

## Zorgen over de COTAN-beoordelingsprocedure voor het neuropsychologisch testgebruik

Reactie van de auteurs van de COTAPP naar aanleiding van het oordeel van de COTAN, 14 januari 2021

Op 7 januari 2021 publiceerde de COTAN (Commissie Testaangelegenheden Nederland) haar beoordeling van de neuropsychologische test COTAPP (COgnitieve TestAPPLICatie). De betrouwbaarheid en validiteit zijn als onvoldoende beoordeeld. Hoewel de beoordeling zeer teleurstellend was voor ons als auteurs, is dat niet de hoofdreden voor het schrijven van deze reactie.

Al langere tijd worden de nodige kanttekeningen geplaatst bij de geschiktheid van de COTAN-beoordelingscriteria om de kwaliteit van neuropsychologische tests te beoordelen (zie o.a. p93-95, Hendriks et al., 2017; p318-319, Swaab et al., 2017; p99-100, Verbraak et al., 2011). De algemeen geformuleerde criteria van de COTAN houden te weinig rekening met de neuropsychologische praktijk (p100, Verbraak et al., 2011). Een groot aantal tests is als 'onvoldoende' beoordeeld op één of meer essentiële aspecten, terwijl het werkveld deze tests wel als klinische goed bruikbaar acht. Het omgekeerde komt ook voor: een positieve beoordeling door de COTAN op alle punten, terwijl het de vraag is of dit terecht is als rekening zou worden gehouden met de neuropsychologische toepassing (p95, Hendriks et al., 2017). Hierdoor is al geruime tijd de aanbeveling vanuit het werkveld om grotendeels het eigen kritische oordeel te volgen zolang de procedure van de COTAN onvoldoende rekening houdt met de klinisch neuropsychologische toepassingen (p.95, Hendriks et al., 2017).

De COTAN schrijft op haar [website](#): *'De COTAN Documentatie is bedoeld voor testgebruikers ter ondersteuning bij het maken van een objectieve en verantwoorde keuze voor een geschikt (psychodiagnostisch) instrument, maar ook voor mensen die getest zijn en meer willen weten over de kwaliteit van het instrument dat zij hebben ingevuld.'* Een 'onvoldoende' zou volgens de COTAN niet betekenen dat een test onbruikbaar is. Dergelijke tests kunnen volgens de COTAN toch nuttige informatie opleveren. Het oordeel 'onvoldoende' voor de deskundige testgebruiker zou vooral de functie moeten hebben van een waarschuwingssignaal om zelf de geschiktheid voor het toepassingsdoel te beoordelen.

De realiteit is echter dat veruit de meeste testgebruikers een 'onvoldoende' niet als waarschuwingssignaal interpreteren maar als 'niet gebruiken' en de vaak technische en complexe toelichting van de COTAN niet tot nauwelijks zullen lezen. De meeste testgebruikers ontbreekt het hiervoor simpelweg aan tijd, naast dat de toelichtingen een dusdanig technisch niveau hebben dat zij voor de gemiddelde gebruiker niet leesbaar zijn. Zij vertrouwen er (terecht) op dat het COTAN-oordeel houvast geeft in de testkeuze, omdat dat is wat de COTAN zelf ook expliciet als doel formuleert (zie boven). Van 'mensen die getest zijn' kan al helemaal niet verwacht worden dat ze tot een dergelijke verdieping overgaan en dit kan in de praktijk tot vervelende situaties leiden wanneer de clinicus verantwoording moet afleggen aan de geteste vanwege het gebruik van 'onvoldoende' geschikt bevonden tests. De COTAN-beoordeling is in de praktijk daarmee heel vaak beslissend in of een test wel of niet gebruikt wordt. Doordat zo weinig neuropsychologische tests de COTAN-proeve goed doorstaan, ofwel helemaal niet ter beoordeling worden voorgelegd vanwege de reële kans op een negatieve beoordeling, ontstaat de ongewenste situatie dat testgebruikers weinig houvast hebben. Hierdoor blijven tests gebruikt worden die sterk verouderd, matig valide en/of betrouwbaar zijn en blijven tests ongebruikt die veel beter geschikt zijn voor klinisch gebruik. Ontwikkelingen op neuropsychologisch testvlak worden eveneens geremd doordat uitgevers terughoudend zijn om te investeren in het ontwikkelen van toekomstig instrumentarium anders dan 'veilige' investeringen, zoals vragenlijsten of een Nederlandse bewerking van een intelligentie-testbatterij. De neuropsychologische praktijk blijft dan met lege handen staan, terwijl de wetenschappelijke kennis van neuropsychologische diagnostiek zo veel verder is.

