

De Test-Hertestbetrouwbaarheid van Instrumenten voor Executieve Functies bij Kleuters

Bachelorthesis - 200600042

Bacheloropleiding Pedagogische Wetenschappen

Universiteit Utrecht

M. T. van Diest (4118464)

M. van Slobbe (4175476)

Thesisbegeleider: M. W. Röttger

Datum: 25-06-2017

Abstract

The preschool period is a period of significant advances in the development of executive functions. Reliable measures are essential for evaluating intra- and inter group differences in the development of executive functions. However little is known about the psychometric qualities, especially about the test-retest reliability, of measures for executive functions in preschool children. This study examined the test-retest reliability of four executive function tasks (Corsi-blocktest, DCCS, updatingtest and HTKS) in preschool children. The influence of age on test-retest reliability and possible practice effects were evaluated. The four tests were administered to 40 preschool children between the ages of four and six on two testing occasions approximately one week apart. The results show that the tests differ in terms of test-retest reliability, with an inconsistent pattern concerning the different aspects of executive function. Age does not have an influence on the test-retest reliability. The test-retest reliability was high when the instruments were being interpreted as a task battery. Significant practice effects were limited to 5-year olds on the updatingtest and HTKS. These findings contribute to the future establishment of reliable measures for preschool children. Implications of the current study are discussed and recommendations for further research are given.

Keywords: test-retest reliability, executive functions, preschoolers, practice effects

De Test-Hertestbetrouwbaarheid van Instrumenten voor Executieve Functies bij Kleuters

In de laatste decennia is er interesse ontstaan in de ontwikkeling van executieve functies (EF) bij jonge kinderen (Best, Miller, & Jones, 2009). Inzichten op neurobiologisch gebied laten zien dat EF begint te ontwikkelen rond het einde van het eerste levensjaar (Hughes, 1998; Garon, Bryson, & Smith, 2008). De kleutertijd, voornamelijk tussen driejarige en zesjarige leeftijd, is een periode waarin EF sterk ontwikkelt (Carlson, 2005; Diamond, 2001; Rothbart & Posner, 2001). De ontwikkeling van EF vormt een belangrijke basis voor hogere cognitieve processen in de volwassenheid (Garon et al, 2008). Afgelopen jaren zijn verschillende EF-instrumenten ontwikkeld voor kleuters (Carlson, 2005; Garon, et al., 2008; Zelazo & Müller, 2010). Er is echter weinig onderzoek gedaan naar de test-hertestbetrouwbaarheid van EF-instrumenten voor kleuters (Müller, Kerns, & Konkin, 2012). Betrouwbare instrumenten zijn essentieel voor onderzoek naar oorzaken, correlaties en consequenties van individuele- en groepsverschillen in de ontwikkeling van EF (Beck, Schaefer, Pang, & Carlson, 2011). Het doel van dit onderzoek is om een bijdrage te leveren aan onderzoek naar de betrouwbaarheid, specifiek de test-hertestbetrouwbaarheid, van EF-instrumenten voor kleuters in de leeftijd van vier tot zeven jaar.

EF is een overkoepelende term voor onderling gerelateerde cognitieve vaardigheden die een rol spelen bij de controle en coördinatie van informatie die van

belang is voor doelgericht gedrag (Miller & Cohen, 2001). Volgens het model van Miyake et al. (2000) is EF een multidimensioneel construct bestaande uit drie kerncomponenten. Kerncomponenten van EF zijn werkgeheugen, cognitieve flexibiliteit en inhibitiecontrole. Werkgeheugen wordt gedefinieerd als het in gedachten houden en 'updaten' van informatie tijdens het uitvoeren van een bewerking op die informatie. Cognitieve flexibiliteit betreft het vermogen om snel en flexibel gedrag aan te kunnen passen aan veranderende situaties (Davidson, Amso, Anderson, & Diamond, 2006) en inhibitiecontrole verwijst naar de mogelijkheid om een automatische respons te onderdrukken om een meer geschikte reactie in een bepaalde situatie uit te kunnen voeren (Diamond, 2013).

De test-hertestbetrouwbaarheid geeft aan in hoeverre het instrument consistent hetzelfde construct meet over tijd en geeft daarmee aan hoeveel variatie het gevolg is van de invloed van dagelijks wisselende omstandigheden, zoals humeur of weersomstandigheden (Anastasi & Urbina, 1997; Gray, 2003). Voldoende test-hertestbetrouwbaarheid verzekert dat de prestatie op een taak geen gevolg is van errorvariantie, wat een voorwaarde is voor het valide kunnen meten van een construct. Ondanks dat een test-hertest design bepaalde limitaties kent, zoals het mogelijk optreden van leereffecten, geeft het belangrijke inzichten over in hoeverre de prestatie op een test een stabiele weergave is van de huidige capaciteiten (Willoughby & Blair, 2011).

