)acute fase afasie gehad, geoperationaliseerd als een afwijkende/pathologische score op één of meerdere afasietests (niet verder gespecificeerd). Bijlage 13 geeft een overzicht van de in- en exclusiecriteria op basis waarvan deze 24 participanten zijn geselecteerd. In het huidige onderzoek is één extra inclusie criterium toegevoegd: de aanwezigheid van fatische kenmerken op het moment van de ANTAT-afname, vastgesteld op basis van klinisch oordeel. Dit is enerzijds beoordeeld door het onderzoeksteam van Piai en anderszijds door de auteur van het huidige onderzoek. Het hanteren van het aanvullende inclusie criterium heeft aanvankelijk geleid tot een groep van 18 participanten. De fatische kenmerken zijn per participant in Bijlage 14 onder de kolom 'klinisch oordeel' terug te vinden. Voor participant 48 waren op het moment van de ANTAT-afname weliswaar fatische kenmerken aanwezig, echter omwille van een onvolledige testafname is deze participant alsnog geëxcludeerd voor de huidige studie.

Samengevat zijn uit het corpus van Piai aldus 40 participanten geanalyseerd. De leeftijd van de 23 PZA, van wie 12 man waren, varieerde tussen de 22 en 75 jaar ($M = 59$, $SD = 11.72$). De overige 17 participanten, van wie 11 man waren, zijn in de huidige studie als PMA gelabeld. De leeftijd van deze groep varieerde tussen de 22 en 78 jaar ($M = 56.71$, $SD = 14.08$). Hoewel de exacte tijd post onset (TPO) onbekend was voor participant 43 en 50, bleek hun afasie in een chronisch stadium te zijn (minstens 6 maanden na aanvang). De gemiddelde TPO voor de overige 15 participanten bedroeg 6.54 ($SD = 7.88$). In Bijlage 14 is een uitgebreid overzicht van de participantkarakteristieken van deze groep opgenomen.

2.1.5 Audio-opnames dataset 2

In de audio-opnames uit de studie 'Taal in het brein, van links naar rechts' is op één meetmoment parallelversie 1 van de ANTAT-CU afgenomen. Een overzicht van de 10 scenario's die tot parallelversie 1 behoren, is opgenomen in Bijlage 12. Bij de scenario's 1, 4 en 9 werd naast de mondelinge instructie een foto gepresenteerd. Voor één van de personen met afasie (participant 40) ontbrak de afname van het scenario 'afpraak'.

2.1.6 Participanten dataset 3

Het derde gedeelte van de participanten was afkomstig uit een lopende pilot-studie genaamd SimpTell (Trial NL7855), onder leiding van dr. M.B. Ruiter, dr. A. Roelofs en dr. V. Piai. Deze studie doet onderzoek naar het effect van de therapeutische app SimpTell⁶ op de verbale functionele communicatie. De onderzoeksgroep bestond uit 10 Nederlandstalige PMA. Alle PMA bevonden zich in de chronische fase (ten minste 6 maanden na aanvang) en hadden ten gevolge van afasie een gestoorde zinsproductie. De ernst van de afasie is uitgedrukt aan de hand van de ruwe score op de Verkorte Token Test (VTT; De Renzi & Faglioni, 1978). Bij een enkele participant was de VTT score niet beschikbaar of was de afasie niet te objectiveren middels de VTT. Deze participanten vielen uit op andere afasietests, zoals de Nederlandse Comprehensive

⁶ SimpTell is een afkorting voor Semi-onafhankelijke Interactieve Multimodale Productietraining van ELLipsen en leert mensen met chronische Broca-afasie hun blijvende zinsproductieproblemen te compenseren door in ellipsen te spreken (Ruiter et al., 2019).

Aphasia Test (CAT-NL; Visch-Brink, Vandenborre, De Smet & Mariën, 2014). Bij 4 participanten was naast afasie ook nog sprake van spraakapraxie. Echter, bij deze participanten was ten minste 95% van de spraakoutput verstaanbaar. Bijlage 15 geeft een volledig overzicht van de in- en exclusiecriteria voor deelname aan het SimpTell onderzoek. Het opleidingsniveau is niet gedefinieerd. De groep bestond uit 8 mannen en 2 vrouwen tussen de 44 en 78 jaar ($M = 56.39$, $SD = 9.57$). De gemiddelde TPO was 5.77 ($SD = 4.63$) en de gemiddelde score op de VTT bedroeg 25.94 ($SD = 6.13$), wat kan worden geïnterpreteerd als een lichte afasie. In Bijlage 16 is een uitgebreid overzicht van de participantkarakteristieken van deze groep opgenomen.

2.1.7 Audio-opnames dataset 3

Van de 10 participanten uit de SimpTell studie zijn 8 participanten op twee meetmomenten, met een tussenperiode van 6 weken, getest. Om te controleren voor een mogelijk leereffect is op meetmoment 1 parallelversie I en op meetmoment 2 parallelversie II afgenomen. In de tussenperiode hebben de participanten 30 uur telegramstijltherapie (SimpTell) gekregen, verspreid over 6 weken. Van de andere 2 participanten zijn audio-opnames aangeleverd, waarbij door persoonlijke redenen enkel parallelversie I van de ANTAT op één meetmoment is afgenomen. Naast de mondelinge instructie werd bij de scenario's 1, 4, 9 in zowel parallelversie I als in parallelversie II een foto getoond. In verband met de coronapandemie heeft de ANTAT-afname voor de participanten 9 en 10 online plaatsgevonden. Voor deze participanten zijn video-opnames gedeeld. Tabel 2 geeft een schematische weergave van de onderzoeksopzet van de SimpTell studie.

Tabel 2. Schematische weergave van de onderzoeksopzet van de SimpTell studie.

	T1	Tussenperiode	T2
n=8 (PMA)	Afname 10 scenario's ANTAT (parallelversie I)	6 weken; 30 uur telegramstijltherapie middels SimpTell	Afname 10 scenario's ANTAT (parallelversie II)
n=2 (PMA)	Afname 10 scenario's ANTAT (parallelversie I)	n.v.t.	n.v.t.

2.1.8 Orthografische transcripties

Naast de audio-opnames zijn voor alle drie de datasets orthografische transcripten aangeleverd. Iedere audio-opname is op basis van het protocol Orthografische Transcriptie (Corpus gesproken Nederlands, 2000) in afzonderlijke Word-bestanden getranscribeerd.

