DE NEDERLANDSE BENOEM TEST VOOR EEN VLAAMSE POPULATIE: EEN AANGEPASTE CORRECTIESLEUTEL

Dorien Vandenborre[1,2], Lien Van Rompaey[1], Eline Verwilligen[3], en Frank Paemeleire[3,4]

[1] Thomas More hogeschool, opleiding Logopedie en Audiologie

[2] Thomas More hogeschool, expertisecentrum Mens en welzijn

[3] AZ Maria Middelares ziekenhuis, Gent

[4] Arteveldehogeschool, Gent

Functiegerichte diagnostiek met de Nederlandse Benoem Test (NBT) is waardevol om een anomie kwantitatief en kwalitatief in kaart te brengen. Een betrouwbare scoring bij een Vlaamse populatie is moeilijk op basis van de huidige Nederlandse correctiesleutel omdat regionale taalverschillen een impact kunnen hebben op de score.

Bij 248 volwassen Vlaamse proefpersonen zonder hersenletsel werd een mixed method onderzoek uitgevoerd met een kwalitatief luik (i.e. opstellen van een Vlaamse correctiesleutel) en kwantitatief observationeel cross-sectioneel luik (i.e. bepalen van betrouwbaarheid van testscore). Studenten verzamelden de reacties van de proefpersonen zonder hersenletsel. Op basis van de scoringscriteria van de oorspronkelijke NBT werd een Vlaamse correctiesleutel opgesteld. Een intraclass correlatiecoëfficiënt werd berekend om de betrouwbaarheid van deze aangepaste correctiesleutel na te gaan.

Een aangepaste, Vlaamse correctiesleutel werd opgesteld met meer dan 400 antwoordmogelijkheden. Het scoren van de NBT aan de hand van de aangepaste correctiesleutel levert een hogere intraclass correlationcoëfficiënt op (ICC =.85) in vergelijking met de scores bekomen aan de hand van de Nederlandse correctiesleutel (ICC =.57).

De aangepaste correctiesleutel zorgt ervoor dat de testleider op een meer betrouwbare, consistente en vlottere manier de totaalscore kan berekenen.

Trefwoorden
Volwassenen, anomie, diagnostiek

Inleiding

Woordvinding

Het ophalen van woorden is normaal gezien een snel en efficiënt proces. Dit blijkt uit het feit dat we twee tot drie woorden per seconde kunnen produceren met niet meer dan één of twee fouten per duizend woorden (Cuetos & Alija, 2003). Volwassenen met een niet-aangeboren hersenletsel hebben vaak een anomie (Palmer et al., 2017). Anomie duidt op het onvermogen om toegang te krijgen tot een woord bij een intacte objectherkenning en/of om het te gebruiken in mondelinge/schriftelijke communicatie (Macoir & Lavoie, 2021). Een anomie leidt tot beperkingen in de communicatie, een kernactiviteit in ons dagelijks leven (Herbert et al., 2008). Daarenboven hindert beperkt communiceren het contact met de sociale omgeving, wat kan leiden tot sociaal isolement (Berns et al., 2015). Typische klinische symptomen die wijzen op anomie zijn zelfcorrecties, (lange) pauzes, het gebruik van 'vullers', inhoudloze spraak en semantische of fonematische parafasieën (Macoir & Lavoie, 2021). Een grondig assessment naar de aard en de ernst van de anomie vormt een waardevol aanknopingspunt om samen met de persoon behandeldoelen op maat te formuleren.

Woordvinding in kaart brengen

Om de woordvinding van personen met niet-aangeboren hersenletsel kwantitatief (aantal fouten) en kwalitatief (soort fouten) in kaart te brengen bestaan er verschillende onderzoeksinstrumenten op participatie-, activiteiten- en functieniveau. Op **participatieniveau** kan de testleider de persoon met afasie observeren in verschillende communicatieve situaties (bv. in gesprek met de arts/de partner/een familielid, in een bepaalde context (ziekenhuissetting of thuis) en onder bepaalde voorwaarden (wel/geen tijdsdruk). Op **activiteitenniveau** kan de testleider **gebonden taal** beoordelen. Gebonden taal is een continue opeenvolging van uitingen die een spreker verbaal (mondeling of schriftelijk) produceert om gedachten en ideeën op een betekenisvolle manier over te brengen (Crystal, 1980). Als testleider verzamel je gebonden taal op verschillende manieren. Zo kan je een spontaan gesprek voeren met de persoon, de persoon kan een logische reeks of situatieplaat beschrijven of de persoon kan een verhaal navertellen. Op **functieniveau** kan de persoon een verbale woordvlotheidstaak uitvoeren, een woord ophalen bij een omschrijving of een afbeelding benoemen (Brysbaert & Ellis, 2015). Bij een verbale woordvlotheidstaak somt de persoon binnen een bepaalde tijd zoveel mogelijk woorden op passend bij een bepaald criterium. Er zijn semantische (bv. "noem