Wij vinden het zorgelijk dat de COTAN zich niet bewust lijkt van deze forse negatieve impact van hun beoordelingspraktijk op het neuropsychologisch werkveld, ondanks dat dit signaal de COTAN toch herhaaldelijk ter ore moet zijn gekomen gezien de vele kanttekeningen over de COTAN in handboeken voor klinische professionals (Hendriks et al., 2017; Swaab et al., 2017; Verbraak et al., 2011). Wij hebben geprobeerd met de COTAN hierover in gesprek te komen, maar hier was de COTAN niet toe bereid omdat dit niet volgens de procedure zou zijn. Dit brengt ons bij ons belangrijkste punt: de beoordelingsprocedure is opgesteld door de COTAN én wordt uitgevoerd door de COTAN, waarbij inspraak van testauteurs tot het minimum beperkt wordt. In feite heeft de COTAN hiermee de rol van zowel 'wetgevende' als 'rechtsprekende' macht, machten die normaliter gescheiden zijn en elkaar controleren om tot een evenwichtige oordeelsvorming te komen. Naast de ervaring dat de beoordelingsmatrix die de COTAN hanteert zo matig bleek te passen op het daadwerkelijk klinisch gebruik van de COTAPP (zie toelichting onder), was het voor ons als testauteurs zeer wrang om te moeten constateren dat hele delen van de psychometrische onderbouwing van de COTAPP (bijna gehele sectie Validiteit, grote delen van de nieuw aangeleverde resultaten voor Betrouwbaarheid en representativiteitsanalyses van de Normen) niet zijn meegenomen in de definitieve beoordeling om voor ons niet goed te begrijpen redenen. Verontrustend was dat enige vorm van dialoog hierover niet mogelijk bleek: de definitieve beoordeling was al online geplaatst.

We hebben daarom gekozen voor deze openbare reactie. We hopen hiermee te bereiken dat -zoals de aanbeveling in het werkveld al was- gebruikers van neuropsychologische tests het eigen kritische oordeel blijven volgen zolang de procedure van de COTAN onvoldoende rekening houdt met de klinisch neuropsychologische toepassingen (p.95, Hendriks et al., 2017). Ook hopen we in Bijlage 1 duidelijk te maken hoe het oordeel van de COTAN t.a.v. de COTAPP tot stand is gekomen en waarom wij menen dat de Betrouwbaarheid en Validiteit als minimaal 'voldoende' mogen worden beschouwd.

We beseffen dat ons verweten kan worden niet objectief te zijn als zijnde auteurs van de COTAPP. Daarom hechten wij eraan te **benadrukken dat wij geen enkel financieel of wetenschappelijk belang hebben bij de COTAPP**. De COTAPP is in de afgelopen jaren door de auteurs in eigen tijd ontwikkeld zonder financiële vergoeding. De reden voor deze inzet: intrinsieke motivatie om bij willen dragen aan de klinische praktijk middels het ontwikkelen van nieuwe generatie verfijndere en efficiëntere neuropsychologische tests.

We zullen gebruikers van de COTAPP binnenkort informeren wanneer we enkele kleine verbeteringen hebben aangebracht in de COTAPP n.a.v. de aanvullende analyses die we ten behoeve van de COTAN-beoordeling hebben uitgevoerd. De aanvullende analyses die t.b.v. de COTAN beoordeling zijn aangeleverd bij de COTAN, zijn via deze [link](#) te downloaden.

prof. dr. Nanda Rommelse, GZ-psycholoog i.o.t. klinisch neuropsycholoog (BIG 99922055325)  
dr. Catharina Hartman, universitair hoofddocent  
drs. Annemieke Brinkman, GZ-psycholoog (BIG 09928678725)  
dr. Dorine Slaats-Willemse, klinisch neuropsycholoog (BIG 19052218825)  
dr. Patrick de Zeeuw, GZ-psycholoog i.o.t. klinisch psycholoog (BIG 39920594125)  
dr. Marjolein Luman, universitair hoofddocent

## **Bijlage 1: Illustratie van de beperkte geschiktheid van de COTAN-beoordelingscriteria om de kwaliteit van neuropsychologische tests te beoordelen**