Tijdens het scoren van de orthografische transcripties, werd opgemerkt dat voor 8 participanten⁷ uit dataset 2 twee verschillende Word-bestanden met orthografische transcripten waren gedeeld. Voor deze participanten heeft de auteur van het huidige onderzoek het Word-bestand met de zuiverste transcripties als basis genomen en indien nodig (op basis van de audio-opnames) aangepast. Om ervoor te zorgen dat de transcripten zo zuiver mogelijk waren, is de opname tijdens het transcriberen onbeperkt teruggeluisterd. Om te voorkomen dat het meermaals beluisteren van de audio-opnames bias zou veroorzaken tijdens het scoren volgens de ANTAT-CU RT scoringswijze, zijn de audio-opnames 4 weken vóór de orthografische scoring beluisterd.

2.2 Procedure

Om de onderzoeksvragen van de huidige studie te beantwoorden is alle data zowel Real Time (RT), middels audio-opnames, als Orthografisch (OT), middels transcripten, gescoord. In deze paragraaf wordt achtereenvolgens de opbouw van het scoringsschema, de scholing, de werkwijze van de scoring en de dataverwerking toegelicht.

⁷ Het gaat hierbij om de participanten 31, 32, 40, 41, 42, 43, 45 en 46.

2.2.1 Opbouw scoringsschema

Het scoringsschema (zie Bijlage 3) bestaat voor ieder afzonderlijk scenario uit scoringstabellen waarin Content Units (CU) zijn opgenomen. Een CU kan bestaan uit een inhoudswoord (zelfstandig naamwoord, werkwoord, bijwoord, bijvoeglijk naamwoord) of een zinsdeel (b.v. ‘op de tafel’). Voor ieder scenario is achtereenvolgens een aanleiding en een verzoek voor het communicatieve doel beschreven. De aanleiding betreft een introductie of constatering. Het verzoek bestaat doorgaans uit een mededeling, een uitnodiging, het aanzetten tot een handeling of een voorstel om een situatie op te lossen. Bij een aantal scenario’s is het geven van een aanleiding niet van toepassing. In dat geval is in het scoringsschema enkel een tabel voor het verzoek opgenomen.

Iedere rij in het scoringsschema staat voor een propositie. Een propositie duidt op een groepering van CUs die ten grondslag ligt aan een bepaalde uiting (Ruiter et al., 2011). Voor het gebruiksgemak zijn de aanleidingen en verzoeken in de eerste kolom als volledige zin gepresenteerd, zie Tabel 3 als voorbeeld. In de overige kolommen zijn de afzonderlijke concepten weergegeven volgens de ‘wie’, ‘wat’, ‘waar’ en ‘wanneer’ structuur (of wijzigingen van deze vorm). Onder ieder concept zijn in de witte cellen enkele voorbeelden van CUs – die binnen de desbetreffende concepten passen – opgenomen. Het is van belang dat de testleider zich realiseert dat de concepten verschillend verbaal uitgedrukt kunnen worden en de voorbeelden slechts dienen ter verduidelijking van het concept, de lijst is namelijk niet uitputtend. De testleider kruist uiteindelijk de concepten aan die door de persoon in CUs worden uitgedrukt.

Tabel 3. Scoringsschema van de aanleiding voor het scenario ‘boeket’.

Aanleiding					
Ik kom een boeket uitzoeken voor een kennis.	O WIE	O VOOR WIE	O WAT	O DOEN	Score ___/4
	-Ik	-Voor een kennis	-Bloemen	-Uitzoeken -Laten bezorgen	

In de scoringsschema’s zijn voor sommige scenario’s de woorden ‘EN’ en/of ‘OF’ opgenomen. ‘EN’ houdt in dat een persoon beide proposities moet produceren om het maximaal aantal punten te behalen. ‘OF’ houdt in dat de productie van verschillende proposities kan leiden tot het behalen van het maximaal aantal punten, zie Tabel 4 voor een voorbeeld.

Tabel 4. Scoringsschema van de aanleiding voor het scenario ‘betalingsherinnering’ met de opties ‘EN’ en ‘OF’.

Aanleiding					
Ik heb een betalingsherinnering gekregen.	O WIE	O WAT	O DOEN	Score _____/6	
	-Ik	-Betalingsherinnering/ -Rekening	-Gekregen		
EN					
<i>Keuze uit onderstaande opties</i>					
Ik heb het bedrag al betaald.	O WAT	O BEPALING VAN GESTALDHEID	O DOEN		
	-Bedrag -Rekening -Verwijzing	-Al (lang) -Wel	-Betaald -Overgemaakt		
OF					
Ik kan het bankafschriftbewijs laten zien.	O WIE	O WAT	O DOEN		
	-Ik	-Bankafschrift -Betalingsbewijs	-Hebben -Laten zien -Doorsturen		

Het is echter niet toegestaan om verschillende opties bij ‘OF’ te combineren. Enkel de CUs voor de propositie waarbinnen de meeste CUs geproduceerd zijn, mogen gescoord worden. Wanneer iemand bij het scenario ‘vriend’ bijvoorbeeld de uiting “wat leuk, heb je lang niet

gezien” produceert, is de scoring in Tabel 5 dus niet toegestaan. In dit voorbeeld worden in de eerste aanleiding meer CUs geproduceerd dan in de tweede aanleiding. In Tabel 6 is de correcte scoring voor de genoemde voorbeelduiting weergegeven.

Tabel 5. Foutief ingevuld scoringsschema voor het scenario ‘vriend’ bij de uiting “wat leuk, heb je lang niet gezien”.

Aanleiding					
Ik heb jou al lang niet meer gezien.	O WIE	<input checked="" type="radio"/> WIE <i>Aangesproken persoon</i>	<input checked="" type="radio"/> HOE LANG	<input checked="" type="radio"/> DOEN	Score 4/4
	-Ik/we	-Je/elkaar [je]	-Lange tijd [lang]	-Niet gezien [niet gezien]	
OF					
Ik vind het leuk jou weer te zien.	<input checked="" type="radio"/> BEPALING VAN GESTELDHEID	O WIE <i>Aangesproke n persoon</i>	O BIJWOORDELIJKE BEPALING VAN TIJD	O DOEN	
	-Leuk [leuk]	-Je/elkaar	-Weer	-Te zien	

Tabel 6. Correct ingevuld scoringsschema voor het scenario ‘vriend’ bij de uiting “wat leuk, heb je lang niet gezien”.

