



university of
groningen

‘Zie jij wat ik niet zie?’

*De betrouwbaarheid van het Gedragstaxatie-Instrument Revised
(GTI-R) onderzocht*

Nikki van Kempen

Studentnummer: sS3997839
Datum: 30-05-2025
Eerste Beoordelaar: Marleen Wessels
Tweede Beoordelaars: Arjen van Assen
Mastertrack: Orthopedagogiek
Faculteit: Gedrags- en Maatschappijwetenschappen
Woordenaantal: 10869

Abstract

This thesis examines the reliability of the Behavioral Appraisal Scales- revised (BAS-R), an instrument developed to assess the functional abilities of individuals with Profound Intellectual and Multiple Disabilities (PIMD). The goal of this study was 1) to determine whether the BAS-R produces consistent scores when used by different raters (inter-rater reliability) and by the same rate over time (intra-rater reliability) and 2) to explore where differences in scores present themselves and why.

A sequential explanatory design is used in which quantitative data was first collected through completed BAS-R scoring sheets. This was followed by qualitative data derived from raters' notes, a reflective discussion between raters and final written reports.

The findings demonstrate a high reliability: percentage agreement between raters ranged from 80–98%, and between one rater over time from 90–100%. Scoring differences occurred primarily in observational subscales such as visual and motor exploration. Content analysis revealed that these differences often stemmed from interpretational discrepancies, item ambiguities, or contextual factors.

It can be concluded that the BAS-R produces reliable results when used by different or the same raters, but improvements can be made by clarifying certain items, requiring raters' comments, and repetition of the study with a larger group of participants and raters. This would allow for the use of alternative reliability measurements that may offer a more accurate reflection of the instrument's reliability.

Samenvatting

De onderzoek richt zich op de betrouwbaarheid van het Gedragstaxatie-Instrument Revised (GTI-R), een meetinstrument dat is ontworpen om de functionele mogelijkheden van mensen met een (Zeer) Ernstige Verstandelijke en Meervoudige Beperking (ZEVMB) in kaart te brengen. Het doel van het onderzoek was 1) te bepalen wat de betrouwbaarheid van het GTI-R was wanneer het werd gebruikt door verschillende beoordelaars (inter-beoordelaarsbetrouwbaarheid) en door dezelfde beoordelaar over tijd (intra-beoordelaarsbetrouwbaarheid) én 2) ontdekken waar mogelijke verschillen in resultaten vandaan komen.

Het onderzoek is uitgevoerd volgens een verklarend sequentieel design. Eerst is kwantitatieve data verzameld via ingevulde scoretabellen van GTI-R-afnames bij vijf participanten. Vervolgens is kwalitatieve data verzameld via opmerkingen op scoreformulieren, een reflectief gesprek tussen beoordelaars en een analyse van de eindverslagen.

De resultaten tonen aan dat het GTI-R een hoge mate van betrouwbaarheid heeft: de percentage overeenkomst lagen tussen 80–98% en intra-beoordelaarsbetrouwbaarheid tussen 90–100%. Verschillen in scoring kwamen vooral voor bij observatieschalen zoals visueel en motorisch exploratief gedrag. Deze verschillen konden vaak verklaard worden door interpretatieverschillen, onduidelijkheden in de items of contextuele factoren.

De conclusie is dat het GTI-R betrouwbare resultaten oplevert bij afname door verschillende of dezelfde beoordelaars, maar dat er winst te behalen valt in verduidelijking van bepaalde items, scoringsopmerkingen te verplichten en het onderzoek te herhalen met een grotere groep participanten en beoordelaars. Dit maakt het mogelijk om alternatieve maten van betrouwbaarheid te gebruiken die mogelijk een nauwkeuriger beeld geven van de betrouwbaarheid.

Inhoud	
Inleiding	5
Methode	11
Design	11
Steekproef	11
Instrument	13
Procedure van dataverzameling	16
Procedure van filmen en scoren	17
Analyseplan Inter- beoordelaarsbetrouwbaarheid	18
Analyseplan Intra-beoordelaarsbetrouwbaarheid	19
Resultaten	20
Inter-beoordelaarsbetrouwbaarheid	20
Intra-beoordelaarsbetrouwbaarheid	25
Conclusie	28
Discussie	30
Literatuur	34
Bijlage 1	38
Bijlage 2	41
Bijlage 3	43

Inleiding

In wetenschappelijke literatuur, maar ook in beleidsdocumenten wordt in toenemende mate benadrukt hoe belangrijk effectiviteit en efficiëntie van zorg is (Van der Plasschen et al., 2009). Een manier om efficiëntie en effectiviteit te bevorderen is door zorg te bieden die werkt volgens de principes van Evidence Based Practice (EBP). Dit is zorg die werkt en 1) aansluit bij de behoeften van de client, 2) gebaseerd is op wetenschappelijk onderbouwde methoden en 3) wordt ondersteund door de expertise van professionals (Elser et al., 2025).

Het waarborgen van zorg die werkt via deze principes is van belang voor alle doelgroepen in de zorg, maar in het bijzonder voor mensen met een (Zeer) Ernstige Meervoudige Beperking (Z)EVMB). Dit komt doordat mensen met (Z)EVMB veel directe ondersteuning nodig hebben bij alle aspecten van hun leven door een combinatie van een zeer ernstige verstandelijke en ernstige of zeer ernstige motorische beperkingen (Van der Putten et al., 2017). Mensen met een (Z)EVMB kennen vaker gezondheidsproblemen zoals epilepsie, slikproblematiek en ademhalingsproblemen, vanwege ernstige hersenschade of dysfunctie (Van Timmeren et al., 2017b). Ook is er veelal sprake van een visuele en/of auditieve beperking (Van Timmeren, 2019). Een groot deel van deze groep mensen kan bovendien niet communiceren met gesproken taal, maar gebruikt hiervoor geluiden, mimiek (Van der Putten et al., 2017). Ook kunnen fysiologische reacties, zoals versnelde ademhaling communicatieve signalen zijn. Door deze beperkte en subtiele communicatiemogelijkheden zijn mensen met een (Z)EVMB afhankelijk van de directe ondersteuning in het herkennen en signaleren van gezondheidsproblemen (Van Timmeren et al., 2017a). Het is belangrijk aan te merken dat er binnen de groep mensen met een (Z)EVMB verschillende mogelijke combinaties van beperkingen een forse spreiding bestaat in functioneren (Van der Putten et al., 2017).

Volgens van der Putten et al. (2017) speelt de kwaliteit van zorg een belangrijke rol in de ervaren kwaliteit van leven van deze mensen. Gezien de intensieve zorgvraag van mensen met een (Z)EVMB is het van groot belang dat zij zorg ontvangen die daadwerkelijk aansluit bij hun unieke wensen, behoeften en voorkeuren. Hiervoor dient informatie verzameld te worden, ook wel *assessment* genoemd. Assessment is het proces van informatie verzamelen over een persoon en diens situatie om deze persoon goed te kunnen begrijpen (Appl, 2000; Thompson et al., 2018). Het is belangrijk dat er tijdens dit proces goed gekeken wordt naar de omstandigheden en de individuele behoeften van de persoon (Mansell & Beadle-Brown, 2005). De informatie die hieruit volgt kan gebruikt worden om richting te geven aan de ondersteuning (Wessels, 2021).

Uitdagingen voor assessment bij mensen met (Z)EVMB

Voor de doelgroep (Z)EVMB brengt het proces van assessment moeilijkheden met zich mee. Zo is het afnemen van een vragenlijst of een interview vrijwel altijd afhankelijk van een ouder of begeleider die de persoon met een (Z)EVMB goed kent. Dit vanwege de gebrekkige (verbale) communicatievaardigheden van deze persoon zelf (Wessels et al., 2020). Hierdoor is informatie afhankelijk van de context en hoe iets wordt geïnterpreteerd (Wessels et al., 2020). Bovendien is het in de ondersteuning van mensen met een (Z)EVMB vaak zo dat veel verschillende mensen uit diverse disciplines betrokken zijn, die allemaal een eigen perspectief hebben (Vlaskamp et al., 2015). Volgens (Lyons et al., 2016) is het belangrijk dat bij het proces van assessment deze verschillende perspectieven worden meegenomen, maar dit maakt het verzamelen en interpreteren van informatie ingewikkeld. De perspectieven kunnen uiteenlopen en subjectief van aard zijn.

Bij het vinden van de juiste, (ontwikkelingsstimulerende) interventie speelt het proces van assessment een belangrijke rol (Cline, 2018; Visser, 2014). Het is hierbij van belang om instrumenten te gebruiken die geschikt zijn voor deze doelgroep en die inzicht geven in hun specifieke ontwikkelingsmogelijkheden en ondersteuningsbehoeften. Het vinden van dergelijke instrumenten blijkt echter niet eenvoudig. Instrumenten die ontwikkeld zijn voor een andere doelgroep zijn vaak niet geschikt, omdat niet bekend is op wat voor manier mensen met een (Z)EVMB zich ontwikkelen en of dit significant verschilt van andere doelgroepen (Visser et al., 2017). Algemeen gehanteerde normen over de ontwikkeling zijn dan ook niet van toepassing binnen de groep mensen met een (Z)EVMB (Wessels, 2021). Bestaande instrumenten zijn doorgaans gebaseerd op een algemeen ontwikkelingskader, waarin een vaste volgorde van moeilijkheid wordt gebruikt. Veel van deze instrumenten maken gebruik van zogenoemde afbreekregels: wanneer een bepaald aantal opeenvolgende items niet wordt behaald, stopt de test. Het is echter niet bekend of dit ook passend is voor mensen met een (Z)EVMB, vanwege hun mogelijk atypische ontwikkelingspatroon (Wessels, 2021). Het is bijvoorbeeld denkbaar dat iemand een 'moeilijk' item wél beheerst, terwijl eenvoudigere items niet worden gehaald.

Bovendien kent deze doelgroep een enorme complexiteit van beperkingen, wat maakt dat items uit bestaande instrumenten vaak niet toepasbaar zijn. Items kunnen bepaalde motorische of visuele vaardigheden veronderstellen, die door deze complexiteit lang niet altijd haalbaar zijn voor deze doelgroep. Een voorbeeld hiervan is het meten van communicatie door het maken van oogcontact. Voor mensen die een ernstige visuele beperking hebben is dit (te) lastig, waardoor zij lager zullen scoren op dit item. Dit terwijl blind zijn helemaal niet hoeft te betekenen dat de communicatievaardigheden minder zijn, maar mogelijk wordt contact op een andere manier gelegd, zoals door de ander aan te raken (Wessels, 2021).

Naast het beperkte aanbod van instrumenten die geschikt zijn voor deze doelgroep, ontbreekt het bij de bestaande instrumenten vaak aan onderzoek naar de psychometrische eigenschappen (Carnaby, 2007). Dit terwijl psychometrische onderbouwing wel van belang is om in adequate assessment te kunnen voorzien (Wessels, 2021).

Samengevat kent het proces van assessment bij mensen met een (Z)EVMB verschillende uitdagingen. Door de beperkte communicatiemogelijkheden van een persoon met (Z)EVMB is informatie vaak afhankelijk van de context en hoe iets wordt geïnterpreteerd (Wessels et al., 2020). Daarbij is het aanbod van instrumenten dat specifiek is ontwikkeld voor deze doelgroep beperkt. Bestaande instrumenten sluiten vaak onvoldoende aan vanwege een niet-passend theoretisch kader, een atypisch ontwikkelingsverloop en de complexiteit van beperkingen (Visser et al., 2017, Wessels, 2021). Van de instrumenten die wél geschikt zijn voor deze specifieke doelgroep ontbreekt in veel gevallen onderzoek naar de psychometrische kwaliteit (Carnaby, 2007). Toch is adequate assessment juist bij deze groep van essentieel belang, aangezien alleen passende en betekenisvolle ondersteuning geboden kan worden wanneer er inzicht is in de wensen, behoeften en mogelijkheden van de persoon met een (Z)EVMB (Wessels, 2021).

Het GTI

Eén van de weinige instrumenten die zowel specifiek is ontwikkeld voor de doelgroep (Z)EVMB als onderzocht is op psychometrische kwaliteit, is het Gedrags Taxatie-Instrument (GTI). Dit instrument is ontworpen om de functionele mogelijkheden van mensen met een (Z)EVMB in kaart te brengen (Wessels, 2024). Het GTI onderscheidt zich van andere assessmentinstrumenten doordat het is afgestemd op complexe ondersteuningsbehoeften van de doelgroep en flexibel kan worden afgenomen. De eerste versie van dit instrument dateert uit 1999 en heeft als doel het in kaart brengen van de functionele mogelijkheden van kinderen en volwassenen met een (Z)EVMB (Vlaskamp et al., 1999). Met ‘functionele mogelijkheden’ wordt de mogelijkheid om alledaagse vaardigheden uit te voeren, bedoeld (Vlaskamp et al., 2005). Door de functionele mogelijkheden van een persoon in kaart te brengen biedt het GTI aanknopingspunten voor het vormgeven van begeleiding waarin zowel rekening wordt gehouden met de functionele mogelijkheden als de beperkingen van een persoon (Vlaskamp et al., 1999).