zoveel mogelijk dieren"), fonologische (bv. "noem zoveel mogelijk woorden beginnend met een d-") en alternerende woordvlotheidstaken (bv. "noem alternerend zoveel mogelijk dieren en woorden met d-"). Als de persoon een woord ophaalt bij een omschrijving, dan geeft de testleider een semantische omschrijving van een object of van een actie en de persoon roept het passende doelwoord op. Een afbeelding benoemen, ook wel een visuele confrontatietaak genoemd, is de meest gebruikte functietaak (Boyle, 2014). De testleider kan statische (prent) of dynamische (video) afbeeldingen gebruiken, in zwart-wit of in kleur, van voorwerpen of van acties (Cuellar & Del Toro, 2017). Bij het kiezen van testitems wordt rekening gehouden met psycholinguïstische factoren. Zo kunnen de testresultaten een foutenpatroon duidelijk maken, waaraan de behandelaar samen met de persoon concrete behandeldoelen koppelt. Een groot voordeel van een visuele confrontatietaak is dat de testleider weet wat het doelwoord is en zo accurater de onderliggende stoornis kan nagaan (Friedman et al., 2013). Een nadeel is dat de resultaten negatief beïnvloed worden door visuele perceptiestoornissen (Paemeleire, 2005). Een tweede nadeel is dat een visuele confrontatietaak geïsoleerd kijkt naar het benoemen van objecten/acties en geen volledig beeld geeft van het communiceren in een alledaagse situatie of het actief participeren in de maatschappij (Hofs-van Kats, 2019). Om een beeld te krijgen van de daadwerkelijke woordvinding in het dagelijks leven en de mate waarin de persoon daar zelf hinder van ondervindt, is een holistische kijk nodig.

Visuele confrontatietaken maken deel uit van uitgebreide stoornisgerichte tests, zoals de Comprehensive Aphasia Test (CAT-NL: Visch-Brink et al., 2014). Er bestaan ook geïsoleerde visuele confrontatietaken, zoals de Boston Benoem Taak (BBT: Roomer et al., 2011) en de Nederlandse Benoemtest (NBT: van Ewijk et al., 2018b). De BBT bestaat uit 60 zwart-wit lijntekeningen en is een letterlijke vertaling van de Amerikaans-Engelse Boston Naming Test (BNT: Kaplan et al., 1983). In 2011 maakten Roomer et al. een aangepaste scoringssleutel voor het Nederlands. De testitems zijn geordend volgens woordfrequentie in de Amerikaanse taal en de afbeeldingen zijn afgestemd op de Amerikaanse cultuur. Noch de taal, noch de cultuur zijn aangepast aan een Nederlandse en/of Vlaamse context. Daarnaast plaatsen onderzoekers ook andere kanttekeningen bij de BNT: (1) de test voldoet niet aan de meest recente wetenschappelijke inzichten rond woordvinding, waarbij woordfrequentie en verwervingsleeftijd een rol spelen (bv. Bastiaanse et al., 2016; Brysbaert & Ellis, 2015); (2) de psychometrische eigenschappen zijn onvoldoende (bv. Harry & Crowne, 2014); en (3) de normen zijn verouderd (bv. Shao et al., 2014). Om een antwoord te bieden op deze kritieken ontwikkelden van Ewijk et al. (2018b) de NBT.

Nederlandse Benoem Test (NBT)

Deze test bestaat uit 92 afbeeldingen in kleur en is opgebouwd volgens de woordfrequentie en verwervingsleeftijd van de Nederlandse taal (Schaerlaekens, 2007; Gerhand & Barry, 1999). De test is tussen 2018 en 2019 genormeerd voor een Nederlandse populatie. Aanvullend op de normtabellen stelden Dijkhuis et al. (2020) in een whitepaper ook een correctietabel op om de invloed van de demografische factoren 'leeftijd' en 'opleidingsniveau' te neutraliseren. In 2021 brachten Dijkhuis et al. een tweede whitepaper uit met een uitgebreide correctiesleutel en licht aangepaste scoringsregels. De auteurs van dit artikel gingen na het verschijnen met deze nieuwe test in hun klinische praktijk aan de slag. Een aantal voordelen werd duidelijk: de aantrekkelijke vormgeving, de handige presentatiewijze, de eenvoudige afname, de snelle scoring en de sensitiviteit voor detectie van lichte stoornissen. Er werden uiteraard ook beperkingen duidelijk. Ten eerste heeft de test een hoge aankoopprijs (momenteel 527 euro) wat een drempel vormt voor logopedisten die slechts met een beperkt aantal personen met afasie werke. Ten tweede de lengte van de test: met zijn 92 testitems is het veruit de langste benoemtest op de markt. Intussen maakten Alons et al. (2022) een screeningsversie (met vijf tot tien items) van de test. Een derde beperking was dat voor bepaalde items er geen 1:1 correspondentie tussen de visuele stimulus en het doelwoord was en de Nederlandse correctiesleutel deze scoringsonduidelijkheden niet altijd oploste. Nederlanders en Vlamingen spreken dan wel dezelfde taal, ze kenmerken zich door eigen culturele waarden en een specifiek woordgebruik (Gerritsen & Claes, 2016). Dit wekte het vermoeden dat de normtabellen en correctiesleutel onvoldoende op maat zijn van een Vlaamse populatie.