De 'Hoofd-Tenen-Knieën-Schouders' (HTKS) is ontwikkeld als maat voor voornamelijk inhibitie, maar ook als maat voor werkgeheugen en gerichte aandacht (Ponitz et al., 2008). In deze taak moeten kinderen het tegenovergestelde doen van wat de testleider zegt. Kinderen moeten bijvoorbeeld de tenen aanraken wanneer de testleider zegt dat zij het hoofd aan moeten raken en andersom. Er is geen informatie bekend over de test-hertestbetrouwbaarheid van de HTKS, daarom zal worden gekeken naar de test-hertestbetrouwbaarheid van andere inhibitietesten voor kleuters.

De HTKS komt overeen met go/no-go testen omdat de HTKS voornamelijk responsinhibitie meet. Bij responsinhibitie dient een natuurlijke/dominante respons geremd te worden om correct te reageren (Rubia et al., 2001). Uit onderzoek blijkt dat de test-hertestbetrouwbaarheid van go/no-go testen varieert van onvoldoende tot goed (Floyd & Kirby, 2001; Müller et al., 2012; Nampijja et al., 2010). Vier van de zes testen uit de onderzoeken van Floyd en Kirby (2001), Müller en collega's (2012) en Nampijja en collega's (2010) hebben minimaal voldoende test-hertestbetrouwbaarheid. Onderzoek naar het optreden van leereffecten op go/no-go testen laat zien dat er geen sprake is van significante leereffecten (Müller et al., 2012; Nampijja et al., 2010). Een kanttekening bij het onderzoek van Müller en collega's (2012) is dat zij gebruik hebben gemaakt van een gecomputeriseerde test waardoor de resultaten van hun onderzoek met voorzichtigheid

gegeneraliseerd moeten worden naar onderzoeken met taken die niet gecomputeriseerd zijn. Ook de resultaten van Nampijja en collega's (2010) zijn beperkt in de zin dat de instrumenten in hun onderzoek vertaald en aangepast zijn voor afname bij Ugandese kleuters.

Een test die wereldwijd ingezet wordt in onderzoek naar cognitieve flexibiliteit in de vroege kindertijd betreft de DCCS (Frye, Zelazo, & Palfai, 1995; Zelazo, 2006). In deze test moeten kaarten gesorteerd worden op basis van een bepaalde dimensie (bijvoorbeeld kleur) en daarna op basis van een andere dimensie (bijvoorbeeld vorm). Onderzoek naar de test-hertestbetrouwbaarheid van de DCCS, met een tijdsinterval van 15 minuten tussen beide testmomenten, laat zien dat de test-hertestbetrouwbaarheid van voldoende kwaliteit is (Beck et al., 2011). De conclusies van dit onderzoek zijn echter beperkt tot metingen die op dezelfde dag uitgevoerd worden. Ander onderzoek naar de test-hertestbetrouwbaarheid van de DCCS, waarbij ongeveer drie weken tussen de testmomenten zat, laat daarentegen zien dat de test-hertestbetrouwbaarheid van de DCCS zwak is (Müller et al., 2012). Er is slechts één onderzoek gevonden naar het optreden van leereffecten op de DCCS (Müller et al., 2012). In dit onderzoek zijn geen significante leereffecten gevonden.

Een veelgebruikt instrument voor het meten van het werkgeheugen is de Corsi-blokkentest. In het eerste deel van de test, De Forward Corsi, moeten kinderen een serie blokken aanwijzen die de testleider heeft voorgedaan. In het tweede deel van de test, de Backward Corsi, wordt gevraagd om het patroon in omgekeerde volgorde aan te wijzen (Kessels, van den Berg, Ruis, & Brands, 2008). Een recent ontwikkeld instrument, de updatingtest (Wijnroks), richt zich op een specifieke functie van het werkgeheugen, namelijk updating. Updating is het coderen en evalueren van binnenkomende informatie aan de relevantie van de taak die op dat moment uitgevoerd wordt. Bij updating wordt niet alleen passief informatie vastgehouden in het werkgeheugen, maar wordt de informatie ook actief bewerkt (Baddeley, 1996; Miyake et al., 2000).

Er is nog geen informatie bekend over de test-hertestbetrouwbaarheid van de Corsi-blokkentest en de recent ontwikkelde updatingtest. Het is interessant om te kijken naar de test-hertestbetrouwbaarheid van andere werkgeheugen instrumenten voor kleuters omdat deze testen overlap in structuur vertonen met de Corsi-blokkentest. Onderzoek naar werkgeheugen instrumenten voor kleuters laten over het algemeen een goede test-hertestbetrouwbaarheid zien (Nampijja et al., 2010; Müller et al., 2012). Echter moet dit gegeven voorzichtig geïnterpreteerd worden omdat het werkgeheugen uit meerdere aspecten bestaat (Baddeley, 1996). Het is mogelijk dat verschillende werkgeheugen instrumenten een ander aspect van werkgeheugen meten dan de Corsi-blokkentest en de updatingtest. Onderzoek naar het optreden van leereffecten bij werkgeheugentaken laat inconsistente resultaten zien, een mogelijk gevolg van het