Onderzoek naar de psychometrische kwaliteit van het GTI

Er is beperkt onderzoek gedaan naar de psychometrische kwaliteit van het oorspronkelijke GTI. De constructvaliditeit en inter-beoordelaarsbetrouwbaarheid zijn onderzocht door Vlaskamp, van der Meulen & Zijlstra (2002). De convergente validiteit van het instrument, bestaande uit 122 items, is onderzocht door de correlaties per subschaal te vergelijken met die van een ander vergelijkbaar instrument. Voor de inter-beoordelaarsbetrouwbaarheid zijn 44 afnames van het GTI onafhankelijk gescoord door twee beoordelaars. In dit onderzoek werd een acceptabele constructvaliditeit en goede betrouwbaarheidscoëfficiënten gevonden. De auteurs benadrukten het belang van verder onderzoek naar constructvaliditeit (Vlaskamp et al., 2002). In een vervolgstudie onderzochten Wessels, Paap en Van der Putten (2020) de factorstructuur, itembias en convergente validiteit van het GTI. Hieruit bleek dat het merendeel van de items goed binnen de beoogde subschalen viel, wat de factorstructuur ondersteunt. Wel correleerde 16–18% van de items sterker met een andere subschaal, wat wijst op ruimte voor verbetering in de schaalindeling. Verder werd er een acceptabele convergente validiteit gevonden, maar werd er ook genoemd dat wanneer bepaalde items verwijderd of veranderd worden, deze mate van samenhang vergroot kan worden (Wessels et al., 2020). Tot slot is er in een recentere Delphi studie onderzoek gedaan naar de inhoudsvaliditeit van het GTI (Wessels et al., 2022). Met inhoudsvaliditeit wordt de mate waarin een meetinstrument representatief is voor het meten van alle aspecten van een bepaald construct bedoeld (Wessels et al., 2022). Het doel van de studie was het evalueren van de inhoudsvaliditeit, maar ook het waar mogelijk verbeteren ervan. In deze studie gaven ouders van mensen met een (Z)EVMB, zorgprofessionals werkzaam in deze sector en wetenschappers die zich richten op deze doelgroep, feedback over de items van het GTI. Een punt van feedback was bijvoorbeeld dat een aantal items niet van toepassing waren voor mensen met een (Z)EVMB, omdat deze afhankelijk waren van andere vaardigheden, bijvoorbeeld op visueel of motorisch vlak. Deze en meer aspecten zijn vervolgens verwerkt in het, volgens betrokken experts, meer representatieve GedragTaxatie-Instrument-Revised. Vervolgens is tussen het GTI en het GTI-R de inhoudsvaliditeit vergeleken, waaruit is gebleken dat dat deze hoger was bij de aangepaste tweede versie, het GTI-R.

Het GTI-R

Het GTI-R heeft evenals het GTI het doel om de functionele mogelijkheden van kinderen en volwassenen met een (Z)EVMB in kaart te brengen. Dit instrument is als een van de weinige wél onderzocht op psychometrische kwaliteit en is specifiek ontwikkeld voor de doelgroep (Z)EVMB. De test bestaat uit zes subschalen met een bijpassend aantal items. Ieder item beschrijft een bepaalde vaardigheid (Wessels et al., 2023). Een afname van het GTI-R is flexibel, waardoor rekening wordt gehouden met de complexiteit van de beperking van mensen met een (Z)EVMB. Zo ligt de volgorde

van de items niet vast. Ook ligt het aantal keer dat een object mag worden aangeboden niet vast. Bij het GTI-R zit een tas met standaard materiaal, maar hier mag van worden afgeweken als het kind of de volwassene voorkeur heeft voor eigen materiaal heeft (Wessels et al., 2023). Dit kan bijvoorbeeld een bepaalde knuffel zijn waar iemand waarde aan hecht, maar ook een ander voorwerp dat de persoon sensorisch gezien als prettig ervaart. Gebaseerd op de afname wordt een eindverslag geschreven met daarin de bevindingen per subschaal en eventuele aanbevelingen voor de begeleiding. Een dergelijk eindverslag kan vervolgens gebruikt worden om de ondersteuning op af te stemmen. Deze verslagen vormen de brug tussen testresultaten en daadwerkelijke ondersteuning in de praktijk.

Probleemstelling

Het GTI-R is gebaseerd op de input van de experts tijdens het Delphi onderzoek, maar naast dit onderzoek heeft geen onderzoek naar de psychometrische kwaliteit van het instrument plaatsgevonden. Vanwege vele wijzigingen van het GTI naar het GTI-R, is verder onderzoek naar de psychometrische kwaliteit van deze nieuwe versie wenselijk. Een belangrijke wijziging is de toevoeging van open vragen, in tegenstelling tot de voornamelijk dichotome scores van het oorspronkelijke GTI. Zo wordt nu bijvoorbeeld gevraagd op welke manier iemand een bepaalde vaardigheid laat zien. Dit kan aanvullende informatie opleveren over bijvoorbeeld de frequentie van gedragingen en of de persoon met een (Z)EVMB bepaald gedrag buiten de testsetting juist wel of normaliter nooit laat zien (Wessels, 2023). Een voorbeeld hiervan is dat de informant vertelt dat de persoon met (Z)EVMB normaal gesproken wel zijn/haar hoofd draait om een object te lokaliseren, maar dat nu door de aanwezigheid van de (onbekende) testleider niet lijkt te doen. Ook wordt met open vragen in kaart gebracht hoe iemand zich uit, bijvoorbeeld hoe je kunt zien dat de persoon blij of boos is. Deze kwalitatieve informatie draagt bij aan een rijker en genuanceerder beeld van de persoon met een (Z)EVMB. Tegelijkertijd maakt dit de objectieve scoring lastiger: de antwoorden zijn immers gebaseerd op de ervaring en interpretatie van de informant. Deze subjectiviteit vormt een mogelijke bedreiging voor de betrouwbaarheid van het GTI-R, omdat scores sterk afhankelijk kunnen zijn van degene die de antwoorden geeft en ook invult. Daarnaast bestaat het GTI-R voor een groot deel uit observaties. Ook dit heeft gevolgen voor de betrouwbaarheid van de resultaten (Brentnall & Bundy, 2009.) Observaties zijn namelijk gevoelig voor contextuele factoren zoals locatie, tijdstip, de fysieke toestand of positionering van de persoon (Maes et al., 2021). Er is dan ook een afweging nodig tussen enerzijds procedurele betrouwbaarheid en anderzijds het bieden van ruimte voor aanpassing aan de individuele kenmerken van de persoon met een (Z)EVMB.

Aangezien het eindverslag mede gebaseerd is op deze scores en observaties, is het relevant om niet alleen te onderzoeken in hoeverre er verschillen zijn in scoring tussen of binnen

beoordelaars, maar ook in hoeverre deze verschillen mogelijk doorwerken in het totaalbeeld dat in het verslag wordt geschetst. Dit is relevant, aangezien dergelijke eindverslagen in de praktijk een belangrijke rol spelen bij het vormgeven van de ondersteuning, en daarmee mede bepalen of verschillen tussen beoordelaars of tijdstippen betekenisvolle gevolgen kunnen hebben.

Samenvattend kan gesteld worden dat doordat het GTI-R meer gebruik maakt van open informatie dan het oorspronkelijke GTI, het belangrijk is om te kijken naar de (veranderde) betrouwbaarheid van het instrument. Dit levert tot dusverre ontbrekende informatie op waar in dit onderzoek de focus op ligt. Er wordt op een andere manier gekeken naar de mate van betrouwbaarheid dan destijds in het onderzoek van Vlaskamp, van der Meulen & Zijlstra (2002) voor het oorspronkelijke GTI. In dit artikel uit 2002 is met een grote steekproef gekeken naar óf er sprake is van verschillen in resultaten als verschillende personen dezelfde meting verrichten (inter-beoordelaarsbetrouwbaarheid). In dit onderzoek wordt gebruik gemaakt van een kleinere steekproef en ligt de focus meer op waar de mogelijke verschillen in resultaten liggen en mogelijke verklaringen hiervoor. Naast inter-beoordelaarsbetrouwbaarheid is er gekeken naar of er sprake is van verschillen in resultaten als dezelfde persoon dezelfde meting verricht over tijd (intra-beoordelaarsbetrouwbaarheid). Aangezien een afname van het GTI-R niet alleen resulteert in scores, maar tevens leidt tot een eindverslag waarin de sterke en zwakke kanten van een persoon met (Z)EVMB worden geïnterpreteerd en vertaald naar aanbevelingen, is het van belang te onderzoeken of eventuele verschillen in scoring ook daadwerkelijk leiden tot een ander totaalbeeld tussen de verslagen van twee beoordelaars.

De onderzoeksvragen die in dit onderzoek centraal staan zijn als volgt:

- 1) In hoeverre is er sprake van dezelfde resultaten bij een afname van het GTI-R als verschillende personen dezelfde meting verrichten? (Inter-beoordelaarsbetrouwbaarheid)
 - a) In welke mate worden de scoretabellen van het GTI-R hetzelfde ingevuld door twee beoordelaars?
 - b) Waar zitten de mogelijke verschillen in scoring en waarom?
 - c) In hoeverre komen de verslagen op basis van GTI-R afnames overeen tussen twee beoordelaars?
- 2) In hoeverre is er sprake van overeenkomst tussen de uitkomsten van een GTI-R afname wanneer dezelfde onderzoeker dezelfde afname, dus exact hetzelfde fragment, binnen vier weken nogmaals scoort? (intra-beoordelaarsbetrouwbaarheid)
 - a) In welke mate worden de GTI-R scoretabellen hetzelfde ingevuld door dezelfde beoordelaar over tijd?
 - b) Waar zitten de mogelijke verschillen in scoring en waarom?

Methode

Design

Om de onderzoeksvragen van dit onderzoek te beantwoorden is gewerkt via een verklarend sequentieel design. Bij een dergelijk design vindt de kwantitatieve dataverzameling plaats voordat wordt begonnen met de kwalitatieve analyse (George, 2024). Het idee hierachter is dat de kwalitatieve analyse helpt om de kwantitatieve data te verklaren. Er is voor dit design gekozen, omdat het aansluit bij de aard van het instrument en de doelgroep. De kwantitatieve analyse geeft inzicht in de mate van overeenstemming tussen beoordelaars, maar biedt nog geen verklaring voor eventuele verschillen in scoring. Omdat het GTI-R deels bestaat uit observaties en interpretaties van open informatie, is het belangrijk om niet alleen te weten *óf* er verschillen bestaan, maar ook *waarom*. Door na de kwantitatieve fase over te gaan op een kwalitatieve analyse, kunnen deze verschillen in context worden geplaatst (Fetters, 2013). Voor dit onderzoek zijn GT-R afnames gedaan en gescoord door twee beoordelaars en door één beoordelaar op twee momenten, waarbij de mate van overeenstemming is geanalyseerd en verklaringen voor verschillen zijn onderzocht.

In dit onderzoek bestaat de kwantitatieve data uit de binaire variabele 'ja' 'nee' per item op de ingevulde scoreformulieren. De kwalitatieve data bestaat uit de eventuele opmerkingen die bij de items op het scoreformulier zijn geschreven, de eindverslagen die zijn opgesteld en een reflectief gesprek tussen de beoordelaars.

Steekproef

Participanten zijn geschikt voor dit onderzoek wanneer zij voldoen aan de volgende kenmerken:

- *(Zeer)ernstige verstandelijke beperking*. Individuen met zulke ernstige verstandelijke beperkingen dat er geen bestaande gestandaardiseerde tests toepasbaar zijn voor een geldige schatting van hun niveau van verstandelijke capaciteit (Nakken & Vlaskamp, 2007). Hiervoor wordt in dit onderzoek voor volwassenen een ontwikkelingsleeftijd tussen de 0 en 24 maanden aangehouden (Maes et al., 2021). Voor kinderen geldt dat wanneer hun functioneren (meer dan) de helft lager wordt geschat dan hun leeftijdsgenoten ze in aanmerking komen (Vig & Sanders, 2007).
- *Zeer ernstige of ernstige motorische beperking*: Volgens het Gross Motor Function Classification System (GMFCS) Niveau IV (beperkingen in de eigen mobiliteit, dus bijvoorbeeld ondersteuning nodig hebben om te kunnen zitten of staan) of niveau V (geen enkele vorm van zelfstandige mobiliteit en weinig controle in het bewegen (Palisano et al., 1997)).
- Er is toestemming verkregen voor dit onderzoek van een wettelijk vertegenwoordiger

Voor de participanten geldt geen leeftijdsgrens. De participanten zijn woonachtig in zorginstellingen in Nederland die ondersteuning bieden aan mensen met een (Z)EVMB. De twee beoordelaars zijn beide werkzaam binnen de zorg voor mensen met een beperking en hebben ervaring met deze doelgroep.

Vanwege de tijdinvestering die een afname van het GTI-R en het schrijven van het afsluitend verslag per afname in beslag neemt, is gekozen voor een steekproef van vijf participanten. Hoewel dit een relatief klein aantal is, wordt dit als voldoende geacht om een beeld te vormen van waar en waarom er mogelijke verschillen in scoring bestaan tussen twee beoordelaars en tussen één beoordelaar over tijd, binnen de beperkte tijd die beschikbaar is voor dit onderzoek. Omdat het onderzoek een verklarend en verkennend karakter heeft, ligt de nadruk niet op generaliseerbaarheid, maar op het verkrijgen van inzicht in de mogelijke variatie in interpretatie tussen twee beoordelaars. Malterud et al. (2016) stelt dat informatiewaarde van de data belangrijker is dan de omvang van de steekproef. De gekozen deelnemers vormen een representatieve afspiegeling van de doelgroep (Z)EVMB. Zo is er sprake van verschillende geslachten, leeftijden, beperkingen (te zien in Tabel 1) en voldoen ze allen aan de eerder genoemde kenmerken.