Woordvinding psychometrisch correct in kaart brengen

Ivanova en Hallowell (2013) beschrijven een protocol om een taaltest in eender welke taal te ontwikkelen. Naast beïnvloedende factoren bij testontwikkeling, testafname en scoring beschrijven ze ook beïnvloedende factoren bij het opstellen van normen en het nagaan van betrouwbaarheid en validiteit. Om normen te kunnen bepalen is het belangrijk dat onderzoekers de test afnemen bij een voldoende grote steekproef van mensen zonder een hersenletsel. Franzen (2003) spreekt over een normgroep van tenminste 100 personen. Onderzoekers (bv. Mitrushina et al., 2005) geven aan dat demografische factoren (bv. leeftijd, opleidingsniveau en sociaal-economische status)

onderzoeksresultaten kunnen beïnvloeden, waardoor de normgroep representatief moet zijn voor de doelgroep voor wie de test bedoeld is.

Betrouwbaarheid verwijst naar de consistentie, stabiliteit en nauwkeurigheid van een meetinstrument. Als het gaat om interbeoordelaarsbetrouwbaarheid, dan ga je na in hoeverre de onderzoeker een invloed heeft op de testprestatie van een persoon. Gedetailleerde scoringsrichtlijnen kunnen de interbeoordelaarsbetrouwbaarheid verhogen (Ivanova & Hallowell, 2013).

Onderzoeksvragen

We formuleren een onderzoeksdoel, een onderzoeksvraag en twee deelvragen.

Onderzoeksdoel: het opstellen van een Vlaamse correctiesleutel op basis van testafnames bij een Vlaamse populatie zonder hersenletsel.

Onderzoeksvraag: Is de Nederlandse correctiesleutel voldoende betrouwbaar om in te zetten bij Vlaamse proefpersonen zonder hersenletsel?

Is er een overeenstemming tussen de scores van de verschillende student-onderzoekers met de oorspronkelijke Nederlandse correctiesleutel?

Is er een overeenstemming tussen de scores van de verschillende student-onderzoekers met de aangepaste Vlaamse (Nederlands-Belgische) correctiesleutel?

Methodiek

Dit onderzoek is een *mixed method onderzoek met een kwantitatief observationeel cross-sectioneel luik voor het bepalen van de interbeoordelaarsbetrouwbaarheid en een kwalitatief luik voor het opstellen van een Vlaamse correctiesleutel.*

Proefpersonen

De NBT werd afgenomen bij proefpersonen zonder hersenletsel tussen 20 en 90 jaar met Vlaams (Belgisch-Nederlands) als moedertaal, zonder voorgeschiedenis van een neurologisch of psychiatrisch ziektebeeld en met een normaal gehoor- en gezichtsvermogen (eventueel aangepast met hoorapparaat en/of bril). Handvoorkeur, geslacht, leeftijd, opleidingsniveau en regio werden mee opgenomen, maar vormden geen exclusiecriteria voor deelname aan het onderzoek.

Aan de hand van een poweranalyse werd de gewenste grootte van de steekproef bepaald. We bepaalden een foutenmarge van 5%, een betrouwbaarheidsniveau van 90%, een onderzoekspopulatie van 6 miljoen en een mate van spreiding van 50%. Dit leverde een steekproefgrootte van 272 proefpersonen op. Via het Belgisch Statistiekbureau (Statbel, 2021) werd een verdeling opgesteld van de steekproefgroep (Tabel 1). Er werd gefocust op deelgroepen volgens geslacht (man/vrouw), leeftijd (20- tot 29-jarigen, 30- tot 49-jarigen, 50- tot 69-jarigen en 70-jarigen of ouder), opleidingsniveau (hoog-/laagopgeleid) en regio (woonachtig in West-Vlaanderen, Oost-Vlaanderen, Antwerpen, Vlaams-Brabant en Limburg). Hoogopgeleide proefpersonen werden vastgelegd als personen die tenminste twee jaren hoger onderwijs succesvol afrondden.

Tabel 1. Streefcijfers voor dataverzameling (Statbel, 2021)

Regio	Oost-Vlaanderen		West-Vlaanderen		Vlaams-Brabant		Antwerpen		Limburg	
Voor 272 proefpersonen	64		48		48		76		36	
Geslacht	32 (M)	32 (V)	24 (M)	24 (V)	24 (M)	24 (V)	38 (M)	38 (V)	18 (M)	18 (V)
Leeftijdscategorie (20 tot 29 jaar)	4 (L) 1 (H)	4 (L) 1 (H)	3 (L) 1 (H)	3 (L) 1 (H)	4 (L) 1 (H)	4 (L) 1 (H)	5 (L) 1 (H)	5 (L) 1 (H)	2 (L) 0 (H)	2 (L) 0 (H)
Leeftijdscategorie (30 tot 49 jaar)	6 (L) 5 (H)	6 (L) 5 (H)	4 (L) 4 (H)	4 (L) 4 (H)	4 (L) 4 (H)	4 (L) 4 (H)	7 (L) 7 (H)	7 (L) 7 (H)	4 (L) 3 (H)	4 (L) 3 (H)
Leeftijdscategorie (50 tot 69 jaar)	7 (L) 5 (H)	7 (L) 5 (H)	5 (L) 4 (H)	5 (L) 4 (H)	5 (L) 3 (H)	5 (L) 3 (H)	7 (L) 6 (H)	7 (L) 6 (H)	4 (L) 3 (H)	4 (L) 3 (H)
Leeftijdscategorie (≥70 jaar)	2 (L) 2 (H)	3 (L) 1 (H)	2 (L) 1 (H)	2 (L) 1 (H)	2 (L) 1 (H)	2 (L) 1 (H)	3 (L) 2 (H)	4 (L) 1 (H)	2 (L)	2 (L)