**De COTAN-beoordelingsmatrix om de kwaliteit van een instrument te beoordelen blijkt ver af te staan van het daadwerkelijk gebruik van neuropsychologische tests in de klinische praktijk.**

Zoals uitgebreid in de handleiding is beschreven en in lijn is met de hoeksteen van neuropsychologische diagnostiek, is de COTAPP ontworpen om een cognitief sterkte-zwakke profiel te verkrijgen. De gebruiker wordt middels het '[Stappenplan interpretatie Basisrapportage](#)' handvatten geboden hoe een sterkte-zwakke cognitief profiel beschreven kan worden en richting kan geven aan handelingsadviezen. Op geen enkele manier suggereren wij in dit stappenplan -of in de handleiding- dat een interpretatie op het niveau van losse parameters van enige waarde is. Ter verduidelijking van dit principe gebruiken we hier een foto als metafoor. We willen liefst 'het hele kind' op de foto (breed cognitief profiel i.p.v. één cognitief domein) om zo goed mogelijk het kind te kunnen beschrijven aan de hand van typerende kenmerken (validiteit). Dit moeten we met voldoende scherpheid doen (betrouwbaarheid), rekening houdend met de praktische realiteit en toepasbaarheid (i.e. het neuropsychologisch onderzoek moet plaatsvinden binnen een aanvaardbare tijd). Uit deze metafoor blijkt hoezeer de psychometrische eigenschappen van een test onderling samenhangen én tevens afhankelijk zijn van klinisch gebruik. In de COTAN-beoordelingsmatrix wordt dit aspect, de psychometrie in relatie tot de manier waarop een test klinisch wordt ingezet, in het geheel niet meegewogen. Ieder psychometrisch aspect wordt losstaand beoordeeld waarbij het geheel er niet toe doet. Oftewel, in termen van de foto-metafoor: als we neuropsychologische test-keuze baseren op basis van de COTAN-beoordelingsmatrix, zouden we in de klinisch neuropsychologische praktijk losse foto's maken van telkens een stukje van het kind en daar ruim de tijd voor nemen, zodat we van ieder stukje een heel scherp beeld hebben. Of de stukjes vervolgens aan elkaar passen, enigszins het geheel dekken, en of het maken van zoveel losse foto's op deze scherpheid realistisch is gegeven de tijd die dat kost, wordt in het geheel niet meegewogen. Het behoeft geen verdere uitleg dat deze werkwijze van ver af staat van de praktijk en daarmee voorbij gaat aan een cruciaal aspect van testbeoordeling: de (toegevoegde) waarde voor de klinisch neuropsychologische praktijk (DePascale & Betebenner, 2015).

Binnen de COTAPP hebben we het verkrijgen van een klinisch bruikbaar cognitief sterkte-zwakke profiel in de praktijk voor testgebruikers op meerdere manieren willen faciliteren. De taak duurt relatief kort (gemiddeld 35 minuten), is makkelijk af te nemen en self-paced, waardoor een kind nooit onder- of overvraagd wordt qua werktempo en scores daardoor niet vertekend worden. De taak bestaat uit meerdere onderdelen die op veel (maar niet alle) aspecten overeenkomsten vertonen. Door het berekenen van verschilcores tussen onderdelen én overstijgende parameters krijgt de testgebruiker inzicht in respectievelijk specifieke cognitieve processen onafhankelijk van generieke cognitieve processen als werktempo en – accuratesse. Dit is klinisch van grote meerwaarde, omdat specifieke cognitieve processen vaak niet goed beoordeeld kunnen worden met bestaande taken door de dominante invloed van generieke processen, terwijl beoordeling van specifieke cognitieve processen voor de diagnostische conclusie en behandelindicatie de doorslag kunnen geven. De Basisrapportage bevat parameters die zijn geselecteerd op basis van betrouwbaarheid, validiteit, en klinische bruikbaarheid, met de insteek een zo breed mogelijk cognitief profiel te willen bemeten. Ter illustratie: enkele (voornamelijk verschilscore-)parameters in de Basisrapportage hebben een betrouwbaarheid van rond de  $r=.60$  (zie Bijlage 2, beneden), te laag volgens de COTAN, maar zijn vanwege positieve validiteitsresultaten en klinische bruikbaarheid toch opgenomen. Eén parameter heeft een lage betrouwbaarheid (.26), maar dit bleek te wijten aan het niet parallel zijn van de testhelften voor specifiek deze parameter. Vanwege een goede test-hertest betrouwbaarheid ( $r=.75$ ), validiteit en klinische bruikbaarheid is deze parameter toch opgenomen. In deze weloverwogen weging zijn we voortdurend uitgegaan van de aanname dat testgebruikers op basis van het gehele profiel van scores een conclusie trekken, omdat dit immers de gouden standaard is binnen het klinisch neuropsychologisch werkveld (Swaab et al., 2017; Hendriks et al., 2017). In termen van de foto-

metafoor: ondanks dat de afbeelding hier en daar minder scherp is, heeft de testgebruiker in afzienbare tijd nagenoeg het hele kind op de foto.