Tabel 1

Overzicht kenmerken participanten steekproef 1

Nr.	Leeftijd	Geslacht	GMFCS- score	Visuele beperking	Auditieve beperking	Epilepsie
1	16	Vrouw	= 5 Zit altijd in een rolstoel, kan deze niet zelfstandig voortbewegen	Ja. Ernstige visuele beperking.. Heeft ook Myokymie (wiebelende oogleden). Ziet mogelijk beter aan de linkerkant dan aan de rechterkant.	Nee. Er zijn geen aanwijzingen geweest om dit te onderzoeken	Ja. In beide hersenhelften is er sprake van epileptische activiteit.
2	21	Man	= 5 Zit altijd in een rolstoel, kan deze niet zelfstandig voortbewegen	Ja. Ernstige visuele beperking.	Nee. Er zijn geen aanwijzingen geweest om dit te onderzoeken.	Ja. Maar geen recente aanvallen meer gehad.
3	44	Man	= 5 Zit altijd in een rolstoel, kan deze niet zelfstandig voortbewegen	Nee. Er zijn geen aanwijzingen geweest om dit te onderzoeken.	Nee. Er zijn geen aanwijzingen geweest om dit te onderzoeken.	Nee

Nr.	Leeftijd	Geslacht	GMFCS- score	Visuele beperking	Auditieve beperking	Epilepsie
4	46	Man	= 4 (Zit niet altijd in een rolstoel, kan zich enigszins voortbewegen door over de grond te schuiven).	Ja. Ziet beter van dichtbij dan verder weg.	Ja. Is onderzocht en slechthorend verklaard.	Nee
5	60	Vrouw	= 5 (Zit altijd in een rolstoel, kan deze niet zelfstandig voortbewegen).	Ja. Ernstige visuele beperking kokervisus. Ziet rechts beter dan links.	Nee. Er zijn geen aanwijzingen geweest om dit te onderzoeken	Nee

Noot: Naast deze kenmerken hebben alle participanten een (zeer) ernstige verstandelijke beperking

Instrument

Bij het GTI-R zit een handleiding, een scoretabel en een rugzak met materiaal. Voorafgaand aan de afname dient er informatie over de participant verzameld te worden om een zo reëel mogelijk beeld van iemands functioneren te krijgen. Er worden hiervoor vragen gesteld aan een informant (iemand die de participant goed kent), maar informatie mag ook uit het dossier gehaald worden. Onderwerpen voor deze vragen zijn te vinden in de scoretabel. Voorbeeldvragen zijn: 'Welke voorkeursobjecten heeft iemand die gebruikt kunnen worden tijdens de afname? Is er sprake van mogelijke invloed van epilepsie op de afname? Het is belangrijk dat wordt vermeld hoe aan de informatie gekomen is.

Deze informatie voor het GTI-R wordt op verschillende manieren verzameld, afhankelijk van de subschaal. In Tabel 2 (zie pagina 15) zijn de genoemde subschalen met omschrijving en het aantal bijbehorende items van het GTI-R te vinden. De eerste drie subschalen, emotioneel communicatief gedrag, receptief taalgedrag en algemeen communicatief gedrag, zijn grotendeels gebaseerd op informatie die via uitvragen bij een informant wordt verkregen. De laatste drie subschalen, visueel exploratief gedrag, motorisch exploratief gedrag en auditieve exploratie, zijn daarentegen gebaseerd op observatie. Deze observatie bestaat uit zowel het spontaan observeren van gedrag in de vrije ruimte als het doelgericht uitlokken van gedrag met behulp van specifiek materiaal. Op deze manier wordt geprobeerd een zo volledig mogelijk beeld van het functioneren te verkrijgen, waarbij ook rekening wordt gehouden met contextuele invloeden en interpretatieverschillen.

Op het scoreblad staat per item een bepaald criteriumgedrag, zoals bijvoorbeeld 'een object volgen met de ogen' zo precies mogelijk omschreven die de participant moet laten zien. Ook staan er handelingen genoteerd die testleider moet uitvoeren, evenals de

kenmerken van de te gebruiken materialen. Indien het criteriumgedrag zich voordoet wordt een vinkje in de kolom 'ja' geplaatst in de scoretabel. Indien dit niet het geval is, wordt een vinkje geplaatst in de kolom 'nee'. In Tabel 3 is een voorbeeld van een item uit de scoretabel te zien. Vervolgens kunnen de vakjes waar 'ja' gescoord is worden opgeteld per subschaal om zo een score per subschaal te berekenen. Deze totaalscore geeft het vaardigheidsniveau van een persoon aan op een subschaal, waarbij een hogere score een hoger vaardigheidsniveau betekent. Er zijn geen normscores. Door het scoren wordt duidelijk welke vaardigheden iemand beheerst en welke vaardigheden nog niet (Wessels et al., 2023). Naast deze dichotome scoring (ja/nee) maakt het GTI-R ook gebruik van open vragen. Dit geeft aanvullende informatie, bijvoorbeeld over hoe bepaald gedrag eruitziet bij de persoon en hoe emoties kunnen worden herkend. Met deze informatie uit de open vragen en de dichotome scores wordt een verslag geschreven met daarin de bevindingen van het onderzoek. Ook wordt er een aantal aanbevelingen voor de begeleiding van de participant genoemd. Dit verslag dient als het eindproduct van een GTI-R afname.

Het GTI-R is gebaseerd op input van experts tijdens een Delphi onderzoek van Wessels et al. (2022), maar naast dit onderzoek heeft geen onderzoek naar de psychometrische kwaliteit van het instrument plaatsgevonden. Naar het oorspronkelijke GTI is wel onderzoek gedaan. Hier zijn onder andere een acceptabele constructvaliditeit en goede betrouwbaarheidscoëfficiënten gevonden (Vlaskamp et al., 2002).

Tabel 2*Subschalen met omschrijving van het GTI-R*

Subschaal	Omschrijving	Aantal items
Emotioneel communicatief gedrag	Deze subschaal bevat items die aangeven hoe een persoon met (Z)EVMB gevoelens uitdrukt	6
Receptief taalgedrag	In deze subschaal komen reactieve taalvaardigheden aan de orde. Het gaat om de manier waarop het kind/de volwassene reageert op de communicatieve bedoelingen van een ander. Aan de orde is vooral het tonen van begrip van communicatieve bedoelingen van een ander en het begrip van symbolische waarde van gebaren, activiteiten en afbeeldingen.	8
Algemeen communicatief gedrag	In deze subschaal gaat het om de mogelijkheden van het kind of de volwassene om zich te kunnen uiten in de sociale omgeving. Omdat sommige personen met (Z)EVMB met name non-verbaal communiceren (bijvoorbeeld met mimiek, gebaren) en anderen verbaal (geluiden, woorden), is deze subschaal opgedeeld in non-verbale communicatie en verbale communicatie.	19 (9 non-verbaal en 10 verbaal)
Visueel (exploratief) gedrag	Het gaat in deze subschaal om de reactie op visuele prikkels en het visueel onderzoeken van de omgeving/objecten	17
Motorisch exploratief gedrag	Het gaat in deze subschaal om het onderzoeken van de omgeving en objecten door het grijpen, reiken, voelen en manipuleren van objecten. Het type object (voorkeursobjecten, grootte, kleur) kan worden aangepast aan de mogelijkheden van het kind of de volwassene, waardoor hij of zij maximaal kan scoren.	22
Auditieve exploratie	In deze subschaal gaat het om reactie op geluiden.	7
Totaal aantal items		79

Noot. Overgenomen uit Handleiding GTI-R, door M.D Wessels et al., p. 15. Copyright 2023 door Academische Werkplaats EMB

Tabel 3

Voorbeeld uit scoretabel van scoring item behorende bij de subschaal Visueel exploratief gedrag.

Itemnr.	Item	Instructie	Ja	Nee
4	De persoon blijft kijken naar een object wanneer het dichterbij en verder weg bewogen wordt	Een object wordt in het centrum van het gezichtsveld aangeboden. Wanneer de persoon naar het object kijkt, wordt het drie keer langzaam dichterbij en verder weg van de persoon bewogen. Sommige personen zien juist in het midden van hun gezichtsveld niet goed, het object mag in dat geval ook op een andere plek (links of rechts) worden aangeboden		

Opmerkingen

Procedure van dataverzameling

De Ethische Commissie van de faculteit GMW van de Rijksuniversiteit Groningen heeft toestemming voor dit onderzoek verleend. Zodra deze toestemming werd verleend is er begonnen met het werven van participanten. Om participanten voor een afname te werven, is in het eigen netwerk van de onderzoekers gezocht naar personen met (Z)EVMB die voldoen aan de eerder genoemde inclusiecriteria (zie pagina 11). Er zijn verschillende zorgorganisaties voor mensen met een verstandelijke beperking benaderd. Zij ontvingen een informatiebrief met informatie over dit onderzoek en daarbij de vraag of de organisatie mee zou willen werken. Als het antwoord hierop 'ja' was zijn gedragswetenschappers binnen deze organisaties op zoek gegaan naar participanten die voldoen aan de vereiste kenmerken voor dit onderzoek. Voor die participanten is vervolgens contact gezocht met de wettelijk vertegenwoordigers. Zij ontvingen een informatiebrief waarin het doel van het onderzoek beschreven staat en een toestemmingsformulier die zij bij akkoord voor deelname konden tekenen. Deze informatiebrief is te vinden in Bijlage 1 en het toestemmingsformulier in Bijlage 2. De uiteindelijke afnames zijn gedaan door een testleider (beoordelaar 1) en een observant (beoordelaar 2). Gedurende de vijf afnames is deze verdeling hetzelfde gebleven om zo consistent mogelijk te blijven. Dit is belangrijk in onderzoek naar de betrouwbaarheid om het voorkomt dat verschillen in scoring worden veroorzaakt door een variatie in rolverdeling (Avrumson, 2025). De testleider (beoordelaar 1) was verantwoordelijk voor het actief uitvoeren van de afname, het aanreiken van materialen en het uitlokken van gedrag. De observant bediende de video- camera en

ondersteunde waar nodig gedurende de afname. Naast de twee beoordelaars was per afname een participant en een betrokken begeleider aanwezig. Per afname zijn de eerste drie subschalen uitgevraagd aan deze betrokken begeleider. De laatste drie subschalen zijn gescoord via observaties gedurende de afname of met video-opnames achteraf. Alle afnames vonden plaats in de woning van de participant. De periode van dataverzameling betreft een periode van drie maanden tussen Januari en Maart 2025.

Procedure van filmen en scoren

Beide beoordelaars hebben de GTI-R scoretabellen per participant ingevuld en hebben per item in de scoretabel 'ja' ingevuld als het criteriumgedrag zich voordeed en 'nee' ingevuld als dit niet het geval is. Als het item niet aan bod is gekomen gedurende de afname konden de beoordelaars er ook voor kiezen om niets in te vullen. Naast deze 'ja' en 'nee' of het leeglaten hebben de beoordelaars ook bij zoveel mogelijk van de items een opmerking genoteerd. Deze opmerkingen vormden de antwoorden op de open vragen en dienden als toelichting op de gegeven score. De eerste drie subschalen, die uitgevraagd worden aan iemand die de persoon met (Z)EVMB goed kent, zijn tijdens de afname gescoord en de andere subschalen, die door observatie worden ingevuld, zijn gescoord op basis van de video opnames. De videobeelden zijn beveiligd gedeeld via een Y-schijf van de RUG. Dit is een gedeelde netwerkschijf om onderzoeksgegevens (intern) op te slaan (University Workplace Storage, 2025). Tijdens het beoordelen mochten de beoordelaars de opname op pauze zetten en stukken terugspoelen. De twee beoordelaars mochten hun observaties tijdens de afname niet met elkaar bespreken omdat dit de resultaten mogelijk beïnvloedt (Evers et al., 2010). De video-opnames maakten het ook mogelijk om een afname een tweede keer te scoren zonder dat de persoon met (Z)EVMB en de informant nog een keer aanwezig hoefden te zijn. Hiervoor is gekozen vanwege de belasting van de participant en informant. Het maken van video opnames was alleen mogelijk voor de subschalen die op basis van observatie worden gescoord, namelijk: *Visueel exploratief gedrag*, *Auditief exploratief gedrag* en *Motorisch exploratief gedrag*. Voor de intra-beoordelaarsbetrouwbaarheid is door beoordelaar 2 dezelfde afname, via video-opnames na een periode van vier weken nogmaals gescoord.

Nadat beide beoordelaars de GTI-R scoretabellen in zijn volledigheid hebben ingevuld, is er een reflectieve discussie geweest tussen de twee beoordelaars. Deze discussie richtte zich op de items die verschillen in scoring lieten zien.

Tot slot hebben beide beoordelaars voor elke afname een eindverslag geschreven met daarin de bevindingen van de afname en hun adviezen.