Legende: M=man, V=vrouw, L=laagopgeleid, H=hoogopgeleid

Dataverzameling

Tussen september 2022 en januari 2023 verzamelden vier studenten van Thomas More hogeschool (TM) en vier studenten van de Arteveldehogeschool (AH) data. Ze rekruteerden elk 30 proefpersonen via sociale media of gerichte mailing, rekening houdend met de diversiteit van de steekproef. De student-onderzoekers oefenden vooraf het testprotocol om een gestandaardiseerde afname te

waarborgen. Onduidelijkheden of vragen werden genoteerd en besproken met de onderzoekers. Voor het eigenlijke testprotocol stemden de proefpersonen aan de hand van een geïnformeerde toestemming en informatiebrief formeel in met het testprotocol.

Het testprotocol bestond uit drie delen: een vragenlijst, de Nederlandse Benoem Test (NBT: van Ewijk et al., 2018) en de Montréal Cognitive Assessment (MoCA, versie 7: Dautzenberg & de Jonghe, 2004). De testleider mocht de proefpersoon niet cuen tijdens de afname en er was geen afbreeknorm. Elke proefpersoon doorliep het volledige testprotocol. De *vragenlijst* peilde aan de hand van 18 vragen naar de demografische gegevens en medische voorgeschiedenis van de proefpersoon. De *NBT* bestaat uit 92 testitems en is opgebouwd volgens de woordfrequentie en verwervingsleeftijd van de aangeboden items (Schaerlaekens, 2007; Gerhand & Barry, 1999): de test start met 21 vroeg verworven hoog frequente woorden (EHF), dan volgen 17 vroeg verworven midden frequente woorden (EMF), 11 laat verworven hoogfrequente woorden (LHF), 14 vroeg verworven laagfrequent woorden (ELF), 13 laat verworven middenfrequente woorden (LMF) en tot slot 15 laat verworven laagfrequente woorden (LLF). De testitems werden digitaal aangeboden in de vorm van een PowerPointpresentatie waarbij elk item conform de afnamerichtlijnen 10 seconden zichtbaar bleef. De *MoCA* is een screeningsinstrument dat de cognitieve vaardigheden van de proefpersonen nagaat. De proefpersoon scoort maximaal 30 punten. Als de proefpersoon onder de normscores opgesteld door Kessels et al. (2022) scoorde, dan werd deze uitgesloten. De testafname gebeurde op verschillende locaties, maar telkens in een prikkelarme omgeving. Het volledige testprotocol duurde 30 tot 60 minuten. De student-onderzoekers maakten een audio-opname met de smartphone om de data achteraf woordelijk te transcriberen. Na de transcripties werden de opnames verwijderd. Alle transcripties werden uitgeschreven en in een gedeeld Excel-document bewaard.

Dataverwerking

Kwalitatief luik: de correctiesleutel

Een kwalitatieve analyse werd uitgevoerd met het oog op het ontwikkelen van een aangepaste Vlaamse correctiesleutel. Bij het scoren met de scoringsregels (Tabel 2) en de oorspronkelijke Nederlandse correctiesleutel (Dijkhuis et al., 2021) noteerden de student-onderzoekers onduidelijkheden en inconsistenties. Vervolgens werden de 92 testitems evenredig verdeeld over de student-onderzoekers. Iedere student-onderzoeker voerde voor de toegewezen 15 testitems een grondige analyse uit aan de hand van de gegevens van alle proefpersonen. Ze noteerden per testitem alle mogelijke responsen op de vierpuntschaal en noteerden bijkomende onduidelijkheden en inconsistenties. De student-onderzoekers, auteurs van dit artikel (onderzoekers) en een van de oorspronkelijke auteurs van de NBT bespraken samen alle twijfelgevallen om een uniforme scoringsmethode te bekomen. Om een consistente beoordeling te waarborgen, raadpleegden ze drie online woordenboeken, i.e. het Van Dale woordenboek, het Synoniemenwoordenboek en het woordenboek van de Vlaamse Dialecten. Ze bereikten een consensus voor de deelscores van alle 92 testitems. Na consensus finaliseerde elke student-onderzoeker de correctiesleutel voor de 15 toegewezen testitems. Dit leverde een tabel op met per testitem het doelwoord, de responsen en de score. De onderzoekers voegden bij elke respons in de Vlaamse correctiesleutel een motivatie voor de scoring toe (zie tabel 2). Ze baseerden zich bij de motivering op de scoringsregels uit Whitepaper 2 van de Nederlandse Benoem Test (Dijkhuis et al., 2021). Deze tweede scoringsronde resulteerde in een aangepaste, Vlaamse correctiesleutel.