Aanleiding					
Ik heb jou al lang niet meer gezien.	O WIE	<input checked="" type="radio"/> WIE <i>Aangesproken persoon</i>	<input checked="" type="radio"/> HOE LANG	<input checked="" type="radio"/> DOEN	Score 3/4
	-Ik/we	-Je/elkaar [je]	-Lange tijd [lang]	-Niet gezien [niet gezien]	
OF					
Ik vind het leuk jou weer te zien.	O BEPALING VAN GESTELDHEID	O WIE <i>Aangesproken persoon</i>	O BIJWOORDELIJKE BEPALING VAN TIJD	O DOEN	
	-Leuk	-Je/elkaar	-Weer	-Te zien	

De handleiding die Otters (2019) naar aanleiding van het scoringsschema heeft geschreven en in deze studie voor beide scoringswijzen is gehanteerd, is opgenomen in Bijlage 4. In deze handleiding zijn een aantal algemene regels omtrent het scoren samengevat.

2.2.2 Scholing

De interbeoordelaarsbetrouwbaarheid is voor beide scoringswijzen tussen twee vrouwelijke beoordelaars berekend. Beoordelaar 1 (26 jaar) heeft in 2018 de opleiding logopedie afgerond en in 2019 de master Taal- en Spraakpathologie. Deze beoordelaar heeft het nieuwe ANTAT-CU scoringsschema vormgegeven (Otters, 2019) en heeft zich op deze manier het scoren eigen gemaakt. Beoordelaar 2 (24 jaar) – de auteur van deze scriptie – heeft de opleiding logopedie in 2020 afgerond en volgt op het moment de master Taal- en Spraakpathologie.

Om de betrouwbaarheid van het scoren te waarborgen, heeft beoordelaar 2 voorafgaand aan de scoring een scholing gevolgd. Deze scholing is in het voorjaar van 2021 door dr. M.B. Ruiter voorgedragen en heeft online via Zoom plaatsgevonden. Het eerste gedeelte van de scholing bestond uit een presentatie, waarin zowel de ontwikkeling van het ANTAT-beoordelingssysteem als de ANTAT-CU handleiding centraal stond. Het tweede gedeelte van de scholing bestond uit het praktisch oefenen met beide scoringswijzen. Voor het oefenen werden twee verschillende scenario’s gebruikt (respectievelijk scenario ‘schoen’ en scenario ‘voetballen’), die afkomstig waren van participant 13 uit dataset 2. Er werd gestart met de ANTAT-CU RT scoringswijze, waarbij gelijktijdig met het afspelen van de audio-opname het bijbehorende scoringsschema op de PowerPoint dia werd getoond. In de audio-opname was de instructie van het scenario te horen, gevolgd door de respons van de participant. Nadat de

scoring was uitgevoerd werd ook de ANTAT-CU OT scoringswijze geoefend. Om tijd te besparen waren de audio-opnames voorafgaand aan de scholing reeds getranscribeerd. Op de PowerPoint dia werd de orthografische transcriptie van de respons en het bijbehorende scoringsschema getoond.

Om tot een goede behendigheid van het scoren te komen, is na afloop van de scholing besloten om als oefening de volledige afname van de ANTAT bij één participant middels de ANTAT-CU RT scoringswijze te scoren. Dit betrof participant 1, meetmoment 1, afkomstig uit dataset 1. Nadat alle scenario's van deze participant volledig waren gescoord, zijn de scores afgezet tegen de scores van dr. M.B. Ruiten. In een overleg is vervolgens besproken hoe enkele verschillen zijn ontstaan en wat de correcte manier is om bepaalde taaluitingen te scoren.

2.2.3 Werkwijze scoring

Otters, tevens beoordelaar 1, heeft een gedeelte van de data al gescoord in het kader van haar onderzoek in 2019. In het huidige onderzoek is omwille van het kunnen vaststellen van de interbeoordelaarsbetrouwbaarheid dezelfde werkwijze aangehouden⁸.

Alle aangeleverde audio-opnames en orthografische transcripten zijn in een beveiligde online werkmap van de Radboud Universiteit geplaatst en waren toegankelijk middels een VPN-verbinding. Om te zorgen dat de RT-scoring zo zuiver mogelijk kon worden uitgevoerd, is ervoor gekozen om eerst de RT-scoring uit te voeren, alvorens de OT-scoring werd uitgevoerd. Het lezen van de orthografische transcripten vóór het scoren van de uitingen aan de hand van de audio-opnames zou er namelijk voor kunnen zorgen dat de uitingen van de participanten onbewust bij de onderzoeker zouden beklipen, wat kan leiden tot vertekende resultaten. De orthografische transcripties zijn om deze reden pas nadat de RT-scoring was afgerond in de beveiligde online map geplaatst. Tussen de uitvoering van de twee scoringswijzen is een periode van twee weken aangehouden.

Tijdens de RT-scoring zijn alle audio-opnames via een vrije en opensource-mediaspeler (VLC media player) één voor één beluisterd en direct gescoord. Om de scoring te controleren was het toegestaan om de audio-opname naderhand nog één keer opnieuw te beluisteren en de score indien nodig aan te passen. Tijdens de OT-scoring zijn alle uitingen van de participanten nogmaals gescoord. Het verschil met de RT-scoring is dat alle uitingen nu niet aan de hand van audio-opnames gescoord werden, maar middels de orthografische transcripties. Voor iedere 10 scenario's kostte het gemiddeld 10 minuten om de uitingen middels één scoringswijze te scoren.

Om de tweede hoofdvraag te beantwoorden zijn zowel tijdens de OT- als de RT-scoring mogelijke interpretatieverschillen – die zich bij het scoren van de geproduceerde uitingen hebben voorgedaan – geïnventariseerd (hierna te noemen: twijfelgevallen). Alle twijfelgevallen zijn samengebracht in een tabel. Voor elk twijfelgeval is het scenario, participantnummer en de dataset vermeld. Tevens is van elk twijfelgeval ter illustratie een voorbeeld gegeven en is toegelicht hoe de uiting uiteindelijk is gescoord⁹.

2.2.4 Dataverwerking

De afhankelijke variabele van de huidige studie is de verbaal-mondelinge effectiviteitsscore. Om tot deze score te komen, is dezelfde berekening gehanteerd zoals dit in eerdere studies is gedaan (Lotgering, 2009; Ruiten et al., 2011; Zwartjens, 2017; Otters, 2019). De totaalscore per scenario is berekend door de behaalde CUs van de aanleiding en het verzoek bij elkaar op te tellen. Het maximaal aantal haalbare CUs op de scenario's is variabel en loopt uiteen van 3 CUs bij het scenario 'arm' tot 9 CUs bij het scenario 'betalingsherinnering'. Voor de 20 scenario's samen kon een maximale score van 126 tot 127 punten behaald worden. Deze variabele score is afhankelijk van het scenario 'kat' (scenario 18), waarbij twee verschillende formuleringen

⁸ De scoring door beide beoordelaars heeft onafhankelijk van elkaar plaatsgevonden.