Analyseplan Inter- beoordelaarsbetrouwbaarheid

Om de data te analyseren is er als eerste met kruistabellen gekeken of er sprake was van verschillen in de ingevulde scoretabellen als beoordelaar 1 en beoordelaar 2 dezelfde meting verrichten (inter-beoordelaarsbetrouwbaarheid). Er is daarvoor gekeken naar verschillen in de dichotome scoring 'ja' of 'nee' tussen de twee beoordelaars. Met deze gegevens uit de scoretabellen is het percentage van overeenkomst berekend per subschaal. Het percentage van overeenkomst is berekend door voor elk item het totale aantal van overeenkomsten te delen door het totale aantal onenigheden (Vlaskamp et al., 2002). Volgens McHugh (2012) is deze methode geschikt als eerste stap in het vaststellen van de mate van overeenstemming, vooral in situaties met binaire scores en een kleine steekproefomvang. Een percentage van 80% of hoger wordt doorgaans beschouwd als een acceptabele mate van overeenstemming en is ook voor dit onderzoek aangehouden (McHugh, 2012).

Als volgende stap is gekeken naar mogelijke verklaringen voor de verschillen in beoordeling tussen twee beoordelaars en tussen één beoordelaar over tijd, door middel van een kwalitatieve inhoudsanalyse. Hiervoor zijn de opmerkingen in de scoretabellen geanalyseerd. Met deze opmerkingen hebben beide beoordelaars gepoogd om een 'ja' of 'nee' scoring te onderbouwen en/of antwoord te geven op een open vraag. Om deze data te kunnen interpreteren is een codeboek opgesteld. Daarmee kunnen de oorzaken van de verschillen tussen twee beoordelaars geclassificeerd worden. Er is hiervoor alleen gebruik gemaakt van inductieve codes. Het volledige codeboek met toelichtingen en voorbeelden is te vinden in Bijlage 3.

Om beter zicht te krijgen op de overwegingen achter de gegeven scores, zijn de beoordelaars met elkaar in gesprek gegaan over alle gevonden verschillen in scoring tussen de twee beoordelaars. Per verschillend gescoord item is besproken waarom een bepaalde score is toegekend. Deze gesprekken hielpen om beter te begrijpen hoe verschillen tot stand kwamen, en boden de mogelijkheid om de gemaakte coderingen te controleren, aan te vullen of aan te scherpen. Dit proces kan worden gezien als een vorm van *team-based reflexivity*, waarbij gezamenlijk reflecteren op interpretaties en besluitvorming bijdraagt aan de kwaliteit van de analyse (Rankl et al., 2021). Zoals Vandesande et al. (2020) beschrijven, kan het bespreken van deze observaties waardevol zijn. Het kan onder andere helpen om te kijken of de beoordelaars hetzelfde gedrag hebben waargenomen. Het gesprek tussen de twee beoordelaars is opgenomen en vervolgens getranscribeerd. Het opnemen en transcriberen van dit gesprek maakte het mogelijk om een aantal thema's te identificeren die inzicht geven in de aard en oorzaak van de verschillen in scoring. Deze thema's zijn te vinden in de resultaten sectie van dit onderzoek.

Tot slot is er aan de hand van kwalitatieve descriptie (Sandelowski, 2000), onderzocht of er naast de verschillen in de ingevulde scoretabellen ook verschillen zichtbaar zijn in het eindproduct

van de afname: het eindverslag. Het gaat er bij een kwalitatieve descriptie om dat onderzoekers dicht bij de data blijven en een rijke, feitelijke beschrijving hiervan geven (Sandelowski, 2000). Voor dit onderzoek is per participant gekeken in hoeverre beide beoordelaars dezelfde sterke en zwakke(re) punten van een participant in hun eindverslag hebben genoemd. De verschillen en overeenkomsten zijn te vinden in de resultatensectie.

Analyseplan Intra-beoordelaarsbetrouwbaarheid

Het analyseplan voor de intra-beoordelaarsbetrouwbaarheid volgt grotendeels dezelfde structuur als die voor de inter-beoordelaarsbetrouwbaarheid, maar kent enkele belangrijke verschillen. Zo zijn er minder stappen uitgevoerd en is de analyse beperkt tot de drie subschalen die op basis van observatie worden gescoord: *Visueel exploratief gedrag*, *Auditief exploratief gedrag* en *Motorisch exploratief gedrag*. Alleen deze subschalen zijn opgenomen op video en konden daardoor een tweede keer door dezelfde beoordelaar worden gescoord. Om de intra-beoordelaarsbetrouwbaarheid te bepalen, is door middel van kruistabellen gekeken of er verschillen zijn tussen de scores die beoordelaar 2 over tijd (met maximaal vier weken ertussen) heeft ingevuld voor dezelfde afname. Hierbij is gekeken naar verschillen in de dichotome scoring ('ja' of 'nee') tussen het eerste en tweede scoringsmoment. Vervolgens is, net als bij de inter-beoordelaarsbetrouwbaarheid, per subschaal het percentage van overeenkomst berekend. Een overeenkomst van 80% of hoger wordt beschouwd als acceptabel en is ook hier als grenswaarde aangehouden (McHugh, 2012).

Daarnaast is, net als bij de inter-beoordelaarsbetrouwbaarheid, een kwalitatieve inhoudsanalyse uitgevoerd. Hiervoor zijn de opmerkingen in de scoretabellen geanalyseerd. Deze opmerkingen bevatten zowel onderbouwingen van de gegeven scores als beantwoording van de open vragen. Hier is gebruik gemaakt van dezelfde inductieve codering als bij de inter-beoordelaarsbetrouwbaarheid. Het volledige codeboek, inclusief toelichtingen en voorbeelden, is opgenomen in Bijlage 3.

In tegenstelling tot de inter-beoordelaarsbetrouwbaarheid is er bij de intra-beoordelaarsbetrouwbaarheid geen reflectief gesprek gevoerd, aangezien het hier gaat om de beoordeling door één persoon over tijd.

Resultaten

Inter-beoordelaarsbetrouwbaarheid

Om te onderzoeken of de scoretabellen van het GTI-R hetzelfde ingevuld worden door twee beoordelaars, zijn er per participant kruistabellen gemaakt en is het percentage overeenkomst per subschaal berekend. In Tabel 4 is het percentage overeenkomst per subschaal te zien.

Tabel 4

Percentage van overeenkomst in de scoretabellen per subschaal

Subschaal	Inter-beoordelaars
Emotioneel communicatief gedrag	96,6%
Receptief taalgedrag	87,5%
Algemeen communicatief gedrag (non-verbaal)	96,0%
Algemeen communicatief gedrag (verbaal)	98,0%
Visueel exploratief gedrag	80,0%
Motorisch exploratief gedrag	84,5%
Auditieve exploratie	88,5%

Tabel 4 laat het percentage overeenkomst per subschaal zien tussen de scores van beoordelaar 1 en 2 zien. De inter-beoordelaarsovereenkomsten variëren tussen de 80,0% en 98,0%, waarbij de hoogste overeenstemming is gevonden bij *Algemeen communicatief gedrag (verbaal)* en de laagste bij *Visueel exploratief gedrag*. Alle gemeten percentages van overeenkomst liggen boven de grens van 80%, wat binnen dit onderzoek als acceptabel wordt beschouwd.

In dit onderzoek zijn er 48 verschillen in scoring in de scoretabellen tussen twee beoordelaars gevonden. In Tabel 5 is het aantal verschillen per subschaal en participant (P1 tot en met P5) tussen de twee beoordelaars te zien.

Tabel 5*Frequentietabel aantal verschillen inter-beoordelaars per subschaal en participant*

Subschaal	P1	P2	P3	P4	P5	Totaal aantal verschillen (totaal aantal items)	Percentage van het totaal aantal items
Emotioneel communicatief gedrag (6 items)	0	0	0	1	0	1 (30)	3,3%
Receptief taalgedrag (8 items)	2	1	1	1	0	5 (40)	12,5%
Algemeen communicatief gedrag (non-verbaal) (9 items)	0	1	0	1	0	2 (45)	4,4%
Algemeen communicatief gedrag (verbaal) (10 items)	2	0	0	0	0	2 (50)	4,0%
Visueel exploratief gedrag (17 items)	5	2	3	4	3	17 (85)	20%
Motorisch exploratief gedrag (22 items)	0	0	4	8	5	17 (110)	15,5%
Auditieve exploratie (7 items)	1	1	1	1	0	4 (20)	11,4%
Totaal aantal verschillen per participant	10	5	9	16	8	48 (395)	12,2%

Tabel 5 laat zien dat 12,2% van alle items verschillend gescoord zijn tussen twee beoordelaars. De subschalen waar het vaakst verschillen in scoring gevonden zijn betreffen *Visueel exploratief gedrag* en *Motorisch exploratief gedrag*. Ook is in Tabel 5 te zien dat bij participant 1 en 4 het vaakst verschillen in scoring inter-beoordelaars gevonden zijn. De items die inter-beoordelaars het vaakst verschillend zijn gescoord, zijn:

- Item 14 onder *Visueel exploratief gedrag*: ‘De persoon kijkt naar uw gezicht en/of maakt oogcontact.’ (3 X verschillend gescoord)
- Item 16 (*Visueel exploratief gedrag*): ‘De persoon kijkt in het rond in een nieuwe omgeving’ (3 X verschillend gescoord)

Inhoudsanalyse scoretabel

Door middel van een inhoudsanalyse is gekeken naar de verschillen in de opmerkingen bij de items die verschillend gescoord zijn. De soorten verschillen zijn voorzien van een code. In Tabel 6 zijn de gebruikte codes te zien met een toelichting. Ook is te zien hoe vaak een bepaalde code is gebruikt om verschillen in scoretabellen te beschrijven.

Tabel 6*Frequentietabel gebruikte codes*

Code	Toelichting code	Freq. Inter-beoordelaars
Interpretatie informant	Er is in verschil in hoe de twee beoordelaars de informatie van de informant, gedurende de afname, interpreteren.	4
Interpretatie item	Er is, tussen de beoordelaar(s), een verschil in de interpretatie van welk gedrag passend is bij het item. Hieronder vallen ook situaties waarin de beoordeling wordt beïnvloed door persoonlijke verwachtingen van wat iemand 'zou moeten kunnen' of hoe gedrag idealiter zou verlopen.	10
Onbekend	Er is (g)een (duidelijke) opmerking genoteerd door de beoordelaar(s). Hierdoor is het niet mogelijk om uitspraken te doen over het verschil in scoring.	12
Vershil in observatie	Er is door de beoordelaars geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname.	15
Vershil in voorwaarde	Er is een verschil in of de beoordelaar(s) het item 'scoorbaar' achtte(n) gedurende de afname	2

In tabel 6 is te zien dat bij de verschillen in scoring tussen twee beoordelaars de codes '*Vershil in Observatie*', '*Onbekend*' en '*Interpretatie item*' het vaakste zijn gebruikt. Het komt dus het vaakste voor dat er door de beoordelaars is geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname. Verder komt het dus ook relatief vaak voor dat een van de twee beoordelaars geen opmerking heeft genoteerd bij diens score waardoor het niet mogelijk werd geacht om iets te zeggen over het verschil en dus 'onbekend' bleef. Tot slot komt het dus ook tien keer voor dat er een verschil is in hoe de hoe de beoordelaars een item interpreteren en welk gedrag ze hierbij passend vinden.

Reflectief gesprek tussen de twee beoordelaars

Uit het reflectieve gesprek tussen de twee beoordelaars kwamen verschillende factoren naar voren die de beoordeling van gedrag bemoeilijkten of beïnvloedden. Deze factoren zijn thematisch gecategoriseerd.

Het eerste thema is: *Moeilijkheden in interpretatie door beperkingen van participant*. Sommige gedragingen bleken lastig te interpreteren vanwege lichamelijke of zintuiglijke beperkingen van de participant. Een voorbeeld hiervan is participant 1, die een visuele beperking heeft en daarnaast myokymie (trillend ooglid). Hierdoor vonden beide beoordelaars het lastig om item 3 van de subschaal Visueel exploratief gedrag 'De persoon kijkt lang(er) naar een object of beeld' te scoren. Het tweede thema is: *Vluchtigheid van gedrag*. Soms komt het voor dat een

participant slechts éénmalig en vluchtig bepaald gedrag laat zien. Deze kortdurende gedragingen werden daardoor niet altijd opgemerkt door de beoordelaars. Bij item 9 van Visueel exploratief gedrag 'De persoon lokaliseert een object door hoofd- en/of oogbewegingen' — draaide participant 2 het hoofd kort in de richting van een object. Beoordelaar 1 zag dit en scoorde 'ja', terwijl beoordelaar 2 het gedrag niet opmerkte en 'nee' scoorde. Omdat het gedrag zich niet herhaalde, was er geen tweede kans om het waar te nemen.

Het derde thema is: *Manier van scoren*. Uit het gesprek bleek ook dat beoordelaar 2 in sommige gevallen nauwkeuriger te werk is gegaan dan beoordelaar 1. Zo heeft beoordelaar 2 bij items waarbij tijd een rol speelt, zoals bij item 3 onder *Motorisch exploratief gedrag* 'De persoon houdt langere tijd, minimaal 30 seconden, een object vast' gebruikgemaakt van een timer, terwijl beoordelaar 1 zich meer baseerde op een inschatting van het kunnen.