Tabel 2. Scoringsregels voor toekennen score bij NBT voor Vlaamse correctiesleutel

Score 3	**Juiste benoeming**, i.e. het doelwoord (in meervoud of enkelvoud); een synoniem of dialectvorm; of een onnodig specifieke benoeming
Score 2	**Goed passende juiste benoeming**, i.e. een semantisch goed passende, foute benoeming; een semantisch goed passende, foute omschrijving; een semantisch bovenbegrip; een groot deel van de afbeelding correct benoemd
Score 1	**Enigszins passende juiste benoeming**, i.e. een semantisch enigszins passende, foute benoeming; een semantisch enigszins passende, foute omschrijving; een te hoog semantisch bovenbegrip; een werkwoordsvorm of adjectief waarvan de betekenis gerelateerd is aan het doelwoord; een klein deel van de afbeelding correct benoemd
Score 0	**Niet-passende benoeming**, i.e. een niet-passende benoeming; of een niet-passende omschrijving

Kwantitatief luik

Direct na afname scoorde elke student-onderzoeker de eigen 30 proefpersonen. Ze bepaalden een ruwe score op de MoCA en een ruwe score op de NBT aan de hand van een vierpuntbeoordelingsschaal (score 3= juiste benoeming; score 2=goed passende benoeming; score 1=enigszins passende benoeming; en score 0 = niet-passende benoeming) (Tabel 2) en de oorspronkelijke Nederlandse correctiesleutel (eerste scoring). Na het opstellen van de aangepaste Vlaamse correctiesleutel scoorde elke

student-onderzoek de testafnames van de eigen 30 proefpersonen opnieuw (tweede scoring).

De ruwe scores werden geanonimiseerd verwerkt aan de hand van IBM SPSS Statistics 27. Voor alle proefpersonen werd via inductieve statistiek het gemiddelde, de standaarddeviatie en de range bepaald. Om de interbeoordelaarsbetrouwbaarheid te meten berekenden de onderzoekers de intraclass correlatiecoëfficiënt zowel voor de ruwe score met de oorspronkelijke, Nederlandse correctiesleutel (eerste scoring) als voor de ruwe score met de aangepaste, Vlaamse correctiesleutel (tweede scoring).

Resultaten

De resultaten zijn gebaseerd op de testafnames van 248 Vlamingen. De proefgroep week af van de vooropgestelde steekproef. Er werden te weinig proefpersonen (n=248) geïncludeerd voor een representatieve steekproef (n=272). Bepaalde demografische subgroepen zijn oververtegenwoordigd in vergelijking met de streeftabel (i.e. proefpersonen uit Oost- en West-Vlaanderen), andere demografische subgroepen zijn ondervertegenwoordigd (i.e. proefpersonen uit Antwerpen en Limburg). Tabel 3 geeft het volledige overzicht van de Vlaamse proefgroep.

Tabel 3. Effectieve steekproef

Regio	Oost-Vlaanderen		West-Vlaanderen		Vlaams-Brabant		Antwerpen		Limburg	
Voor 248 proef personen	70		72		44		48		14	
Geslacht	34 (M)	36 (V)	36 (M)	36 (V)	26 (M)	18 (V)	24 (M)	24 (V)	7 (M)	7 (V)
Leeftijdscategorie (20 tot 29 jaar)	5 (L) 1 (H)	5 (L) 1 (H)	5 (L) 1 (H)	10 (L) 1 (H)	5 (L) 1 (H)	4 (L) 0 (H)	5 (L) 1 (H)	5 (L) 4 (H)	1 (L) 1 (H)	2 (L) 1 (H)
Leeftijdscategorie (30 tot 49 jaar)	12 (L) 4 (H)	9 (L) 3 (H)	9 (L) 3 (H)	10 (L) 0 (H)	2 (L) 3 (H)	4 (L) 3 (H)	4 (L) 4 (H)	6 (L) 6 (H)	1 (L) 1 (H)	1 (L) 2 (H)
Leeftijdscategorie (50 tot 69 jaar)	9 (L) 1 (H)	10 (L) 4 (H)	10 (L) 4 (H)	11 (L) 0 (H)	7 (L) 2 (H)	4 (L) 2 (H)	4 (L) 4 (H)	0 (L) 1 (H)	1 (L) 2 (H)	0 (L) 1 (H)
Leeftijdscategorie (≥70 jaar)	2 (L) 0 (H)	4 (L) 0 (H)	4 (L) 0 (H)	4 (L) 0 (H)	6 (L) 0 (H)	1 (L) 0 (H)	0 (L) 2 (H)	2 (L) 0 (H)	0 (L)	0 (L)

Legende: M=man, V=vrouw, L=laagopgeleid, H=hoogopgeleid

Kwalitatief luik: de correctiesleutel

De responsen bij een Vlaamse populatie zonder niet-aangeboren hersenletsel zijn voor bepaalde testitems vergelijkbaar met de voorbeeldresponsen uit de Nederlandse correctiesleutel (zoals bv. voor de testitems 'gieter', 'paddenstoel' en 'zandbak'). Voor andere testitems wijken de responsen van de Vlamingen systematisch af van het Nederlandse doelwoord en van de Nederlandse scoringsvoorbeelden (Tabel 4). Zo gaat het om Franse leenwoorden die voor het Vlaams niet tellen als 'vreemde taal'. Enkele voorbeelden hierbij zijn 'trotinette' voor 'step', 'bavet' voor 'slab' en 'coffre fort' voor 'kluis'. In andere gevallen gaat het om een verschillend woordgebruik waarbij de oorsprong niet in de andere taal ligt, maar eerder gebaseerd is op beeldspraak (bv. 'muizenstrontjes' voor 'hagelslag' en 'schuifaf' voor 'glijbaan'). Ook andere typische Vlaamse benamingen komen aan bod (zoals 'moor' voor 'theepot', 'wasspeld' voor 'wasknijper' en 'waslapjes' voor 'washandjes'). Voor sommige items was het niet eenvoudig om tot een consistente score te komen. Zo kan je item 87 'kurkentrekker' bijvoorbeeld benoemen als een 'aftrekker'. De vraag is dan: gaat het hier om een dialectvorm (dan krijgt de proefpersoon score 3) of geeft de persoon een semantisch goed passende foute benoeming (dan krijgt de proefpersoon score 2). De twee scoringsrondes resulteerden in een aangepaste, Vlaamse correctiesleutel, waarbij meer dan 400 voorbeelden de scoringsregels verduidelijken. De definitieve correctiesleutel (Paemeleire, van Rompaey, Verwilligen & Vandenborre, 2023) kan gratis gedownload worden via www.neurocom.be.