Wij menen dat bovenstaande goed aansluit bij de behoeftes van testgebruikers en bijdraagt aan verdere ontwikkeling van de klinisch neuropsychologische praktijk. Wanneer we echter primair een goede COTAN-beoordeling zouden hebben nagestreefd, zou de COTAPP er aanmerkelijk anders uit hebben gezien. Uit de Toelichting van de COTAN blijkt namelijk dat alle 22 parameters uit de Basisrapportage losstaand betrouwbaar moeten zijn ( $r \geq .70$ ). Verder is de verwachting dat dit ook het geval is voor betrouwbaarheidsanalyses uitgesplitst voor leeftijdsgroepen (in dit geval 3 x 22 parameters). Daar bovenop blijkt het nodig dat er voor alle 22 parameters specifieke hypothesen geformuleerd worden over de samenhang met relevante gedragsmaten. Dit zou alleen al resulteren in ruim 150 'specifieke' voorspellingen (22 x 8 overkoepelende domeinen) wanneer we het aantal gedragsuitkomsten indikken tot domeinen ('schoolprestaties' i.p.v. specifieke deeldomeinen apart; 'aandachtsproblemen' i.p.v. specifiek volgens ouder en leerkracht apart). Als we deze én-én-én exercitie van de COTAN zouden hebben doorgevoerd, zou de Basisrapportage bestaan uit alleen nog een aantal generieke parameters en daarmee klinische van weinig waarde zijn: we hebben in dat geval een heel scherpe foto van een klein stukje van het kind. Verschilscores (om specifieke processen te meten) zijn namelijk per definitie minder betrouwbaar dan enkelvoudige scores, omdat de meetonzekerheid van twee parameters i.p.v. één meeweegt. Misschien hadden we nog enkele parameters kunnen "redden" door de taak aanmerkelijk langer te maken: een groter aantal trials leidt doorgaans tot een hogere betrouwbaarheid. De vraag is echter of dit de validiteit wel ten goede zou zijn gekomen, omdat bij een te lange taakduur er vooral nog vermoeidheid wordt gemeten. Daarnaast zijn de kosten van lang testen te hoog, waardoor de praktische bruikbaarheid verder onder druk komt te staan.

### **Een test = een test = een test? Is de COTAN-beoordelingsmatrix wel geschikt voor elk type psychodiagnostisch instrument?**

Bij neuropsychologische tests zoals de COTAPP wordt een sterk beroep gedaan op concentratie, reactiesnelheid en geheugen. Factoren als vermoeidheid, motivatie, stress etc. (oftewel 'state' factoren) zorgen voor fluctuaties in de te meten cognitieve processen binnen een -voor dat individueel stabiele range van capaciteiten (stabiele 'trait' factoren). Vragenlijsten (ook met de meetpretentie cognitief functioneren in kaart te brengen, zoals de BRIEF) meten nagenoeg alleen 'trait' factoren en zijn daarmee veel eenvoudiger boven de vereiste betrouwbaarheidsgrens van  $r \geq .70$  te krijgen. Ook met neuropsychologische tests die geen beroep doen op reactiesnelheid en/of langdurig zelfstandige aandachtsregulatie, zoals taken binnen intelligentie-testbatterijen, worden voornamelijk 'trait' factoren gemeten (Danner et al., 2011). Desalniettemin hanteert de COTAN één beoordelingsmatrix voor al het psychodiagnostisch instrumentarium, met als gevolg dat de beoordeling van neuropsychologische tests veel vaker negatief uitvalt in vergelijking tot andersoortige tests.