Thema vier is: *Inconsistentie tussen de itembeschrijving en de instructie*. Bij item 10 onder de subschaal *Visueel exploratief gedrag* staat in de itembeschrijving dat de participant zijn ogen en/of hoofd beweegt om een *persoon* te lokaliseren. In de instructie wordt echter aangegeven dat als een motorische beperking aanwezig is, een positieve score mag worden gegeven wanneer de participant een *object* met de ogen volgt. Dit verschil leidde tot verwarring over het item bij de beoordelaars.

Thema vijf is: *Verschillen in interpretatie van iteminhoud*. De beoordelaars verschilden soms in hoe zij de inhoud of bedoeling van een item begrepen, ook wanneer zij hetzelfde gedrag waarnamen. Een voorbeeld hiervan is item 16 onder de subschaal *Visueel exploratief gedrag* 'De persoon kijkt in het rond in een nieuwe omgeving'. Beoordelaar 1 interpreteerde dit als het reageren op nieuwe elementen in de directe omgeving (zoals het testmateriaal of de aanwezigheid van de beoordelaars), terwijl beoordelaar 2 het item zo opvatte dat de gehele fysieke omgeving nieuw moet zijn voor de participant.

Thema zes is: *Verschillen in interpretatie van subschaal*. Uit het reflectieve gesprek bleek dat de beoordelaars binnen de subschaal *Motorisch exploratief gedrag* de nadruk op verschillende aspecten legden. Beoordelaar 1 legde de nadruk meer op het exploratieve terwijl beoordelaar 2 zich vooral richtte op de motorische vaardigheid van de handeling. Een voorbeeld hiervan is te zien bij item 1 onder deze subschaal 'De persoon exploreert actief de omgeving' tijdens de afname van participant 4. Tijdens de afname was te zien dat participant 4 zichzelf door de ruimte schoof om een voorkeursobject te pakken. Beoordelaar 2 beschouwde dit als actief exploratief gedrag en scoorde 'ja'. Beoordelaar 2 zag het enkel als een gerichte actie naar een bekend object, zonder exploratie en scoorde 'nee'.

Thema zeven is: *Samengestelde items*. Bij item 16 onder de subschaal Motorisch Exploratief gedrag ontstond verwarring onder de beoordelaars. Het item vraagt of de participant een object laat

vallen én daarbij interesse toont in het loslaten of vallen van dat object. Beide gedragingen moeten aanwezig zijn voor een positieve score, maar de beoordelaars hanteerden ook verschillende opvattingen over wat precies als 'interesse' kan worden beschouwd.

Thema acht is: *beperkte observatieomstandigheden*. De beoordelaars noemen dat bij de afname van participant 3 de kamer nogal donker was. Om die reden vond beoordelaar 2 het bijvoorbeeld lastig om item 14 onder *Visueel exploratief gedrag* 'De persoon kijkt naar uw gezicht en/of maakt oogcontact' te scoren met de gemaakte videobeelden.

Tot slot, thema 9 is: *Invloed van verwachtingen*. Beoordelaar 1 merkte tijdens het gesprek op dat bepaald gedrag mogelijk is gemist omdat het niet paste binnen het beeld dat zij van een participant had gevormd. Zo had zij bij participant 4 de indruk dat hij vrijwel alle aangeboden materialen afwees. Door deze verwachting mistte zij bijvoorbeeld dat participant 4 op een gegeven moment wel een knijpfluitje vasthield en hier in kneep. Zij scoorde daardoor 'nee' op item 21 onder de subschaal *Motorisch exploratief gedrag* 'De persoon kan ergens in knijpen'. Beoordelaar 1 had het moment van knijpen wel gezien en 'ja' gescoord.

Kwalitatieve beschrijving eindverslagen

Voor dit onderzoek is per participant gekeken in hoeverre beide beoordelaars dezelfde sterke en zwakke(re) punten van een participant in hun eindverslag hebben genoemd. Voor de leesbaarheid worden de resultaten per participant besproken.

Bij participant 1 noemen beide beoordelaars als kracht de emotionele expressie: participant 1 laat blijdschap, verdriet en boosheid zien via mimiek en geluiden. Ook de gevoeligheid voor auditieve prikkels, zoals muziek en ritselende materialen, wordt in beide verslagen als sterk punt genoemd. Daarnaast geven beide beoordelaars aan dat participant 1 keuzes kan maken tussen twee objecten. Naast deze overeenkomsten zitten de verschillen in de beoordeling van het visuele en motorische functioneren. Beoordelaar 1 beschrijft participant 1 als meer visueel passief en beperkt tot het directe gezichtsveld. Beoordelaar 2 ziet juist dat participant 1 glimmende objecten volgt en gericht bekijkt.

Bij participant 2 noemen beide beoordelaars als sterk punt de gevoeligheid voor auditieve prikkels, vooral muziek. Participant 2 gebruikt woorden zoals "ja", "nee" en korte zinnen. Als zwakker punt zien beide beoordelaars de beperkte motorische exploratie. Ook wordt de visuele aandacht door beide beoordelaars als vluchtig omschreven.

Bij participant 3 noemen beide beoordelaars de emotionele communicatie, met duidelijke uiting van blijdschap en boosheid. Als zwakkere punten worden door beide de sterke onrust en beperkte visuele en auditieve aandacht genoemd.

Bij participant 4 beschrijven beide beoordelaars als kracht dat participant 4 dagstructuren en verwijzers goed herkent, en dat participant 4 zich via non-verbale communicatie duidelijk kan maken. Als zwakker punt noemen beiden de beperkte exploratie, waarbij participant 4 zich vooral richt op het eigen voorkeursmateriaal en ander materiaal vluchtig bekijkt of wegduwt. De verschillen zitten in accenten: beoordelaar 2 ziet meer interactie en initiatief in het gedrag van participant 4, terwijl beoordelaar 1 vooral de beperking tot voorkeursmateriaal lijkt te benadrukken.

Bij participant 5 noemen beide beoordelaars als sterke kanten dat participant 5 woorden in de directe context begrijpt, eenvoudige verzoeken opvolgt en zich via woorden en korte zinnen uitdrukt. Beide beoordelaars noemen daarnaast herkenbare emotionele expressie bij participant 5, waaronder blijdschap, boosheid en verdriet. Als zwakke punten noemen beiden de beperkte visuele exploratie, mogelijk door achteruitgang in visus. Auditieve prikkels dichtbij roepen volgens beide beoordelaars wel een reactie op, maar geluiden op afstand niet.

Intra-beoordelaarsbetrouwbaarheid

Om te onderzoeken of de scoretabellen van het GTI-R hetzelfde ingevuld worden door één beoordelaar over tijd, zijn er per participant kruistabellen gemaakt en is het percentage overeenkomst per subschaal berekend. In Tabel 7 is het percentage overeenkomst per subschaal te zien.

Tabel 7

Percentage van overeenkomst in de scoretabellen per subschaal

Subschaal	Intra-beoordelaar
Visueel exploratief gedrag	90,6%
Motorisch exploratief gedrag	92,6%
Auditieve exploratie	100%

Noot: intra-beoordelaarsbetrouwbaarheid is alleen geanalyseerd voor de subschalen die gescoord worden d.m.v. observatie en gefilmd zijn. Dit zijn de laatste drie subschalen.

Het percentage overeenkomst intra-beoordelaar (alleen gemeten bij de observatieschalen) liggen de tussen de 90,6% en 100%, met *Auditieve exploratie* als hoogste en *Visueel exploratief gedrag* als laatste. Alle gemeten percentages van overeenkomst liggen boven de grens van 80%, wat binnen dit onderzoek als acceptabel wordt beschouwd.

In dit onderzoek zijn er 17 verschillen in scoring in de scoretabellen tussen twee scoringsmomenten door één beoordelaar gevonden. In Tabel 8 is het aantal verschillen per subschaal en per participant intra-beoordelaar te zien.

Tabel 8

Frequentietabel aantal verschillen intra-beoordelaar per subschaal

Subschaal	P1	P2	P3	P4	P5	Totaal aantal verschillen (totaal aantal items)	Percentage van het totaal aantal items
Visueel exploratief gedrag (17 items)	1	1	3	2	1	8(85)	9,4%
Motorisch exploratief gedrag (22 items)	0	1	4	3	1	9(110)	8,1%
Auditieve exploratie (7 items)	0	0	0	0	0	0(20)	0,0%
Totaal aantal verschillen per participant	1	2	7	5	2	17 (215)	7,9%

Noot: Alleen de subschalen op basis van observatie zijn een tweede keer gescoord door beoordelaar 2. Dat zijn de laatste drie subschalen van het GTI-R.

In Tabel 8 is te zien dat 7,9% van de items van de laatste 3 subschalen verschillend is gescoord intra-beoordelaar. De subschaal *Motorisch exploratief gedrag* is het vaakst verschillend gescoord. Bij de subschaal *Auditieve exploratie* zijn geen verschillen gevonden. Bij participant 3 zijn het vaakst verschillen in scoring intra-beoordelaar gevonden.

De items die intra-beoordelaar het vaakst verschillend zijn gescoord, zijn:

- Item 8 onder *Motorisch exploratief gedrag*: 'De persoon kan ergens op slaan' (3 X verschillend gescoord)
- Item 10 onder *Visueel exploratief gedrag*: 'De persoon draait zijn ogen en/of hoofd om een persoon te lokaliseren' (2 X verschillend gescoord)

Inhoudsanalyse scoretabellen

Door middel van een inhoudsanalyse is gekeken naar de verschillen in de opmerkingen bij de items die verschillend gescoord zijn. De soorten verschillen zijn voorzien van een code. In Tabel 7 zijn de gebruikte codes te zien met een toelichting en een voorbeeld. Ook is te zien hoe vaak een bepaalde code is gebruikt om verschillen in scoretabellen te beschrijven.

Tabel 7

Frequentietabel gebruikte codes

Code	Toelichting code	Freq. Intra-beoordelaar
Interpretatie item	Er is, tussen de beoordelaars, een verschil in de interpretatie van welk gedrag passend is bij het item. Hieronder vallen ook situaties waarin de beoordeling wordt beïnvloed door persoonlijke verwachtingen van wat iemand 'zou moeten kunnen' of hoe gedrag idealiter zou verlopen.	2
Onbekend	Er is (g)een (duidelijke) opmerking genoteerd door de beoordelaar(s). Hierdoor is het niet mogelijk om uitspraken te doen over het verschil in scoring.	5
Vershil in observatie	Er is door de beoordelaars geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname.	4
Vershil in informatie	Er is een verschil in informatie die de tweede beoordelaar heeft tussen het eerste en tweede scoringsmoment	3
Vershil in voorwaarde	Er is een verschil in of de beoordelaar(s) het item 'scoorbaar' achtte(n) gedurende de afname	3

Aan de hand van Tabel 7 kan gesteld worden dat de codes '*Onbekend*' en '*Vershil in observatie*' het vaakste voorkomen tussen twee scoringsmomenten van één beoordelaar. Het komt dus het vaakste voor dat dat beoordelaar 2 tijdens een van de scoringsmomenten geen opmerking heeft genoteerd bij diens score waardoor het niet mogelijk werd geacht om iets te zeggen over het verschil en dus 'onbekend' bleef. Verder komt het even vaak voor dat er door de beoordelaars is geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname. De code '*Vershil in informatie*' is een specifieke code voor de verschillen in scoring tussen de twee scoringsmomenten van beoordelaar 2. Deze code is toegekend wanneer beoordelaar 2 tijdens het eerste scoringsmoment informatie kreeg via de observant, terwijl die bij het tweede scoringsmoment, dat via videobeelden verliep, ontbrak. Deze code is drie keer toegekend.

Conclusie

Dit onderzoek had als hoofddoel om de betrouwbaarheid van het Gedragstaxatie-Instrument Revised (GTI-R) te onderzoeken, specifiek gericht op de inter- en intra-beoordelaarsbetrouwbaarheid. Daarbij is gekeken naar de mate van overeenkomst tussen ingevulde scoretabellen door verschillende beoordelaars (inter-beoordelaars), en dezelfde beoordelaar over tijd (intra-beoordelaar). Verder is onderzocht waar verschillen in scoring vandaan komen. Voor de intra-beoordelaarsbetrouwbaarheid is dit laatste onderzocht middels een inhoudsanalyse van de GTI-R scoreballen. Voor de inter-beoordelaarsbetrouwbaarheid werd dit uitgebreid met een reflectief gesprek tussen de beoordelaars en een analyse van de eindverslagen.

Inter-beoordelaarsbetrouwbaarheid (onderzoeksvraag 1)

Uit de analyse van de scoretabellen dat het percentage overeenkomst tussen twee beoordelaars per subschaal, tussen de 80,0% en 98,0% ligt. Dit valt ruim binnen de acceptabele mate die gesteld is binnen dit onderzoek (>80%).

De meeste verschillen in scoring werden gevonden in de subschalen *Visueel exploratief gedrag* en *Motorisch exploratief gedrag*. De participanten waar de meeste verschillen werden gevonden zijn participant 1 & 4.

De kwalitatieve inhoudsanalyse van de verschillen in scoring leverde extra inzicht in de aard van deze verschillen. Hieruit bleek dat het vaakste voorkwam dat er door de beoordelaars is geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname. Ook kwam het vaak voor dat een van de twee beoordelaars geen opmerking had genoteerd bij diens score waardoor het niet mogelijk werd geacht om iets te zeggen over het verschil en dus 'onbekend' bleef.