Tabel 4. Enkele scoringsvoorbeelden uit beide correctielijsten

Item	Respons	Score o.b.v. Nederlandse correctiesleutel (Dijkhuis et al., 2021)	Score o.b.v. Vlaamse correctiesleutel
4. Jurk	kleed, vrouwenkleed	1	3
16. Koning	man met een cape, ridder, lakei, keizer	1	1
24. Banaan	geel, krom	0	1
38. Ijsje	likken, koud, smelten	1	1

Kwantitatief luik

De ruwe scores op de NBT zijn eerst bepaald met behulp van de Nederlandse correctiesleutel. De proefpersonen behalen een gemiddelde score van 265,13 op 276 met een standaarddeviatie van 8,02. De proefpersonen behalen een minimumscore van 231 en een maximumscore van 274. Om een antwoord te formuleren op de onderzoeksvraag werd de intraclass correlatiecoëfficiënt berekend. Uit de resultaten bleek dat er een matige overeenkomst is tussen de scores van de verschillende testleiders (ICC=.57). Deze overeenkomst bleek niet significant tot op het 1%-niveau (p=.381) (Tabel 5).

De ruwe scores van de NBT zijn opnieuw bepaald met behulp van de aangepaste Vlaamse correctiesleutel. De proefpersonen behalen dan een gemiddelde score van 268 met een standaarddeviatie van 5.31. De proefpersonen behalen een minimumscore van 234 en een maximumscore van 276. De intraclass correlatiecoëfficiënt werd opnieuw berekend. Uit de resultaten bleek dat er een goede overeenkomst is tussen de scores van de verschillende testleiders (ICC=.85). Deze overeenkomst bleek niet significant tot op het 1%-niveau (p=.590) (Tabel 5).

Tabel 5. Descriptieve maten

	Gemiddelde score	Standaard-deviatie	Minimum	Maximum	95% CI
Nederlandse correctiesleutel	265,13	8,02	231	274	.57 (p=.381)
Student-oz 1	268,60	7,95	251	274	
Student-oz 2	267,60	7,87	260	274	
Student-oz 3	259,80	9,08	231	273	
Student-oz 4	263,97	9,90	232	273	
Aangepaste, Vlaamse correctiesleutel	268	5.31	234	276	.85 (p=.590)
Student-oz 1	269,67	3,00	265	274	
Student-oz 2	269,47	5,00	257	275	
Student-oz 3	267,17	4,32	242	276	
Student-oz 4	267,68	5,01	234	274	

Legende: student-oz = student-onderzoeker

Discussie

Het opzet van dit onderzoek was om een aangepaste Vlaamse correctiesleutel op te stellen voor de NBT en om na te gaan of deze aangepaste correctiesleutel effectief zorgt voor een meer betrouwbare totaalscore.

De meerwaarde van een Vlaamse correctiesleutel sluit aan bij bevindingen in de literatuur en uit de klinische praktijk. Gerritsen en Claes (2017) en Shao et al. (2014) tonen aan dat regionale taalverschillen een impact kunnen hebben op de score. Een aangepaste, Vlaamse correctiesleutel focust specifiek op Vlaams woordgebruik.

Een kwantitatieve analyse, m.n. het berekenen van de intraclass correlatiecoëfficiënt objectiveert deze meerwaarde. In de whitepaper van de NBT vinden Dijkhuis et al. (2020) een sterke overeenkomst tussen de scores op de NBT tussen twee beoordelaars, een Nederlandse onderzoeker en een onafhankelijke logopedist (ICC =.85). In deze proefgroep bekomen vier Vlaamse student-onderzoekers slechts een matige overeenkomst (ICC=.57) tussen de scores op de NBT met de Nederlandse correctiesleutel. Eenzelfde analyse met de aangepaste Vlaamse correctiesleutel levert wel een goede overeenkomst op (ICC=.85). Hieruit kunnen we concluderen dat de aangepaste Vlaamse correctiesleutel een meerwaarde heeft voor de Vlaamse populatie.