Hierdoor ontstaat vervolgens ook begripsverwarring: zo schrijft de COTAN in hun toelichting aan de auteurs dat de BRIEF beschouwd kan worden als een soortgenoot van de COTAPP. Dit is niet in lijn met de opvatting hierover in het neuropsychologische werkveld, waar consensus is dat deze twee instrumenten verschillende zaken meten; d.w.z. een globale subjectieve beoordeling van executieve functies door een informant (de BRIEF, ouders, leerkracht, of zelfbeoordeling) c.q. een objectieve, directe meting van specifieke cognitieve functies bij het kind zelf (neuropsychologische taak; Boyer et al., 2017; Faridi et al., 2015; Fuermaier et al., 2015; Toplak et al., 2013; Swaab et al., 2017). De COTAN blijft echter bij de veronderstelling dat prestatietests en vragenlijsten soortgenoten zijn (wanneer zij een gelijksoortige meetpretentie hebben), ondanks onze onderbouwing in de handleiding (p.15) en onze aanvullende uitleg aan de COTAN. Doordat we in de analyses niet de soortgenootvaliditeit van de COTAPP met de BRIEF hebben onderzocht, maar daarentegen aantonen dat de COTAPP een betere voorspeller is dan de BRIEF voor meerdere klinisch zeer relevante criteriummaten (o.a. volgen van

speciaal onderwijs, vroege start groep 3, meerdere LVS domeinen) wordt 'niet relevant' bevonden door de COTAN en daarom buiten beschouwing gelaten bij de beoordeling van de Validiteit.

Er blijkt ook verwarring over de toepassing in de praktijk van neuropsychologische tests. Ondanks onze onderbouwde bezwaren, heeft de COTAN haar oordeel gebaseerd op een beoordeling van de losse parameters. De COTAN geeft hierbij aan dat het niet duidelijk is bij welke scores er sprake is van onderpresteren en/of een ontwikkelingsstoornis. Echter, zoals in de vooropleidingen van neuropsychologische testgebruikers gedoceerd wordt, werkt neuropsychologisch testonderzoek niet volgens het model waarin een individuele score "X" één op één samenhangt met een conclusie "onderprestatie"/"stoornis Y". Door onbekendheid met het daadwerkelijk gebruik van neuropsychologische tests in de praktijk, is de vraag hoe relevant de uitkomst van de COTAN-beoordeling voor testgebruikers is.

### Referenties

- Boyer BE, Geurts HM, Van der Oord SJ (2017). Planning skills of adolescents with ADHD. *Atten Disord.* 22(1):46-57
- Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, 39,323–334
- DePascale CA & Betebenner D (2015). Utility vis-à-vis Validity. <https://www.nciea.org/sites/default/files/pubs-tmp/Utility%20vis-a-vis%20Validity.pdf>
- Faridi N, Karama S, Burgaleta M, White MT, Evans AC, Fonov V, Collins DL, Waber DP (2015). Neuroanatomical correlates of behavioral rating versus performance measures of working memory in typically developing children and adolescents. *Neuropsychology* 29(1):82-91.
- Farnsworth JL 2nd, Dargo L, Ragan BG, Kang M. (2017). Reliability of Computerized Neurocognitive Tests for Concussion Assessment: A Meta-Analysis. *J Athl Train*, Sep;52(9):826-833.
- Fuermaier AB, Tucha L, Koerts J, Aschenbrenner S, Kaunzinger I, Hauser J, Weisbrod M, Lange KW, Tucha O (2015). Cognitive impairment in adult ADHD-perspective matters! *Neuropsychology* 29(1):45-58.
- Gregory RJ (2011). Concepts of Reliability. In RJ Gregory, *Psychological testing: History, principles and applications* (6th ed.) (pp. 87-108). Pearson Education, Boston.
- Hendriks, M., Kessels, R., Gorissen, M., Schmand, B., Duits, A. (red.) (2017). *Neuropsychologische diagnostiek. De klinische praktijk*. Boom: Amsterdam.
- Koo, T.K. & Li, M.Y. (2006). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*;15(2):155-63.
- Swaab, H., Bouma, A., Hendriksen, J., Konig, C. (red.) (2017). *Klinische Kinderneuropsychologie* (Editie 3). Boom: Amsterdam.
- Toplak ME, West RF, Stanovich KE (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *J Child Psychol Psychiatry* 54:131-43.
- Verbraak, M., Visser, S., Muris, P., Hoogduin, K. (red.) (2011). *Handboek voor GZ-psychologen* (Editie 1). Boom: Amsterdam.