Om de overwegingen achter de gegeven scores nog beter te kunnen plaatsen, zijn de beoordelaars met elkaar in gesprek gegaan over alle gevonden verschillen in scoring. Hieruit is gebleken dat bepaalde beperkingen van de participant, vluchtig voorkomend gedrag en beperkte observatieomstandigheden (slechte verlichting) de beoordeling mogelijk beïnvloedden. Ook waren er verschillen in werkwijze (bijv. gebruik van een timer) tussen de beoordelaars, werden bepaalde items anders geïnterpreteerd en bleken er in sommige items onduidelijkheden te zitten. In enkele gevallen speelden persoonlijke verwachtingen over de participant mogelijk een rol in de scoring. Een voorbeeld hiervan is dat beoordelaar 1 de verwachting had gevormd dat participant 4 vrijwel alle aangeboden materialen afwees. Door deze verwachting mistte de beoordelaar dat participant 4 op een gegeven moment wel een knijpfluitje vasthield en hier in kneep.

Door middel van een descriptieve analyse is gekeken naar in hoeverre twee beoordelaars dezelfde sterke en zwakke(re) punten van de participanten in de eindverslagen benoemen. Over het algemeen zijn hier grote overeenkomsten gevonden in de genoemde sterke en zwakke(re) punten. Ook waren verschillen zichtbaar, hoewel dit varieerde per participant. Met name bij participant 1 liepen de eindverslagen wat de genoemde sterke en zwakke(re) punten betreft uiteen. Dit is tevens ook een van de twee participanten waar de meeste verschillen in scoring tussen de twee beoordelaars is gevonden. Een mogelijke verklaring hiervoor zit in de visuele beperkingen van de participant (Myokymie) die de scoring bemoeilijkte.

Intra-beoordelaarsbetrouwbaarheid (onderzoeksvraag 2)

Qua intra-beoordelaarsbetrouwbaarheid blijkt uit de analyse van de scoretabellen een hoge mate van overeenkomst in scoring tussen één beoordelaar over tijd, tussen de tussen 90,6% en 100% per subschaal. Dit wijst erop dat het GTI-R stabiel is gescoord door dezelfde persoon over tijd. Dit valt ruim binnen de acceptabele mate die gesteld is binnen dit onderzoek (>80%).

De meeste verschillen intra-beoordelaar deden zich voor binnen de subschaal *Motorisch exploratief gedrag*. Bij participant 3 zijn het vaakst verschillen in scoring intra-beoordelaar gevonden. Bij de subschaal *Auditieve exploratie* zijn geen verschillen in scoring gevonden tussen één beoordelaar over tijd.

Discussie

Een belangrijke bevinding van dit onderzoek is dat er over het algemeen sprake was van een hoge mate van overeenstemming. Zowel inter-beoordelaar als intra-beoordelaar. Tegelijkertijd laat de analyse zien dat, waar verschillen voorkwamen, deze vooral zichtbaar waren binnen de subschalen *Visueel exploratief gedrag* en *Motorisch exploratief gedrag*. Een mogelijke verklaring hiervoor is dat deze subschalen volledig gebaseerd zijn op observatie. Het komt vaker voor dat observationele studies die onderzoek doen naar de doelgroep (Z)EVMB een lagere inter-beoordelaarsbetrouwbaarheid kennen (Maes et al., 2021). Volgens Nicholas, Geers en Rollins (1999) kan de lage overeenstemming tussen beoordelaars bij deze doelgroep mogelijk juist weerspiegelen in hoeverre deze personen werkelijk worden begrepen. Daarnaast is observatie als methode gevoelig voor observatie bias, waarbij de achtergrond en opvatting van de onderzoeker kan beïnvloeden wat hij waarneemt (Zohrabi, 2013).

Uit de resultaten bleek verder dat de subschalen die grotendeels interview-gebaseerd zijn, zoals *Emotioneel communicatief gedrag* en *Algemeen communicatief gedrag*, gemiddeld hogere percentages overeenkomst tonen. Deze subschalen geven de onderzoeker mogelijk meer gelegenheid om door te vragen en betekenis toe te kennen aan antwoorden binnen de context van het gesprek (Zohrabi, 2013). Dit maakt het mogelijk om onduidelijkheden of interpretatieverschillen direct te verhelderen. Daarnaast is het geval dat beide beoordelaars bij de interviewvragen exact dezelfde antwoorden te horen hebben gekregen en er daardoor mogelijk minder ruimte is voor eigen interpretatie dan bij de observatieschalen.

Het onderscheid in mate van overeenkomst tussen observatie- en interviewgebaseerde subschalen onderstreept het belang van een gecombineerde benadering binnen het GTI-R. Door observatie en interview te combineren wordt gedrag vanuit meerdere perspectieven bekeken, wat bijdraagt aan een vollediger en genuanceerder beeld van het functioneren van de persoon. Dit sluit aan bij het eerder in de inleiding genoemde artikel van Lyons et al. (2016), die stellen dat het combineren van verschillende perspectieven essentieel is om een goed beeld te hebben van de wensen, voorkeuren en mogelijkheden.

Verder is bij de interpretatie van zowel de eindverslagen als de ingevulde scoretabellen een contextuele factor mogelijk ook relevant om te noemen: drie van de vijf participanten (1, 3 en 5) waren woonachtig op de zorglocatie waar beoordelaar 2 werkzaam is. Hierdoor bestond voorafgaand aan de afname een zekere mate van voorkennis. Hoewel geprobeerd is om beoordelaar 1 zoveel mogelijk gelijke informatie te geven via dossieranalyse en overdracht (zoals door het GTI-R vereist), kan niet volledig uitgesloten worden dat deze voorkennis invloed heeft gehad op de

interpretatie van gedrag. Dit wordt ondersteund door recent onderzoek van MacLean en Miller (2024), waarin werd aangetoond dat kennis of verwachtingen over participanten het scoringsgedrag van observatoren kan beïnvloeden. Hierop valt wel aan te merken dat de observatie-items in het GTI-R relatief concreet omschreven zijn, wat de invloed van voorkennis mogelijk heeft beperkt. De inzet van het bestaande netwerk was een bewuste keuze om praktische haalbaarheid te garanderen binnen de beperkte tijd en toegankelijkheid van de doelgroep.

Ook is het belangrijk te benadrukken dat het percentage van overeenkomst als gebruikte maat geen correctie toepast voor toevallige overeenstemming. Hierdoor kan de daadwerkelijke betrouwbaarheid iets lager liggen dan de berekende percentages doen vermoeden. Deze maat houdt geen rekening met de kans dat beoordelaars het eens zijn puur op basis van toeval, wat kan leiden tot een overschatting van de werkelijke betrouwbaarheid. Alternatieve maten, zoals Cohen's Kappa (Cohen, 1960), corrigeren voor deze toevallige overeenstemming en bieden daardoor een meer accurate schatting van de inter en intra-beoordelaarsbetrouwbaarheid (McHugh, 2012). Het ontbreken van dergelijke correcties in dit onderzoek vormt een methodologische beperking, maar was niet mogelijk vanwege de kleine steekproef. Tegelijkertijd is deze methode geschikt als eerste stap in het vaststellen van de mate van overeenstemming, vooral in situaties met binaire scores en een kleine steekproefomvang (McHugh, 2012).

Een kracht van dit onderzoek is dat ondanks dat de steekproef klein is, deze wel divers is. De jongste participant is 16 en de oudste 60 jaar. Drie van de participanten zijn man en twee vrouw. Ook zijn de participanten erg verschillend in het wel of niet hebben van epilepsie en de aard van een motorische beperking en visuele of auditieve beperking. Deze variatie weerspiegelt de grote spreiding die kenmerkend is voor de doelgroep mensen met een (Z)EVMB, waarbij de combinatie en intensiteit van beperkingen per individu sterk kan verschillen (Van der Putten et al., 2017). Hoewel generaliseerbaarheid met zo'n kleine steekproef beperkt is, maakt deze diversiteit het wel mogelijk om te onderzoeken of de inter en intra-beoordelaarsbetrouwbaarheid enigszins consistent blijft bij mensen met verschillende soorten beperkingen. Een mogelijke verwachting is dat personen met ernstigere of complexere beperkingen moeilijker te scoren zijn, omdat hun gedrag vaak subtieler is (Van Timmeren et al., 2017a).

Tot slot is het gebruik van een verklarend sequentieel design, waarbij kwalitatieve analyse helpt om kwantitatieve verschillen te duiden een kracht van dit onderzoek. Er is niet alleen gekeken naar of er verschillen in scoring zijn, maar ook waar de verschillen in zitten. Dit maakt het mogelijk om beter te begrijpen welke subschalen, items, gedragingen of contextuele factoren aanleiding geven tot interpretatieverschillen (Fetters, 2013). Die verdieping biedt waardevolle aanknopingspunten voor het verbeteren van het instrument. Op basis van deze inzichten kunnen

kleine aanpassingen aan instructies worden overwogen of kan extra aandacht in training worden besteed aan specifieke onderdelen die relatief vaker tot verschillen in scoring leiden. Hoewel de mate van overeenkomst in dit onderzoek over het algemeen hoog was en het om een kleinschalige studie ging, kunnen dergelijke verfijningen bijdragen aan het verder versterken van de inter- en intra-beoordelaarsbetrouwbaarheid van de GTI-R.

Aanbevelingen voor vervolgonderzoek

In dit onderzoek is ingezoomd op één aspect van de psychometrische kwaliteit van het GTI-R, namelijk de intra- en interbeoordelaarsbetrouwbaarheid. Hoewel dit een belangrijk onderdeel is bij het beoordelen van de bruikbaarheid van het instrument, is het wenselijk dat in vervolgonderzoek ook andere psychometrische eigenschappen worden onderzocht. Gezien de herziening van het instrument is het bijvoorbeeld relevant om te kijken naar de constructvaliditeit en of het GTI-R daadwerkelijk de concepten meet die het beoogt te meten. Psychometrische onderbouwing is, zoals eerder in de inleiding, genoemd namelijk van belang om van adequate assessment te kunnen voorzien (Wessels, 2021). Adequate assessment is vervolgens weer van belang om *zorg die werkt volgens de principes van EBP* te kunnen bieden. Zorg die aansluit bij de behoeften van de cliënt, gebaseerd is op wetenschappelijk onderbouwde methoden en wordt ondersteund door de expertise van professionals (Elser et al., 2025)

Om de betrouwbaarheid van het GTI-R breder te onderzoeken, wordt aanbevolen om het onderzoek te herhalen met een grotere groep participanten en beoordelaars. Dit maakt het mogelijk om ook alternatieve maten zoals Cohen's Kappa (Cohen, 1960) toe te passen, omdat deze corrigeren voor toevallige overeenstemming en daardoor een meer accurate schatting van de inter en intra-beoordelaarsbetrouwbaarheid bieden (McHugh, 2012). Mogelijk is het ook relevant om te onderzoeken wat de invloed van voorkennis, uit een eerdere behandelrelatie, is op de interpretatie van gedrag.

Aanbevelingen voor de praktijk

Uit de kwalitatieve analyse van de beoordelaarsgesprekken blijkt dat verschillen in scoring vaak voortkwamen uit interpretatieverschillen: beoordelaars verschilden in wat zij als 'reiken' beschouwden, of hoe zij 'kijken naar een nieuw object' interpreteerden. Ook bleek dat sommige gedragingen erg vluchtig waren, waardoor ze door één van beide beoordelaars gemist werden. Dit soort verschillen zijn niet ongebruikelijk bij het observeren van mensen met (Z)EVMB, omdat hun gedrag vaak subtiel, weinig voorspelbaar en sterk context gebonden is (Van Timmeren et al., 2017; Visser et al., 2017; Wessels et al., 2020). Daarnaast bleek uit het gesprek tussen de twee beoordelaars dat sommige items zelf aanleiding gaven tot verwarring, bijvoorbeeld door

onduidelijke formuleringen of tegenstrijdigheden tussen de itembeschrijving en scoringsinstructie. Dat laat zien dat niet alleen de observaties maar ook de opzet en duidelijkheid van het instrument invloed hebben op de betrouwbaarheid.

Om deze interpretatieverschillen te verkleinen en de betrouwbaarheid te verhogen, is het aan te bevelen om extra aandacht te besteden aan de onduidelijke items en/of scoringsinstructies die genoemd zijn in het gesprek tussen de twee beoordelaars. Daarnaast kan training van beoordelaars bijdragen aan meer consistentie in beoordeling (Robertson et al., 2018). Er is al een training voor het GTI-R. Door tijdens deze training nog meer gezamenlijk te oefenen met videobeelden, kunnen beoordelaars hun interpretatie van de items op elkaar afstemmen. En komt er waar mogelijk nog meer feedback over items van het GTI-R die onduidelijk zijn.

Tot slot kwam tijdens de inhoudsanalyse van de verschillen in de scoretabellen van dit onderzoek relatief vaak voor dat een van de twee beoordelaars geen opmerking heeft genoteerd bij diens score waardoor het niet mogelijk werd geacht om iets te zeggen over het verschil en dus 'onbekend' bleef. Hierdoor mist data in dit onderzoek. Voor vervolgonderzoek wordt aangeraden dat de beoordelaars in alle gevallen een opmerking schrijven bij de gegeven score. Ook wanneer zij denken dat de observatie voor zich spreekt. Een korte opmerking maakt het mogelijk om achteraf beter te begrijpen vanuit welke overweging een score tot stand is gekomen, wat niet alleen bijdraagt aan de transparantie van het proces, maar ook waardevol is bij het bespreken van verschillen, het verbeteren van de instructie en het versterken van de betrouwbaarheid van het instrument.