⁹ De twijfelgevallen die door Otters (2019) zijn aangedragen, zijn geïntegreerd in de volledige inventarisatie van het huidige onderzoek.

mogelijk zijn die ieder een eigen maximaal aantal te behalen punten heeft (5 óf 6). Voor de 10 scenario's van parallelversie I kon een maximale score van 66 punten behaald worden en voor parallelversie II 60 tot 61 punten. Dit verschil was – net zoals bij de 20 scenario's – afhankelijk van de formulering bij het scenario 'kat'. In Bijlage 10 en 12 is de maximale score per scenario, voor iedere studie weergegeven.

Gezien het feit dat bij het scenario 'kat' een verschillend aantal punten behaald kon worden, is het totaal aantal behaalde CUs omgerekend naar een percentage. De verbaal-mondelinge effectiviteit, oftewel het percentage correct geproduceerde CUs, is berekend door de totaal behaalde CUs te delen door het maximale aantal te behalen CUs x 100%. Deze berekening is voor beide scoringswijzen uitgevoerd. Naast de totaalscore voor de verbaal-mondelinge effectiviteit werd het percentage ook voor ieder afzonderlijk scenario berekend.

2.3 Statistische-analyse hoofdvraag 1

Na de dataverwerking is alle data ingevoerd en verwerkt in SPSS 27 (Statistical Package for the Social Sciences; IBM, 2020). In deze paragraaf worden de statistische analyses voor de deelvragen van hoofdvraag 1 beschreven. Voor iedere deelvraag was de afhankelijke variabele gelijk, namelijk de verbaal-mondelinge effectiviteitsscore in percentage. De onafhankelijke variabele was afhankelijk van de onderzoeksvraag en wordt per deelvraag nader toegelicht.

2.3.1 Known-group validiteit (deelvraag 1)

In de statistische analyse voor de known-group validiteit was groep (PMA vs. PZA) de onafhankelijke variabele en de verbaal-mondelinge effectiviteitsscore de afhankelijke variabele. De known-group validiteit is berekend middels de onafhankelijke t-test, waarbij is onderzocht of de twee steekproefgemiddelden van de PZA- en PMA-groep significant van elkaar verschilden. De onafhankelijke t-test is voor parallelversie I op meetmoment 1, zowel voor de RT- als voor de OT-scoringswijze uitgevoerd. In Tabel 7 is een overzicht opgenomen van de participanten die voor deze statistische analyse zijn meegenomen.

Voordat de onafhankelijke t-test werd uitgevoerd, is gecontroleerd of aan de vier assumpties ('normaal verdeeld', 'niveau van variabelen', 'onafhankelijkheid' en 'homogeniteit van varianties') werd voldaan. Allereerst is de normaliteitstoets uitgevoerd. Voor beide groepen was de Shapiro-Wilk niet-significant, met een *p*-waarde hoger of gelijk aan .05. Met deze bevinding werd voldaan aan de assumptie van 'normaal verdeeld'. Aangezien de afhankelijke variabele van ratio niveau was en de twee groepen onafhankelijk van elkaar waren, werd ook voldaan aan de assumpties 'niveau van variabelen' en 'onafhankelijkheid'. De laatste assumptie 'homogeniteit van varianties' is getoetst middels de Levene's test, waarbij is nagegaan of de variantie van de afhankelijke variabele voor iedere groep ruwweg hetzelfde was. Zoals verwacht, waren de PZA veel homogener, waardoor de significantie van de Levene's Test onder de gebruikelijke .05 ($p < .001$) lag. Met een significantie-waarde kleiner dan .05, wordt de assumptie 'homogeniteit van varianties' geschonden. Om deze reden is in de statistische output gekeken naar de rij Equal variances not assumed (Field, 2009). Voor de interpretatie van de effectgrootte (Cohens *d*) is de richtlijn van Sawilowsky (2009) aangehouden, zie Bijlage 17.

2.3.2 Convergente validiteit (deelvraag 2)

Middels de Pearsons correlatiecoëfficiënt is voor iedere groep apart de samenhang tussen de verbaal-mondelinge effectiviteitsscore, gemeten middels de ANTAT-CU OT en de ANTAT-CU RT, berekend. In Tabel 7 is een overzicht opgenomen van de participanten die voor deze statistische analyse zijn meegenomen.

Voordat de Pearsons correlatiecoëfficiënt werd uitgevoerd, is gecontroleerd of aan de vier assumpties ('schaal', 'normaal verdeeld', 'lineariteit' en 'homoscedasticiteit') werd voldaan. Beide variabelen (zowel de OT- als de RT-score) waren van ratio niveau, waarmee werd voldaan aan de assumptie 'schaal'. Om te toetsen of de afhankelijke variabele van iedere

groep normaal was verdeeld, werd de normaliteitstest uitgevoerd. Voor de PMA-groep was de Shapiro-Wilk voor de RT-effectiviteitsscore niet-significant ($p = .070$) en voor de OT-effectiviteitsscore wel significant ($p = .043$). De Kolmogorov-Smirnov (weliswaar minder conservatief) bleek voor beide scoringswijzen niet-significant te zijn, met een p -waarde hoger dan .05. Bij een niet-significante p -waarde kan worden uitgegaan van normaal verdeelde data (Field, 2009). Voor steekproefgroottes van minstens 30 participanten kan tevens de Centrale Limiet Stelling¹⁰ (CLS; Field, 2009) aangehouden worden. De PMA-groep bestond uit 25 participanten, dit komt erg in de buurt van de 30. De PZA-groep bestond uit 31 participanten. Volgens de CLS kan dus worden aangenomen dat werd voldaan aan de normaliteit. Ter controle is ook voor deze groep de normaliteitstoets uitgevoerd, waarbij de Shapiro-Wilk inderdaad voor beide scoringswijzen niet-significant was ($p > .05$). Ook de laatste twee assumpties met betrekking tot de lineariteit en homoscedasticiteit konden voor beide groepen worden aangenomen. Aan de scatterplots is te zien dat de punten willekeurig en gelijkmatig over de grafiek zijn verspreid (Bijlage 18). Dit patroon wijst erop dat aan de veronderstellingen van lineariteit en homoscedasticiteit was voldaan (Field, 2009).

De correlatie tussen beide scoringswijzen is zowel voor de totaalscores als voor ieder scenario afzonderlijk berekend. Er is hierbij eenzijdig getoetst met een significantieniveau van .05. De correlatiecoëfficiënt is uitgedrukt in Pearson's r . Voor de interpretatie van de Pearsons correlatiecoëfficiënt is de richtlijn van Doorn en Rhebergen (2006) aangehouden, zie Bijlage 19 voor een overzicht van deze richtlijn.