## Bijlage 2 Betrouwbaarheidscoëfficiënten COTAPP na herberekening

De betrouwbaarheid van een meetinstrument geeft aan in hoeverre een score gelijk blijft als een kind onder dezelfde omstandigheden opnieuw getest zou worden. De Pearson correlatie zoals gepresenteerd in de handleiding is niet de beste manier om de betrouwbaarheid uit te drukken. Hoewel van oudsher veel gebruikt voor de betrouwbaarheid, geeft de Pearson correlatie alleen de samenhang tussen de twee testscores weer, maar niet de overeenstemming (Koo & Li, 2006). Wij hebben de betrouwbaarheid van de COTAPP daarom herberekend op basis van een maat die dat wel doet, de Intra Class Correlatie Coëfficiënt (ICC; de zgn. 2-way mixed-effects model variant, Koo & Li, 2006, p.158). Met deze ICC zijn de split-half coëfficiënten en test-hertest (4 tot 6 weken interval) coëfficiënten berekend voor de normgroep. Voor 14 van de 20 uitkomstparameters van de Basisrapportage (split-half coëfficiënt kan niet voor de parameter 'Leersnelheid' berekend worden, 'Responsinhibitie' zal binnenkort worden verwijderd vanwege een te lage betrouwbaarheid, vandaar 20 ipv 22), zijn de split-half coëfficiënten minimaal .70, en veelal boven de .80. Voor 5 van de 6 overige parameters liggen deze rond de .60. Voor 1 parameter (fouten zeer complex werkgeheugen) is de split-half betrouwbaarheid lager (.26), zeer waarschijnlijk samenhangend met het gebrek aan parallelie voor testhelften voor specifiek deze parameter waardoor de uitkomst van de split-half ICC waarschijnlijk niet goed geschat kan worden. Ondersteuning voor deze hypothese volgt vanuit de aanmerkelijk hogere test-hertest betrouwbaarheid voor deze parameter (.75). De 6 parameters met een iets lagere betrouwbaarheid laten verder zonder uitzondering goede resultaten voor validiteit zien en zijn om die reden in de Basisrapportage opgenomen. De aanpassingen in coëfficiënten (zie onderstaande tabel) geven ons inziens aanleiding de Betrouwbaarheid als minimaal voldoende te beoordelen.

	Split-half betrouwbaarheid	Test-hertest betrouwbaarheid
<b>Verwerkingssnelheid</b>		
Reactiesnelheid	.99	.92
Variabiliteit in reactiesnelheid	.89	.76
Reactiesnelheid ten gevolge van cognitieve	.89	.79
Versnellingscapaciteit in reactiesnelheid	.64	.37
Variabiliteit in beslistijd	.80	.71
<b>Aandacht</b>		
Volgehouden aandacht	.87	.36
Arousalregulatie	.60	.62
Onoplettendheid	.59	.66
<b>Executieve controle</b>		
Slordigheid	.99	.52
Responsinhibitie	verwijderd	verwijderd
Interferentiegevoeligheid	.86	.57
Wachttijdafweging	.85	.76
<b>(Werk)geheugen</b>		
Reactiesnelheid eenvoudig	.84	.77
Fouten eenvoudig	.63	.41
Reactiesnelheid complex	.87	.85
Fouten complex	.61	.37
Reactiesnelheid zeer complex	.88	.70
Fouten zeer complex	.26 <sup>1</sup>	.75 <sup>1</sup>
<b>Leersnelheid</b>		
Leersnelheid	Nvt <sup>2</sup>	Nvt <sup>3</sup>
<b>Gedrag tijdens de taak</b>		
Benodigde ondersteuning	.97 <sup>4</sup>	.57
Verbaliseren	.70 <sup>4</sup>	.93
Beweeglijkheid	.74 <sup>4</sup>	.67

Het kleurenspectrum geeft de sterkte van de betrouwbaarheidscoëfficiënt aan, hoe donkerder groen, hoe hoger de coëfficiënt ( $\sim .60$ ;  $\geq .70$ ;  $\geq .80$ ).

<sup>1</sup> split-half coëfficiënt is niet te interpreteren, omdat voor specifiek deze parameter de testhelften niet parallel waren. De test-hertest coëfficiënt is daarom indicatief voor de betrouwbaarheid.

<sup>2</sup> parameter kan niet berekend worden voor twee testhelften gesplitst voor even en oneven genummerde trials.

<sup>3</sup> Test-hertest voor Leersnelheid kan niet beoordeeld worden met een kort test-hertest interval (4-6 weken) omdat kinderen bij hertest de instructie nog kenden van de eerste afname.

<sup>4</sup> Interbeoordelaarsbetrouwbaarheid.