Bij dit onderzoek kunnen we ook enkele kritische bedenkingen formuleren. Sterktes van dit onderzoek zijn dat (1) de aangepaste correctiesleutel uitgebreid is met meer dan 400 mogelijke reacties bij de 92 afbeeldingen; en (2) de aangepaste correctiesleutel ervoor zorgt dat de scoringsregels minder voor interpretatie vatbaar zijn en er een grotere consistentie is tussen verschillende testleiders. Er waren enkele uitdagingen. Ten eerste diende een consistente correctiesleutel over de testitems heen opgesteld te worden. Als bijvoorbeeld een merknaam voor het testitem 'auto' een score 3 krijgt dan moet ook de naam van een bepaalde krant een score 3 krijgen voor het testitem 'krant'. De tweede uitdaging was de tijd die de persoon krijgt om te reageren. De student-onderzoekers hielden

tijdens de testafname enkel rekening met de grens van 10 seconden. In de tweede whitepaper vermelden Dijkhuis et al. (2021) dat een score 2 wordt toegekend als de proefpersoon pas na een stilte van meer dan 5 seconden of een nietszeggende frase van meer dan 10 seconden het passende antwoord geeft.

In een vervolgonderzoek zal de proefgroep verder uitgebreid worden om een representatieve steekproef (leeftijd, regio, opleidingsniveau, geslacht) te bekomen. Momenteel zijn bepaalde demografische subgroepen onder- en andere oververtegenwoordigd. Een representatieve proefgroep zou toelaten om de invloed van demografische variabelen te analyseren en normgegevens op te stellen. Daarnaast kan data verzameld worden bij een proefgroep Vlamingen met hersenletsel en is het aangewezen om bij deze proefgroep de inter- en intrabeoordelaarsbetrouwbaarheid na te gaan.

Dankwoord

De auteurs bedanken de student-onderzoekers van Thomas More Hogeschool (Fee Bourgeois, Ingrid Debruyne, Delphine Meiresonne en Melissa Verweij) en de Arteveldehogeschool (Merel Amez, Marlies Decanniere, Sylvie De Rechter en Lara De Smedt) voor het verzamelen van de data. Onze dank gaat ook uit naar alle proefpersonen voor hun vrijwillige bijdrage aan het onderzoek.

Tot slot bedanken de auteurs ook de oorspronkelijke auteurs voor hun bijdrage bij het opstellen van de correctiesleutel.

Referenties

Alons, E., Dijkhuis, L., van Tuijl, P., & van Ewijk, L. (2022). *Nederlandse Benoem Test - Screening (NBT- Screening).* Houten: Bohn Stafleu van Loghum.

Bastiaanse, R., Wieling, M., & Wolthuis, N. (2016). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology, 30*(11), 1221-1239. https://doi.org/10.1080/02687038.2015.1100709

Berns, P. E. G., Jünger, N., Boxum, E., Nouwens, F., Staaij, M. G. van der, Wessel, S. van, et al. (2015). *Logopedische Richtlijn 'Diagnostiek en behandeling van afasie bij volwassenen'. Woerden: Nederlandse Vereniging voor Logopedie en Foniatrie.*

Boyle, M. (2014). Test-Retest Stability of Word Retrieval in Aphasic Discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966-978. https://doi.org/10.1044/2014_JSLHR-L-13-0171

Brysbaert, M. & Ellis, A. W. (2015). Aphasia and Age of Acquisition: Are Early-Learned Words More Resilient? *Aphasiology*, *vol. 30*, no. 11, pp. 1240–63. https://doi.org/10.1080/02687038.2015.1106439

Cuellar, M.E., Del Toro, C.M. (2017). Time-frequency analysis of mu rhythm activity during picture and video action naming tasks. Brain Science, 7(9), 114. https://doi.org/10.3390/brainsci7090114

Cuetos, F., & Alija, M. (2003). Normative data and naming times for action pictures. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 168-177. https://psycnet.apa.org/doi/10.3758/BF03195508

Crystal, D. (1980). A first dictionary of linguistics and phonetics. London: André Deutsch

Dautzenberg, P.L.J., & de Jonghe, J.F.M. (2004). *Nederlandse bewerking van Montréal Cognitive Assessment,* www.mocatest.org.

Dijkhuis, L., Methorst, T., Hofs-van Kats, M., & van Ewijk, L. (2020). *Whitepaper 1 bij de Nederlandse Benoem Test (NBT).* Testweb. https://testweb.bsl.nl/wp-content/uploads/sites/30/2021/06/A4-Whitepaper-def-1.pdf

Dijkhuis, L., van der Velden, H., & van Ewijk, L. (2021). *Whitepaper 2 bij de Nederlandse Benoem Test (NBT): scoringsvoorbeeldlijst en regels voor scoring.* Testweb. https://testweb.bsl.nl/wp-content/uploads/sites/30/2021/07/Whitepaper-2.NBT-Scoringsvoorbeeldlijst-en-regels-voor-scoring_.pdf

Dongbin, L., Zhiguo, W., Yun, W., Qinte, H., Zhenling, W., Tongdan, C., et al. Disagreement and factors between symptom on self-report and clinician rating of major depressive disorder: A report of a national survey in China. (2019). *Journal of Affective Disorders, 253*, 141-146. https://doi.org/10.1016/j.jad.2019.04.073

Franzen, M.D. (2003). *Reliability and validity in neuropsychological assessment.* New York: Kluwer Academic and Plenum Publishers.