2.3.3 Interne consistentie (deelvraag 3)

Om een uitspraak te doen over de interne consistentie is gekeken naar de waarde van de Cronbachs alfa. Cronbachs alfa is berekend op het moment dat alle scenario's geïnccludeerd waren én op het moment dat bepaalde scenario's verwijderd zouden worden. Deze betrouwbaarheidsanalyse is voor beide scoringswijzen op meetmoment 1, voor zowel PZA als PMA apart in kaart gebracht. In Tabel 7 is een overzicht opgenomen van de participanten die voor deze statistische analyse zijn meegenomen.

Voor de interpretatie van de waarden van Cronbachs alfa is de richtlijn van De Heus et al. (2002) aangehouden. Binnen deze richtlijn worden als het ware drie 'treden' aangehouden: een goede betrouwbaarheid ($\alpha > .80$), een redelijke betrouwbaarheid ($\alpha > .60$ en $< .80$) en een slechte betrouwbaarheid ($\alpha < .60$). Voor het verwijderen van een scenario is de vuistregel aangehouden om een scenario uit de test te verwijderen indien de alfawaarde met meer dan .05 stijgt (De Heus et al., 2002) én de betrouwbaarheid met minimaal één trede toeneemt.

2.3.4 Test-hertestbetrouwbaarheid (deelvraag 4)

Voor de deelvraag gericht op de test-hertestbetrouwbaarheid was de onafhankelijke variabele het tijdstip – moment van afname (T1 vs. T2) – en de afhankelijke variabele de verbaal-mondelinge effectiviteitsscore. Om te onderzoeken of de verbaal-mondelinge effectiviteit bij beide scoringswijzen stabiel was over de tijd heen óf dat er sprake was van een verandering, is voor de 31 PZA uit dataset 1 een gepaarde t -test uitgevoerd. Hierbij zijn de behaalde effectiviteitsscores op meetmoment 1 voor beide scoringswijzen afgezet tegen de behaalde effectiviteitsscores op meetmoment 2.

Voordat de gepaarde t -test werd uitgevoerd, is gecontroleerd of aan de twee assumpties ('normaal verdeelde verschillcores' en 'niveau van variabelen') werd voldaan. De Shapiro-Wilk bleek voor alle metingen niet-significant te zijn, met een p -waarde hoger dan .05. Dit houdt in dat de verschillcores op de afhankelijke variabele tussen de twee condities (T1 en T2) bij de 20 scenario's samen normaal verdeeld waren. Aangezien de onafhankelijke variabele van

¹⁰ De CLS stelt dat bij een steekproefgrootte groter dan 30 participanten mag worden aangenomen dat de data normaal verdeeld is (Field, 2009).

categorisch niveau en de afhankelijke variabele van continu niveau was, werd ook aan de tweede assumptie ‘niveau van variabelen’ voldaan. Voor de interpretatie van de effectgrootte (Cohens *d*) is de richtlijn van Sawilowsky (2009) aangehouden, zie Bijlage 17.

2.3.5 Interbeoordelaarsbetrouwbaarheid (deelvraag 5)

Voor de deelvraag gericht op de interbeoordelaarsbetrouwbaarheid was de onafhankelijke variabele beoordelaar (B1 vs. B2) en de afhankelijke variabele de verbaal-mondelinge effectiviteitsscore. Aangezien de CU-scores voor de verbaal-mondelinge effectiviteit in de huidige studie zijn omgerekend naar percentages behoorde de data tot continue data. Bij continue data wordt de Intraclass Correlation Coefficient (ICC) gebruikt om aan te tonen dat codeurs onafhankelijk tot vergelijkbare conclusies komen bij hetgeen zij observeren (Ranganathan, Pramesh, & Aggarwal, 2017). In SPSS is de ICC-berekening gebaseerd op de terminologie van McGraw en Wong (1996). McGraw en Wong hebben 10 ICC-vormen gedefinieerd op basis van het "model"¹¹, het "type"¹² en de "definitie"¹³ van de relatie die belangrijk wordt geacht. Zij geven aan dat de uiteindelijke methodekeuze van deze drie aspecten wordt bepaald door de onderzoeksopzet. In Bijlage 20 is een schema weergegeven dat het ICC-keuzeprocess stroomlijnt (Koo & Li, 2016). In het huidige onderzoek is dit schema als leidraad aangehouden voor het selecteren van de juiste ICC-vorm.

De twee beoordelaars in het huidige onderzoek zijn vooraf geselecteerd en geschoold in het gebruik van de twee scoringswijzen. Om deze reden moesten ze behandeld worden als ‘fixed factor’. Dit heeft voor het “model” geleid tot de keuze voor two-way mixed. Alle participanten zijn daarnaast door beide beoordelaars gescoord. Omdat de interesse lag in de betrouwbaarheid van de twee beoordelaars als groep is voor het “type” de optie ‘average measures’ geselecteerd. Tot slot is voor de “definitie” gekozen voor ‘absolute agreement’, aangezien de auteur geïnteresseerd was in de exacte overeenstemming tussen de beoordelaars in het toekennen van punten voor geproduceerde CU-eenheden. Echter, opgemerkt moet worden dat het hierbij niet gaat om de exacte overeenkomst van de gemarkeerde CUs, maar om een exacte overeenkomst van het aantal CUs in percentage. Voor de interpretatie van de ICC is de richtlijn van Koo en Li (2016) aangehouden, waarbij de ICC is uitgedrukt in Cronbachs alfa (zie Bijlage 21). De interbeoordelaarsbetrouwbaarheid is voor de gehele test én voor ieder scenario afzonderlijk op meetmoment 1, voor beide scoringswijzen en voor zowel PMA als PZA berekend. In Tabel 7 is weergegeven welke participanten door beide beoordelaars zijn gescoord en voor deze statistische analyse zijn meegenomen.

2.4 Kwalitatieve analyse hoofdvraag 2

Om vast te stellen op welke wijze de handleiding en/of het scoringschema aangepast dient te worden, zijn de geïnventariseerde twijfelgevallen geanalyseerd. In de analyse is gekeken naar de aard van het twijfelgeval. Afhankelijk hiervan is elk twijfelgeval ingedeeld in een bepaalde categorie en is de frequentie van elke categorie vastgesteld. Tot slot is ingezoomd op de frequentie van specifieke twijfelgevallen (voor een bepaalde uiting in een specifiek scenario).