De resultaten van dit onderzoek wijzen erop dat het GTI-R in staat is om zowel inter als intra-beoordelaar, betrouwbaar de functionele mogelijkheden van mensen met een (Z)EVMBB in kaart te brengen. De afname van het GTI-R levert betrouwbare informatie op over sterke kanten en ontwikkelpunten van een persoon, die richting kan geven aan de ondersteuning. Dit maakt het instrument relevant voor praktijktoepassing, zeker gezien het beperkte aantal bestaande instrumenten dat zowel doelgroep-specifiek als psychometrisch onderzocht is (Carnaby, 2007, Wessels et al., 2021)

Literatuur

- Appl, D. J. (2000). Clarifying the preschool assessment process: traditional practices and alternative approaches. *Early Childhood Education Journal*, 27(4), 219-225.
- Avrumson, R., Komatina, N., & Murphy, M. F. (2025). *The art and the science of ensuring inter-rater reliability*. Worldwide Clinical Trials. <https://www.worldwide.com/wp-content/uploads/2022/01/Whitepaper-ArtAndScienceOfInterraterReliability-VF.pdf>
- Brentnall, J., & Bundy, A. C. (2009). The concept of reliability in the context of observational assessments. *OTJR: Occupation, Participation and Health*, 29(2), 63-71.
- Buntinx, W.H.E. and Schalock, R.L. (2010), Models of Disability, Quality of Life, and Individualized Supports: Implications for Professional Practice in Intellectual Disability. *Journal of Policy and Practice in Intellectual Disabilities*, 7: 283-294. <https://doi.org/10.1111/j.1741-1130.2010.00278.x>
- Carnaby, S. (2007). Developing good practice in the clinical assessment of people with profound intellectual disabilities and multiple impairment. *Journal of Policy and Practice in intellectual Disabilities*, 4(2), 88-96. <https://doi.org/10.1111/j.1741-1130.2007.00105.x>
- Cline, T. (2018). *The Assessment of Special Educational Needs: International Perspective* (Vol. 7). Routledge.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Elser, A., Scherer, M., Stadel, M., Garbade, S. F., & Diermayr, G. (2025). Reliability and validity of the German “Evidence-Based Practice Confidence (EPIC) Scale” for allied health professionals. *Zeitschrift Für Evidenz Fortbildung Und Qualität Im Gesundheitswesen*. <https://doi.org/10.1016/j.zefq.2024.12.006>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs- principles and practices. *Health services research*, 48(6 Pt 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- George, T. (2024, 18 juni). *Mixed-methods-onderzoek | Uitleg met voorbeelden*. Scribbr. <https://www.scribbr.nl/onderzoeksmethoden/mixed-methods/>
- Hennink, M., Hutter, I., & Bailey, A. (2020). *Qualitative research methods*. SAGE Publications Limited.
- Hogg, J., & Langa, A. (2005). Introduction: Assessment in perspective. In J. Hogg & A. Langa (Eds.), *Assessing adults with intellectual disabilities: A service providers’ guide* (pp. 1–5). Oxford: The British Psychological Society and Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470773697>

- Legendre, P. (2005). Species associations: the Kendall coefficient of concordance revisited. *Journal Of Agricultural Biological And Environmental Statistics*, 10(2), 226–245.
<https://doi.org/10.1198/108571105x46642>
- Lyons, G. S., De Bortoli, T., & Arthur-Kelly, M. (2016). Triangulated Proxy Reporting: a technique for improving how communication partners come to know people with severe cognitive impairment. *Disability and rehabilitation*, 39(18), 1814-1820.
<https://doi.org/10.1080/09638288.2016.1211759>
- MacLean, C. L., & Miller, G. S. (2024). Trust but verify: The biasing effects of witness opinions and background knowledge in workplace investigations. *Journal Of Safety Research*, 89, 33–40.
<https://doi.org/10.1016/j.jsr.2024.01.009>
- Maes, B., Nijs, S., Vandesande, S., Van keer, I., Arthur-Kelly, M., Dind, J., Goldbart, J., Petitpierre, G., & Van der Putten, A. (2021). Looking back, looking forward: Methodological challenges and future directions in research on persons with profound intellectual and multiple disabilities. *Journal of Applied Research in Intellectual Disabilities*, 34(1), 250–262.
<https://doi.org/10.1111/jar.12803>
- Mansell, J., & Beadle-Brown, J. (2004). Person-centred planning or person-centred action? A response to the commentaries. *Journal of Applied Research in Intellectual Disabilities*, 17, 31–35. <https://doi.org/10.1111/jar.2004.17.issue-1>
- McHugh M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
- Nakken, H., & Vlaskamp, C. (2007). A Need for a Taxonomy for Profound Intellectual and Multiple Disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, 4(2), 83-87.
<https://doi.org/10.1111/j.1741-1130.2007.00104.x>
- Nicholas, J. G., Geers, A. E., & Rollins, P. R. (1999). Inter-rater reliability as a reflection of ambiguity in the communication of deaf and normally-hearing children. *Journal of Communication Disorders*, 32(2), 121–134. [https://doi.org/10.1016/s0021-9924\(99\)00002-7](https://doi.org/10.1016/s0021-9924(99)00002-7)
- Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 39(4), 214-223. <https://doi.org/10.1111/j.1469-8749.1997.tb07414.x>
- Rankl, F., Johnson, G. A., & Vindrola-Padros, C. (2021). Examining What We Know in Relation to How We Know It: A Team-Based Reflexivity Model for Rapid Qualitative Health Research. *Qualitative Health Research*, 31(7), 1358–1370. <https://doi.org/10.1177/1049732321998062>

- Robertson, R. L., Vergis, A., Gillman, L. M., & Park, J. (2018). Effect of rater training on the reliability of technical skill assessments: a randomized controlled trial. *Canadian journal of surgery. Journal canadien de chirurgie*, 61(6), 15917. Advance online publication. <https://doi.org/10.1503/cjs.015917>
- Sandelowski M. Sample size in qualitative research. *Res Nurs Health*. 2000;18(2):179–83.
- Thompson, T., Coleman, J. M., Riley, K., Snider, L. A., Howard, L. J., Sansone, S. M., & Hessel, D. (2018). Standardized assessment accommodations for individuals with intellectual disability. *Contemporary school psychology*, 22(4), 443-457. Doi: 10.1007/s40688-018-0171-4
- University workplace storage. (2025, 27 maart). University Of Groningen. <https://www.rug.nl/digital-competence-centre/it-solutions/store-and-archive/university-workplace-storage?lang=en>
- Van der Putten, A., Vlaskamp, C., Luijkx, J., Poppes, P., & Groningen, R. (2017). Kinderen en volwassenen met zeer ernstige verstandelijke en meervoudige beperkingen: tijd voor een nieuw perspectief. Groningen: Research Centre EMB, Rijksuniversiteit Groningen (A)Van Timmeren, E. A., Waninge, A., Van Schrojenstein Lantman-de, H. M. J., Van der Putten, A. A. J., & Van der Schans, C. P. (2017). Patterns of multimorbidity in people with severe or profound intellectual and motor disabilities. *Research in Developmental Disabilities*, 67, 28-33.
- (B) Van Timmeren, E. A., Schans, C. P., Putten, A. A. J. van der, Krijnen, W. P., Steenbergen, H. A., Schrojenstein Lantman-de Valk, H. M. J., & Waninge, A. (2017). Physical health issues in adults with severe or profound intellectual and motor disabilities: a systematic review of cross-sectional studies. *Journal of Intellectual Disability Research*, 61, 30-49.
- Van Timmeren, E. A. (2019). Physical health in adults with severe or profound intellectual and motor disabilities. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.
- Vanderplasschen, W., Vindevogel, S., & Broekaert, E. (2009). Het meten van effectiviteit en efficiëntie in de integrale jeugdhulp: droom of nachtmerrie. K. De Koster, K, De Vos, D. Kerger, G. Roets, & R. Roose (Reds.), *Handboek Integrale Jeugdhulp*. Politeia.
- Vandesande, S., Van keer, I., Dhondt, A., & Maes, B. (2020). The social-emotional functioning of young children with a significant cognitive and motor developmental delay. *International Journal of Developmental Disabilities*. <https://doi.org/10.1080/20473869.2020.1805574>
- Vig, S., & Sanders, M. (2007). Assessment of Mental Retardation. In M. R. Brassard & A. E. Boehm (Eds.), *Preschool assessment: Principles and practices* (pp. 420-446). Guilford Press.

- Visser, L. (2014). The Bayley-III-NL special needs edition: a suitable developmental assessment instrument for young children with special needs [University of Groningen]. Stichting Kindstudie
- Vlaskamp, C. (2005). Interdisciplinary assessment of people with profound intellectual and multiple disabilities. In J. Hogg & A. Langa (Eds.), *Assessing adults with intellectual disabilities: A service providers' guide* (pp. 39–51). Blackwell Publishing.
<https://doi.org/10.1002/9780470773697.ch3>
- Vlaskamp, C., Meulen, B. F. v. d., Smrkovsky, M., & Stichting Kinderstudies (Groningen). (1999). Gedragstaxatie-Instrument GTI : voor personen met ernstige meervoudige beperkingen. Stichting Kinderstudies.
- Vlaskamp, C., Van der Meulen, B. F., & Zijlstra, H. P. (2002). De instrumentele realisering van het Gedragstaxatie-instrument [The instrumental realisation of the Behavioural Appraisal Scales]. *Tijdschrift voor Orthopedagogiek*, 41(1), 22–31
- Vlaskamp, C., Van Wijck, R., & Poppes, P. (2015). Inventarisatie Persoonsbeeld en Zorg: Versie voor kinderdagcentra [Inventory of the personal Profile and Support: Version for child daycare centres]. Stichting Kindstudies.
- Wessels, M. (2021). Looking at possibilities: research into assessment of people with profound intellectual and multiple disabilities. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.182491814>
- Wessels, M. D., Paap, M. C. S., & Van der Putten, A. A. J. (2020). Validity of an instrument that assesses functional abilities in people with profound intellectual and multiple disabilities: Look what I can do! *Journal of Intellectual & Developmental Disability*, 46(3), 250–260.
<https://doi.org/10.3109/13668250.2020.1785851>
- Wessels, M. D., Paap, M. C. S., & van der Putten, A. A. J. (2022). The content validity of the Behavioural Appraisal Scales in people with profound intellectual and multiple disabilities: A Delphi study. *Journal of Policy and Practice in Intellectual Disabilities*, 19(1), 86–101.
<https://doi.org/10.1111/jppi.12409>
- Wessels, M. D., Vlaskamp, C., Paap, M. C. S., & Van der Putten, A. A. J. (2023). Handleiding Gedragstaxatie-Instrument- Revised. Academische Werkplaats.
- Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and practice in language studies*, 3(2), 254-262.

Bijlage 1

Informatiebrief

Bijlage 1 bevat de informatiebrief die aan de wettelijk vertegenwoordigers van de participanten is gestuurd.

INFORMATIE OVER HET ONDERZOEK

VERSIE VOOR DEELNEMERS

Het bepalen van criteria voor het Gedragstaxatie-Instrument-Revised

PED-2223-S-0002

➤ **Waarom krijg ik deze informatie?**

U krijgt deze informatie omdat de persoon voor wie u verantwoordelijk bent ondersteuning ontvangt van **organisatie X**.

Dit is een onderzoek van de Rijksuniversiteit Groningen. De betrokken onderzoekers zijn Marleen Wessels, Annette van der Putten en Muirne Paap.

➤ **Moet ik meedoen aan dit onderzoek?**

Meedoen aan het onderzoek door de persoon voor wie u verantwoordelijk bent is vrijwillig. Wel is uw toestemming nodig. Lees deze informatie daarom goed door. Stel alle vragen die u misschien heeft, bijvoorbeeld omdat u iets niet begrijpt. Pas daarna besluit u of de persoon voor wie u verantwoordelijk bent meedoet. Als u besluit om niet mee te doen, hoeft u niet uit te leggen waarom, en zal dit geen negatieve gevolgen voor hem of haar hebben. Dit recht geldt op elk moment, dus ook nadat u hebt toegestemd in deelname aan het onderzoek.

➤ **Waarom dit onderzoek?**

Om goed in kaart te kunnen brengen wat de beperkingen en vaardigheden van iemand met (Zeer) Ernstige Verstandelijke en Meervoudige Beperkingen ((Z)EVMB) zijn, gebruiken we in de ondersteuning diagnostische instrumenten. Eén van die instrumenten is het Gedragstaxatie-Instrument-Revised (GTI-R). Het GTI-R is een nieuwe versie van het GTI. Het GTI bestaat uit een observatie, testsituatie en interview aan een betrokkene van iemand met (Z)EVMB. Het GTI-R beoogt functionele vaardigheden in kaart te brengen. Dit zijn vaardigheden die iemand in het dagelijks leven kan gebruiken, zoals het vasthouden van een drinkbeker. Het GTI kon gebruikt worden om op basis van scores een profiel op te stellen van sterke en zwakkere kanten van een persoon. Omdat er veel aanpassingen gedaan zijn aan GTI in de nieuwe versie, zijn de criteria die bij de oude versie gebruikt worden niet meer van toepassing. In dit onderzoek willen we deze criteria opnieuw opstellen voor het GTI-R. Daarvoor hebben we van een groep mensen met (Z)EVMB de scores nodig op het GTI-R. We willen daarnaast kijken of het instrument op een betrouwbare manier gescoord kan worden. Dat betekent dat als twee beoordelaars het instrument scoren, of één beoordelaar de afname twee keer scoort met een paar weken ertussen, de score (bijna) gelijk is.