Friedman, D.I., Grant, T.L., & Digre, K.B. (2013). Revised diagnostic criteria for the pseudotumor cerebri syndorme in adults and children. *Neurology, 81*(13), 1159-1165. https://doi.org/10.1212/wnl.0b013e3182a55f17

Gerhand, S., & Barry, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition, 27*(4), 592-602. https://psycnet.apa.org/doi/10.3758/BF03211553

Gerritsen, M., & Claes, M.T. (2016). *Culturele waarden en communicatie in internationaal perspectief. Bussum: Coutinho.*

Harry, A. & Crowe, S. F. (2014). Is the Boston Naming Test Still Fit For Purpose? *The Clinical Neuropsychologist, vol. 28,* no. 3, pp. 486-504, https://doi.org/10.1080/13854046.2014.892155.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology, 22*(2), 184-203. http://dx.doi.org/10.1080/02687030701262613

Hofs-van Kats, M. (2019). *Severity of Anomia. Severity rating by people with aphasia and experts, using the Dutch Naming Test.* Master thesis, faculty of medicine thesis. Utrecht: Utrecht University.

Ivanov, M.V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: key substantive and psychometric considerations. *Aphas*iology, 27(8), 891-920. https://doi.org/10.1080/02687038.2013.805728.

Ian, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. APA Psych Tests. https://doi.org/10.1037/127208-000.

Kessels, R.P.C., de Vent, N.R., Bruijnen, C.J.W.H., Jansen, M.G., de Jonghe, J.F.M., Dijkstra, B.A.G., & Oosterman, J.M. (2022). Regression-based normative data for het Montreal Cognitive Assessment (MoCA) and its memory index score (MoCA-MIS) for individuals aged 18-91. *Journal of Clinical Medicine, 11*, 4059. https://doi.org/10.3390/jcm11144059.

Macoir, J. & Lavoie, M. (2021). Definitions: Anomia. *Cortex, 144,* pp. 212–212. https://doi.org/10.1016/j.cortex.2021.09.001

Maddy, K.M., Capilouto, G.J., & McComas, K.L. (2014). The effectiveness of semantic feature analysis: an evidence-based systematic review. *Annals of Physical and Rehabilitation Medicine, 57*(4), 254-267. https://doi.org/10.1016/j.rehab.2014.03.002

Mitrushina, M., Boone, K.B., Razani, J., & D'Ella, L.F. (2005). *Handbook of Normative Data for Neuropsychological Assessment*. Second Edition. New York: Oxford University Press.

Orgass, B. (1976). Eine revision des token tests. II: Validitätsnachweis, normierung und standardisierung. *Diagnostica, 22, 141-156.*

Paemeleire, F. (2005). Gebruik van cognitief neuropsychologische taalmodellen bij de diagnostiek van woordvindingsproblemen. *Signaal*, *51*, 4-22.

Paemeleire, F., van Rompaey, L., Verwilligen, E., & Vandenborre, D. (2023). *Vlaamse correctiesleutel voor de Nederlandse Benoem Test (NBT). Uitgave in eigen beheer, verkregen* via www.neurocom.be.

Palmer, R., Hughes, H. & Chater, T. (2017). What Do People with Aphasia Want to Be Able to Say? A Content Analysis of Words Identified as Personally Relevant by People with Aphasia. *PloS One, vol. 12,* no. 3, pp. e0174065–e0174065, https://doi.org/10.1371/journal.pone.0174065

Renzi, E. de, & 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)80006-9*

Roomer, E.K., Brok, S., Hoogerwerf, A.C., & Linn, D.E. (2011). *Literatuuronderzoek naar woordvinding in Nederland.* [handleiding] Kcr Utrecht. https://www.kcrutrecht.nl/wp-content/uploads/2018/09/handleiding-boston-benoemtaak-2011.pdf

Schaerlaekens, A. (2007). *De taalontwikkeling van het kind*. Groningen: Wolters-Noordhoff B.V.

Shao, Z., Roelofs, A., & Meyer, A.S. (2014). Predicting naming latencies for action pictures: Dutch norms. *Behavior Research Methods, 46*(1), 274-283. https://psycnet.apa.org/doi/10.3758/s13428-013-0358-6

Van Ewijk, L., Dijkhuis, L., Hofs-van Kats, M., Hendrickx-Jessurun, M., Wijngaarden, M., & De Hilster, C. (2018a). Handleiding Nederlandse Benoem Test (tweede druk). Houten: Bohn Stafleu van Loghum.

Van Ewijk, L., Dijkhuis, L., Hofs-van Kats, M., Hendrickx-Jessurun, M., Wijngaarden, M., & De Hilster, C. (2018b). Nederlandse Benoem Test: NBT. Houten: Bohn Stafleu van Loghum.

Visch-Brink, E.G., Vandenborre, D., de Smet, H.J., & Mariën, P. (2014). *Comprehensive Aphasia Test-Nederlandse bewerking.* Amsterdam: Pearson bv.

Wilshire, C.E., & Coslett, H.B. (2000). Disorders of word retrieval in aphasia: theories and potential applications. In S.E. Nadeau, L.J.G. Rothi, B. Crosson (Eds.), *Aphasia and language: theory of practice*. New York: Guilford Press.

Zec, R.F., Burkett, N.R., Markwell, S.J., & Larsen, D.L. (2007). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *Clinical Neuropsychologist, 21(4), 578-616. https://doi.org/10.1080/13854040701220028*

Correspondentiegegevens

Dorien Vandenborre
Sint-Andriesstraat 2, 2000 Antwerpen
+ 32 3 241 08 29
dorien.vandenborre@thomasmore.be