¹¹ Model: one-way random model, two-way random model, of two-way mixed model.

¹² Type: één beoordelaar/meting of het gemiddelde van een bepaald aantal beoordelaars/metingen.

¹³ Definitie: consistentie of absolute overeenstemming.

Tabel 7. Participanten per statistische analyse weergegeven.

	known-group validiteit (meetmoment 1)	convergente validiteit (meetmoment 1)	interne consistentie (meetmoment 1)	test-hertest-betrouwbaarheid (meetmoment 1 vs. 2)	interbeoordelaarsbetrouwbaarheid (meetmoment 1)	
	Voor de RT- en OT scoringswijze				OT	RT
PZA	Totaal: 54 participanten (parallelversie I) - 31 afkomstig uit dataset 1 - 23 afkomstig uit dataset 2	Totaal: 31 participanten afkomstig uit dataset 1 (parallelversie I en II)	Totaal: 31 participanten afkomstig uit dataset 1 (parallelversie I en II)	Totaal: 31 participanten afkomstig uit dataset 1 (parallelversie I en II)	Totaal: 14 participanten ^b afkomstig uit dataset 1 (pp 1, 2, 5, 8, 13, 14, 16, 18, 20, 22, 25, 26, 31, 36) (parallelversie I en II)	Totaal: 31 participanten afkomstig uit dataset 1 (parallelversie I en II)
PMA	Totaal: 25 participanten (parallelversie I) - 17 afkomstig uit dataset 2 - 8 participanten afkomstig uit dataset 3 ^a	Totaal: 25 participanten (parallelversie I) - 17 afkomstig uit dataset 2 - 8 participanten afkomstig uit dataset 3 ^a	Totaal: 25 participanten (parallelversie I) - 17 afkomstig uit dataset 2 - 8 participanten afkomstig uit dataset 3 ^a	n.v.t.	OT Totaal: 8 participanten afkomstig uit dataset 2 (pp 31, 32, 40, 41, 42, 43, 45, 46) (parallelversie I)	RT Totaal: 17 participanten (parallelversie I) - 8 afkomstig uit dataset 2 (pp 31, 32, 40, 41, 42, 43, 45, 46) - 9 afkomstig uit dataset 3 (pp 1, 2, 3, 4, 5, 6, 7, 8, 9) ^c

^a Van de 10 participanten die beschikbaar zijn gesteld, zijn uiteindelijk 8 participanten voor deze deelvraag geïnccludeerd. De participanten 3 en 7 uit dataset 3 (SimpTell studie) hebben namelijk ook deelgenomen aan de studie ‘Taal in het brein, van links naar rechts’ (dataset 2; participant 54 en 57). Voor deze statistische analyse zijn de PMA uit beide datasets samengevoegd. Aangezien het niet zuiver is om eenzelfde participant twee keer mee te nemen, zijn de participanten 3 en 7 uit dataset 3 voor deze deelvraag geëxcludeerd.

^b Geen 15 participanten, want bij PZA34 T1 (dataset 1) stond het transcript van T2.

^c Aangezien de eerste beoordelaar enkel de participanten 3 en 7 uit dataset 3 (SimpTell studie) heeft gescoord en niet (dezelfde) participanten 54 en 57 uit dataset 2 (‘Taal in het brein, van links naar rechts’), konden deze participanten voor deze deelvraag wel geïnccludeerd worden.

3. Resultaten

Op verzoek van M.B. Ruiter niet zichtbaar.

4. Discussie

Op verzoek van M.B. Ruiters niet zichtbaar.

5. Conclusie

Ten behoeve van de klinische inzetbaarheid van de ANTAT-CU RT scoringswijze zijn in het huidige onderzoek diverse psychometrische eigenschappen van deze scoringswijze in kaart gebracht en vergeleken met de ANTAT-CU OT scoringswijze en is daarnaast een verbetervoorstel geformuleerd voor de handleiding.

In antwoord op hoofdvraag 1 leveren de onderzoeksresultaten van de afzonderlijke deelvragen evidentie voor het feit dat de ANTAT-CU RT scoringswijze net zo betrouwbaar en valide is als de ANTAT-CU OT scoringswijze om de verbaal-mondelinge effectiviteit bij PMA kwantitatief in kaart te brengen. Ter beantwoording van hoofdvraag 2 zijn suggesties gegeven voor het toevoegen van talige instructies met voorbeeldscores in de handleiding en het doorvoeren van kleine wijzigingen in de scoringsschema's. De verwachting is dat de herziene handleiding en de gewijzigde scoringsschema's testleiders meer handvaten zullen geven en dit de eenduidigheid van het scoren zal optimaliseren.

Concluderend kan worden gesteld dat, hoewel toekomstig onderzoek vereist is om onder andere normscores te ontwikkelen en enkele psychometrische eigenschappen verder te onderzoeken, de ANTAT-CU RT scoringswijze een geschikt alternatief biedt voor de ANTAT-CU OT scoringswijze. Hiermee levert het huidige onderzoek een belangrijke bijdrage aan de klinische toepasbaarheid van de ANTAT-CU RT scoringswijze, die minder arbeidsintensief is en op een snelle en zo objectief mogelijke wijze inzicht geeft in de verbaal-mondelinge effectiviteit bij PMA.