➤ **Wat vragen we van u tijdens het onderzoek?**

Allereerst vragen we u toestemming voor deelname van uw de persoon voor wie u verantwoordelijk bent aan dit onderzoek. Wanneer u toestemming geeft, nemen we het GTI-R af. Dit bestaat uit een aantal vragen voor de begeleider van de persoon voor wie u verantwoordelijk bent, een observatie

en een test situatie, waarbij we met verschillende materialen/speelgoed bekijken hoe hij of zij reageert. De afname wordt gefilmd, zodat het twee keer te scoren is.

➤ **Welke gevolgen kan deelname hebben?**

Op basis van dit onderzoek worden criteria opgesteld voor het GTI-R. Dit betekent dat als het GTI-R wordt afgenomen in de toekomst, er op basis van de scores een profiel voor de persoon met (Z)EVMB kan worden opgesteld, van sterke en minder sterke kanten. Dit vergroot de bruikbaarheid van het instrument. Uiteindelijk beogen we hiermee een passend instrument te hebben waarmee functionele vaardigheden van de persoon met ZEVMB in kaart gebracht kunnen worden en waarmee op basis van het profiel, ondersteuning afgestemd kan worden op de mogelijkheden van de persoon met (Z)EVMB.

Deelname aan dit onderzoek betekent dat er een onderzoeker langs komt voor de GTI-R afname, die bestaat uit vragen voor een begeleider, een observatie en testsituatie voor de persoon met (Z)EVMB. De observatie en testsituatie wordt gefilmd. Een GTI-R afname kent weinig risico's. Het kan zijn dat de afname vermoeiend is of de persoon voor wie u verantwoordelijk bent het niet leuk vindt, dan zal de afname altijd gestopt worden. Er is altijd een bekende begeleider bij de afname, die gevraagd wordt om alert te zijn op de signalen van de persoon voor wie u verantwoordelijk bent. Meestal wordt een GTI-R afname echter gezien als een leuk interactiemoment met interessante materialen.

➤ **Hoe gaan we met uw gegevens om?**

Voor dit onderzoek worden verschillende persoonsgegevens verzameld. De naam van u en de persoon voor wie u verantwoordelijk bent staan op het toestemmingsformulier, zodat we weten voor wie er toestemming is. Dit wordt opgeslagen op een beveiligde map op de server van de Rijksuniversiteit Groningen en is alleen inzichtelijk voor de hoofonderzoeker, Marleen Wessels. Iedere persoon krijgt een identificatienummer, gekoppeld aan de naam. De naam wordt apart opgeslagen van de verdere gegevens en niet gebruikt in de verwerking of analyse van de gegevens.

Daarnaast worden de volgende gegevens verzameld:

-Scores op het GTI-R

-Geslacht

-Leeftijd

-Bijkomende beperkingen: visuele beperking, auditieve beperking, epilepsie, ernst van de motorische beperking.

Deze gegevens worden verzameld om een beeld te krijgen van de mensen die hebben meegedaan aan het onderzoek en omdat sommige kenmerken mogelijk van invloed zijn op de scores op het GTI-R (bijvoorbeeld, iemand met een ernstige visuele beperking scoort op bepaalde onderdelen lager dan iemand die geen visuele beperking heeft). Als laatste nemen de GTI-R afname op, zodat we kunnen bepalen of de afname op een betrouwbare manier gescoord kan worden. Deze data zijn alleen toegankelijk voor het onderzoeksteam. Ook studenten en onderzoeksassistenten kunnen behoren tot het onderzoeksteam.

De gegevens worden gebruikt voor onderzoek naar de kwaliteit van het GTI-R. Wanneer er in de toekomst een vervolgonderzoek gedaan wordt, wordt de data verzameld in het kader van dit onderzoek alleen gedeeld nadat het vervolgonderzoek is goedgekeurd door de ethische commissie.

In vervolgonderzoek wordt alleen data gebruikt die niet te herleiden is tot de persoon voor wie u verantwoordelijk bent. De gegevens worden 10 jaar bewaard en daarna verwijderd.

Tot het moment van data analyse kunt u de gegevens van de persoon voor wie u verantwoordelijk bent inzien, aanpassen of verwijderen. Over dit onderzoek wordt gepubliceerd in een internationaal tijdschrift en er worden fact sheets op de website van de Academische Werkplaats Ernstig Meervoudig Beperkt gepubliceerd. De resultaten worden nooit op individueel niveau gerapporteerd en zullen nooit terug te leiden zijn tot de persoon voor wie u verantwoordelijk bent.

➤ **Wat moet u nog meer weten?**

U kunt altijd vragen stellen over het onderzoek: nu, tijdens het onderzoek, en na afloop. Dit kan door een van de betrokken onderzoekers te e-mailen (m.d.wessels@rug.nl) of te bellen (050 363 6581).

Heeft u vragen/zorgen over uw rechten als onderzoeksdeelnemer of de uitvoering van het onderzoek? U kunt hierover ook contact opnemen met de Ethische Commissie Gedrags- en Maatschappijwetenschappen van de Rijksuniversiteit Groningen: ec-bss@rug.nl.

Heeft u vragen of zorgen over hoe er met uw persoonsgegevens wordt omgegaan? U kunt hierover ook contact opnemen met de Functionaris Gegevensbescherming van de Rijksuniversiteit Groningen: privacy@rug.nl.

Als onderzoeksdeelnemer heeft u recht op een kopie van deze onderzoeksinformatie.

Met vriendelijke groet,

Dr. Marleen Wessels

Dr. Muirne Paap

Prof. dr. Annette van der Putten

Bijlage 2

Toestemmingsformulier wettelijk vertegenwoordiger

Bijlage 2 bevat het toestemmingsformulier dat de wettelijk vertegenwoordigers van de participanten hebben ondertekend bij akkoord aan deelname.

“Het bepalen van criteria voor het Gedragstaxatie-Instrument-Revised PED-2223-S-0002

Hierbij verklaar ik, ouder/wettelijk vertegenwoordiger van:

Naam:.....

Kennis genomen te hebben van de schriftelijke informatie over het onderzoek naar het Gedragstaxatie-Instrument-Revised.

- Ik heb de informatie over het onderzoek gelezen. Ik heb genoeg gelegenheid gehad om er vragen over te stellen.
- Ik begrijp waar het onderzoek over gaat, wat er gevraagd wordt, welke gevolgen deelname kan hebben, hoe er met persoonsgegevens wordt omgegaan, en wat mijn rechten als wettelijk vertegenwoordiger van de deelnemer zijn.
- Ik begrijp dat deelname aan het onderzoek vrijwillig is. Ik kies er zelf voor om mee te doen. Ik kan op elk moment stoppen met meedoen. Als ik stop, hoef ik niet uit te leggen waarom. Stoppen zal geen negatieve gevolgen voor mij of mijn zoon/dochter hebben.
- Ik geef hieronder aan waar ik toestemming voor geef.

Toestemming voor deelname aan het onderzoek:

☐ Ja, ik geef toestemming voor deelname; deze toestemming loopt tot 08-01-2035

☐ Nee, ik geef geen toestemming voor deelname

Toestemming voor het maken van audio/video-opnames tijdens het onderzoek:

☐ Ja, ik geef toestemming voor het maken van video-opnames van degene voor wie ik verantwoordelijk ben.

☐ Nee, ik geef geen toestemming voor het maken van video-opnames van de persoon voor wie ik verantwoordelijk ben.

Toestemming voor de verwerking van de persoonsgegevens van mij en voor degene voor wie ik verantwoordelijk ben:

☐ Ja, ik geef toestemming voor de verwerking van de persoonsgegevens van degene voor wie ik verantwoordelijk ben, zoals vermeld in de onderzoeksinformatie. Ik weet dat ik tot 01-05-2025 kan vragen om de gegevens te laten verwijderen. Ook als ik of degene voor wie ik verantwoordelijk ben besluit om te stoppen met deelname, kan ik hierom vragen. De

gegevens kunnen gebruikt worden in vervolgonderzoek. Dit kan alleen na goedkeuring door de ethische commissie Pedagogische en Onderwijswetenschappen.

[] Nee, ik geef geen toestemming voor de verwerking van de persoonsgegevens van degene voor wie ik verantwoordelijk ben

Volledige naam wettelijk vertegenwoordiger:	Handtekening wettelijk vertegenwoordiger:	Datum:

Volledige naam aanwezige onderzoeker:	Handtekening onderzoeker:	Datum:

De aanwezige onderzoeker verklaart dat de deelnemer uitvoerig over het onderzoek is geïnformeerd.

U heeft recht op een kopie van dit toestemmingsformulier.

Bijlage 3

Codeboek

Bijlage 3 bevat het volledige codeboek dat is gebruikt om de verschillen in de scoretabellen te analyseren.

Code	Toelichting code	Type	Voorbeeld
Interpretatie informant	Er is in verschil in hoe de twee beoordelaars de informatie van de informant, gedurende de afname, interpreteren.	Inductief	Bij item 2 van Receptief taalgedrag gaat het erom of de participant woorden of gebaren over activiteiten in de directe context begrijpt. Beoordelaar 1 scoort 'ja', omdat de begeleider aangeeft dat de participant het woord 'chips' begrijpt. Beoordelaar 2 scoort 'nee', omdat de participant volgens haar heeft gezegd dat de participant enkel reageert op het geluid van de zak chips en niet op het woord zelf.
Interpretatie item	Er is, tussen de beoordelaar(s), een verschil in de interpretatie van welk gedrag passend is bij het item. Hieronder vallen ook situaties waarin de beoordeling wordt beïnvloed door persoonlijke verwachtingen van wat iemand 'zou moeten kunnen' of hoe gedrag idealiter zou verlopen. Hieronder vallen ook situaties waarin de beoordeling wordt beïnvloed door persoonlijke verwachtingen van wat iemand 'zou moeten kunnen' of hoe	Inductief	Bij item 6 van <i>Motorisch exploratief gedrag</i> wordt gevraagd of de participant naar een object reikt. Beoordelaar 1 scoort 'ja', omdat de participant diens arm uitstrekt in de richting van het object. Beoordelaar 2 scoort 'nee', omdat zij vindt dat het enkel om een reflexmatige beweging ging en niet om doelgericht reiken.

Code	Toelichting code	Type	Voorbeeld
	gedrag idealiter zou verlopen.		
Onbekend	Er is (g)een (duidelijke) opmerking genoteerd door de beoordelaar(s). Hierdoor is het niet mogelijk om uitspraken te doen over het verschil in scoring.	Inductief	Bij item 14 van <i>Visueel exploratief gedrag</i> scoort beoordelaar 1 'ja' en beoordelaar 2 'nee', maar bij de tweede beoordelaar is geen toelichting gegeven. Hierdoor blijft onduidelijk waarom zij 'nee' scoorde.
Vershil in observatie	Er is door de beoordelaars geschreven over letterlijk verschillende gedragingen van de participant gedurende de afname.	Inductief	Bij item 14 van <i>Visueel exploratief gedrag</i> wordt beoordeeld of de participant naar het gezicht van de testleider kijkt of oogcontact maakt. Beoordelaar 1 scoort 'ja' en vermeldt dat de participant meerdere keren oogcontact maakt. Beoordelaar 2 scoort 'nee', en noemt dat de participant geen één keer oogcontact heeft gemaakt.
Vershil in informatie	Er is een verschil in informatie die de tweede beoordelaar heeft tussen het eerste en tweede scoringsmoment	Inductief	Bij item 8 van <i>Motorisch exploratief gedrag</i> wordt gevraagd of de participant ergens op kan slaan. In het eerste scoringsmoment scoort beoordelaar 2 'ja' op basis van informatie van de begeleiding. Bij de tweede scoring scoort ze 'nee', omdat slaan tijdens de observatie niet werd gezien en de informatie van de informant bij het tweede scoringsmoment ontbreekt.
Vershil in voorwaarde	Er is een verschil in of de beoordelaar(s) het item 'scoorbaar' achtte(n) gedurende de afname	Inductief	Bij item 11 van <i>Motorisch exploratief gedrag</i> wordt gekeken of de participant een voorwerp laat vallen en er dan geïnteresseerd naar kijkt. Beoordelaar 1 scoort 'ja' omdat de participant dit gedrag toont. Beoordelaar 2 laat het item leeg, omdat dit item volgens haar niet is getest gedurende de afname

Noot: De code verschil in voorwaarde is een specifieke code voor de verschillen in scoring tussen de twee scoringsmomenten van beoordelaar 2. Deze code is toegekend wanneer beoordelaar 2 tijdens

het eerste scoringsmoment informatie kreeg via de observant, terwijl die bij het tweede scoringsmoment, dat via videobeelden verliep, ontbrak.