Referenties

- Bastiaanse, R. (2010). *Afasie*. Houten: Bohn Stafleu van Loghum.
- Barnes, S., & Bloch, S. (2018). Why is measuring communication difficult? A critical review of current speech pathology concepts and measures. *Clinical Linguistics & Phonetics*, 33(3), 219-236. <https://doi.org/10.1080/02699206.2018.1498541>
- Beeke, S., Maxim, J., Best, W., & Cooper, F. (2011). Redesigning therapy for agrammatism: Initial findings from the ongoing evaluation of a conversation-based intervention study. *Journal of Neurolinguistics*, 24(2), 222–236. <https://doi.org/10.1016/j.jneuroling.2010.03.002>
- Beeke, S., Maxim, J., & Wilkinson, R. (2007). Using conversation analysis to assess and treat people with aphasia. *Seminars in Speech and Language*, 28(2), 136–147. <https://doi.org/10.1055/s-2007-970571>
- 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. <https://doi.org/10.1080/02687039008249085>
- Blomert, L. & Koster, C. (2008). *ANTAT: Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden: Handleiding*. Amsterdam: Hogrefe Uitgevers B.V.
- Blomert, L., Koster, Ch. & Kean, M.-L. (1995). *Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheden: Handleiding*. Lisse: Swets & Zeitlinger.
- Boeije, H.R., 't Hart, H. & Hox, J. (2009). *Onderzoeksmethoden*. Den Haag: Boom Lemma uitgevers.
- Bouwens, S. (2010). Neuropsychologisch onderzoek en het alledaags functioneren. *Neuropraxis*, 14, 133–137. <https://doi-org.ru.idm.oclc.org/10.1007/BF03089806>
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). speech and language therapy for aphasia after stroke: An updated systematic review and meta-analyses. *Stroke*, 47(10), E236–E237. <https://doi.org/10.1161/STROKEAHA.116.014439>
- Burns, A. C., & Bush, R. F. (2016). *Principes van marktonderzoek*. Amsterdam: Pearson.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81. <https://doi.org/10.1016/j.jml.2003.08.004>
- Corpus Gesproken Nederlands. (2000). *Protocol voor Orthografische Transcriptie*. Geraadpleegd op 20 juni 2021, van https://ivdnt.org/images/stories/producten/documentatie/cgn_website/doc_Dutch/topics/index.htm
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Dassek, S. (2016). *Eerste stappen richting validiteitsonderzoek en betrouwbaarheidsonderzoek van de ANTAT-CU4: Onderzoek naar de test-hertest sensibiliteit en betrouwbaarheid van de ANTAT-CU4*. Afstudeerproject Master Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- De Heus, P., Van der Leeden, R., & Gazendam, B. (2002). *Toegepaste data-analyse: technieken voor niet-experimenteel onderzoek in de sociale wetenschappen*. 's-Gravenhage: Elsevier.
- Dek, J.E., Braak, M.S.L., & Kooij, A.P. (2013). *Interpretatie van klinische resultaten: het interpreteren van de resultaten van cliënten uit een klinische groep*. Amsterdam: Pearson Assessment & information BV.

- De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex*, *14*(1), 41-49. [https://doi.org/10.1016/s0010-9452\(78\)800069](https://doi.org/10.1016/s0010-9452(78)800069)
- Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology*, *34*(4), 492-514. <https://doi.org/10.1080/02687038.2019.1702848>
- Doorn, P. K. & Rhebergen, M. P. (2006). *Correlatie en regressie*. Leiden: Universiteit Leiden.
- Drenth, P. J., & Sijtsma, K. (2001). *Testtheorie: Inleiding in de theorie.-Van de psychologische test en zijn toepassingen*. Houten: Bohn Stafleu van Loghum.
- Elman, R. J., & Bernstein-Ellis, E. (1995). What is functional? *American Journal of Speech Language Pathology*, *4*(4), 115–117. <https://doi.org/10.1044/1058-0360.0404.115>
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingsstelsel voor de kwaliteit van tests (gewijzigde herdruk)*. Zaandijk: Heijnis & Schipper.
- Field, A. (2009). *Discovering Statistics Using SPSS (3th ed.)*. London: Sage Publications.
- Graetz, P., De Bleser, R., & Willmes, K. (1992). *Akense Afasie Test*. Göttingen: Harcourt Assessment B.V.
- Hallgren, K.A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hattie, J., Cooksey, R.W. (1984). Procedures for Assessing the Validities of Tests Using the "Known-Groups". *Method. Applied Psychological Measurement*, *8*(3), 295–305. <https://doi.org/10.1177/014662168400800306>
- Holland, A. L. (1980). *CADL: Communicative abilities in daily living: A test of functional communication for aphasic adults*. Baltimore: University Park Press.
- Holland, A., Frattali, C., & Fromm, D. (1999). *CADL-2: Communication activities of daily living*. Austin, Texas: Pro-ed.
- Hüsters, S., & Jonkers, R. (2017). Het verschil tussen praten en communiceren: De functionele communicatie bij patiënten met cognitieve communicatiestoornissen. *Stem-, spraak- en taalpathologie*, *22*, 29–50. Geraadpleegd op 24 mei 2021, van <https://www.afasienet.com/wp-content/uploads/HusersJonkers2017.pdf>
- IBM (2020) [Software]. Opgehaald van: <https://www.surfspot.nl/ibm-spss-statistics-24.html>
- Irwin, W. H., Wertz, R. T., & Avent, J. R. (2002). Relationships among language impairment, functional communication, and pragmatic performance in aphasia. *Aphasiology*, *16*(8), 823–835. <https://doi.org/10.1080/02687030244000086>
- Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge, UK: Cambridge University Press.
- Kessels, R., Eling, P., Ponds, R., Spikman, J., & Van Zandvoort, M. (2018). *Klinische neuropsychologie*. Amsterdam: Boom B.V.
- Koo, T. & Li, M. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kooiker, R., Broekhoff, M., & Stumpel, H. (2011). *Marktonderzoek*. Houten: Noordhoff Uitgevers.
- Linnik, A., Bastiaanse, R., & Hohle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology*, *30*(7), 765–800. <https://doi.org/10.1080/02687038.2015.1113489>
- Long, A., Hesketh, A., & Bowen, A. (2009). Communication outcome after stroke: A new measure of the carer's perspective. *Clinical Rehabilitation*, *23*(9), 846–856. <https://doi.org/10.1177/0269215509336055>

- Long, A., Hesketh, A., Paszek, G., Booth, M., & Bowen, A. (2008). Development of a reliable self report outcome measure for pragmatic trials of communication therapy following stroke: The Communication Outcome after Stroke (COAST) scale. *Clinical Rehabilitation*, 22(12), 1083–1094. <https://doi.org/10.1177/0269215508090091>
- Lotgering, L. (2009). *Functionele communicatie bij afasiepatiënten: het meten van de communicatieve efficiëntie met de ANTATcu als operationalisatie van de eenheden van de ANTATprop*. Afstudeerproject Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- Marini, A., Caltagirone, C., Pasqualetti, P., & Carlomagno, S. (2007). Patterns of language improvement in adults with non-chronic non-fluent aphasia after specific therapies. *Aphasiology*, 21(2), 164–186. <https://doi.org/10.1080/02687030600633799>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago and London: The University of Chicago Press.
- NVAT (2015). Afasie Interventie Schema van de Nederlandse Vereniging van Afasietherapeuten. (NAIS). Geraadpleegd op 14 mei 2021, van https://www.afasienet.com/wp-content/uploads/NAIS_maart2015.pdf
- NVLF (2017). *Handreiking: Oefenen met de ICF*. Woerden: Nederlandse vereniging voor Logopedie en Foniatrie.
- Otters, M. (2019). *ANTAT CU Real Time: Een gekwantificeerde beoordeling van de verbaal mondelinge effectiviteit en efficiëntie middels de ANTAT*. Afstudeerproject Master Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.
- Paemeleire, F. (2010). *Personen met afasie én executieve functiestoornissen: implicaties voor diagnostiek en behandeling*. Geraadpleegd op 1 november 2021, van https://www.neurocom.be/uploads/28341443032010_Afasie_en_Executieve_Functiestoornissen_Paemeleire_Logopedie.pdf
- Piai, V. (z.j.). *Projecten Veni: Taal in het brein, van links naar rechts (Veni 451-17-003)*. Geraadpleegd op 11 oktober 2021, van <https://www.nwo.nl/onderzoeksprogrammas/nwo-talentprogramma/projecten-veni/2017>
- Pope, G. (2009). *Psychometrics 101: Item Total Correlation*. Geraadpleegd op 10 januari 2022, van <https://www.questionmark.com/item-total-correlation/>
- Prins, R. S., & Bastiaanse, Y. R. M. (1997). Afasie: Symptomatologie en wetenschappelijke inzichten. In Y. R. M. Bastiaanse, H. Peters, J. Van Borsel, P. H. O. Dejonckere, S. Van der Meulen, & B. J. E. Mondelaers (Eds.), *Handboek Stem-, Spraak- en Taalpathologie*. Houten: Bohn Stafleu Van Loghum.
- Prins, R. S., & Bastiaanse, Y. R. M. (2001). *Spontane-taalanalyse bij afasie. Stem-, spraak- en taalpathologie*, 10(1), 3–23. Geraadpleegd op 14 december 2021, van <https://sstp.nl/article/view/19268>
- Ramsberger, G. (1994). Functional perspective for assessment and rehabilitation of persons with severe aphasia. *Seminars in Speech and Language*, 15(1), 1–16. <https://doi.org/10.1055/s-2008-1064130>
- Ranganathan, P., Pramesh, C.S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in Clinical Research*, 8(4), 187-191. https://doi.org/10.4103/picr.PICR_123_17
- RIVM (2002). *ICF: Nederlandse vertaling van de International Classification of Functioning, Disability and Health*. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu/Nederlands WHO FIC Collaborating Centre.

- Roomer, E.K., Brok, S., Hoogerwerf, A.C., & Linn, D.E. (2011). *Handleiding Boston BenoemTaak*. Geraadpleegd op 10 november, van <https://docplayer.nl/10643585-Boston-benoemtaak-2011.html>
- 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. <https://doi.org/10.1080/02687038.2011.569892>
- Ruiter, M. B., Lotgering, E., & Rietveld, A.C.M. (2016). *Afnameprotocol ANTAT-CU* [Ongepubliceerd manuscript]. Radboud Universiteit, Nijmegen.
- Ruiter, M. B, Piai, V., & Roelofs, A. (z.j.). *Een pilotstudie naar de effectiviteit van SimpTell: Leidt webgebaseerde telegramstijltherapie tot een verbetering in de communicatie bij mensen met afasie? (Trial NL7855)*. Geraadpleegd op 11 oktober 2021, van <https://www.afasienet.com/professionals/wetenschap/overzicht-lopend-onderzoek/simptell-goed-verhaal-korte-zinnen/>
- Ruiter, M. B., Roelofs, A., Piai, V., Datadien, A., Steenbeek-Planting, E., Van Engelen, R., & Hendriks, I. (2019). SimpTell. Stoppen met praten op de automatische piloot [Using Human Language Technology to prevent aphasic speakers from producing disordered language on autopilot]. *DIXIT*, 16, 34-35. Geraadpleegd op 13 november 2021, van <https://repository.ubn.ru.nl/handle/2066/218142>
- Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 467–474. <https://doi.org/doi:10.22237/jmasm/1257035100>
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Georgetown University Roundtable on language and linguistics. Analyzing discourse: Text and talk* (pp. 71-93). Washington D.C.: Georgetown University Press.
- Thompson, C. K. & Worrall, L. (2008). Approaches to aphasia treatment. In C.K. Thompson, L. Worrall & N. Martin (Ed.), *Aphasia Rehabilitation. The Impairment and Its Consequences* (pp. 3–24). San Diego, CA: Plural Publishing, Inc.
- Van der Meulen, I., Van de Sandt-Koenderman, M., Van Gelder-Houthuizen, J., Wiegers, J., & Wielart, S. (2008). *Scenario Test verbale-en non verbale communicatie bij afasie*. Houten: Bohn Stafleu Van Loghum.
- Vander Stichele, R., Van Maele, G., De Paepe, P., Petrovic, M., & Van Bortel, L. (2009). Procentuele verschillen in kleine vergelijkende klinische studies: wanneer zijn verschillen statistisch significant en klinisch relevant? *Tijdschrift voor geneeskunde*, 65(12), 544–547. <https://doi.org/10.2143/TVG.65.12.2000557>
- Vigliocco, G., Permiss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1651), 1-7. <https://doi.org/10.1098/rstb.2013.0292>
- Visch-Brink, E. G. (2006). Rotterdamse Afasie Therapie Studie. *Tijdschrift voor neurologie en neurochirurgie*, 107(4), 168–176.
- Visch-Brink, E., Vandenborre, D., Jung de Smet, H., & Mariën, P. (2014). *CAT-NL, Comprehensive Aphasia Test Nederlandstalige bewerking*. Enschede: Ipskamp Drukkers B.V.
- Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., ... Gauvreau, C. A. (2016). Which outcomes are most important to people with aphasia and their families? an international nominal group technique study framed within the ICF. *Disability and Rehabilitation*, 39(14), 1364–1379. <https://doi.org/10.1080/09638288.2016.1194899>
- WHO (2001). *ICF: International Classification of Functioning, Disability and Health*. Geneva: World Health Organization.

- Wirz, S. L., Skinner, & Dean. (1990). *Revised edinburgh functional communication profile*. Tucson, AZ: Communication Skill Builders.
- Worrall, L., McCooley, R., Davidson, B., Larkins, B., & Hickson, L. (2002). The validity of functional assessments of communication and the Activity/Participation components of the ICIDH-2: do they reflect what really happens in real-life? *Journal of Communication Disorders*, 35(2), 107–137 [https://doi.org/10.1016/s0021-9924\(02\)00060-6](https://doi.org/10.1016/s0021-9924(02)00060-6)
- 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. <https://doi.org/10.1044/jshd.4501.27>
- Zwartjens, J. (2017). *ANTAT-CU4: Een onderzoek naar de betrouwbaarheid van de ANTAT CU4 bij gezonde sprekers*. Afstudeerproject Master Taal- en Spraakpathologie, Radboud Universiteit Nijmegen.

Bijlagen

Op verzoek van M.B. Ruiter niet zichtbaar.