gebruik van verschillende tijdsintervallen tussen de testmomenten (Beck et al., 2011). Zo hebben Holmes en collega's (2009) geen leereffecten gevonden op werkgeheugentesten bij een interval van zes weken, waar ander onderzoek wel leereffecten heeft gevonden bij intervallen van twee tot drie weken (Müller et al., 2012; Kuntsi, Stevenson, Oosterlaan, & Sonuga-Barke, 2001). Een mogelijke verklaring voor het optreden van leereffecten op bepaalde werkgeheugentesten is dat kleuters het concept 'omgekeerd' niet begrijpen (Davis & Pratt, 1995). Een onderdeel als de Backward Corsi zou dan niet begrepen worden tijdens het eerste testmoment, maar door blootstelling aan de test wordt dit concept wel begrepen tijdens het tweede testmoment.

In tegenstelling tot het model van Miyake en anderen (2000), waarbij EF gezien wordt als een multidimensioneel construct bestaande uit drie kerncomponenten, is volgens onderzoek van Wiebe, Espy en Charak (2008) de EF van kleuters ongedifferentieerd. Ander onderzoek naar de structuur van EF bevestigt dat er sprake is van een unidimensioneel construct in de vroege kindertijd (Espy, Sheffield, Wiebe, Clark, & Moehr, 2010). Onderzoek suggereert dat, wanneer er sprake is van een unidimensioneel construct in de vroege kindertijd, een testbatterij de EF-capaciteiten beter zou ondervangen dan individuele testen (Ettenhofer, Hambrick, & Abeles, 2006; Willoughby & Blair, 2011). Naar aanleiding van deze gegevens is het interessant om ook te kijken naar de test-hertestbetrouwbaarheid wanneer de vier EF-instrumenten bekeken worden als testbatterij.

Geconcludeerd kan worden dat het van belang is om meer informatie te verkrijgen over de test-hertestbetrouwbaarheid van EF-instrumenten bij kleuters, zodat onderzoek naar oorzaken, correlaties en consequenties van individuele verschillen in de ontwikkeling van EF mogelijk is (Beck et al., 2011). Hierdoor kan vroegtijdig een beeld gevormd worden van de individuele verschillen tussen kleuters, om eventuele problemen in de ontwikkeling van EF tijdig op te sporen. Om hier een bijdrage aan te leveren staat in dit onderzoek de volgende onderzoeksvraag centraal: Wat is de test-hertestbetrouwbaarheid van de HTKS, DCCS, Corsi-blokkentest en updatintest? Er wordt verwacht dat een goede test-hertestbetrouwbaarheid wordt gevonden voor de twee werkgeheugen testen (Corsi-blokkentest en updatintest) (Müller et al., 2012; Nampijja et al., 2010). Omdat het merendeel van de instrumenten over minimaal voldoende test-hertestbetrouwbaarheid beschikt (Floyd & Kirby, 2001; Müller et al., 2012; Nampijja et al., 2010), is de verwachting dat minimaal een voldoende test-hertestbetrouwbaarheid wordt gevonden voor de HTKS. Er wordt een onvoldoende mate van test-hertestbetrouwbaarheid voor de DCCS verwacht (Müller et al., 2012). Daarnaast wordt de test-hertestbetrouwbaarheid van vier EF-instrumenten als testbatterij onderzocht. Er wordt een goede test-hertestbetrouwbaarheid verwacht voor de testbatterij (Ettenhofer

et al., 2006; Willoughby & Blair, 2011). Ten slotte wordt onderzocht of er sprake is van leereffecten op de vier afzonderlijke instrumenten. Er wordt verwacht dat er geen leereffecten gevonden zullen worden op de HTKS (Müller et al., 2012; Nampijja et al., 2010) en DCCS (Müller et al., 2012). De verwachting is dat er wel leereffecten gevonden zullen worden op de werkgeheugentaken (Kuntsi et al., 2001; Müller et al., 2012).

Methode

Participanten

Huidig onderzoek maakte deel uit van onderzoek naar vier EF-instrumenten waarbij data is verzameld onder 240 kleuters. Bij 40 kleuters werd data verzameld betreffende de test-hertestbetrouwbaarheid. De participanten van huidig onderzoek zijn geworven op vier reguliere basisscholen in verschillende provincies van Nederland door middel van een gemakssteekproef. De scholen en participanten zijn geselecteerd op basis van vrijwilligheid en op basis van bestaande contacten van de onderzoekers (Neuman, 2011). De participanten waren 11 vierjarige kleuters, 18 vijfjarige kleuters en 11 zesjarige kleuters.

Procedure

Voorafgaande aan dit onderzoek is een pilotstudie uitgevoerd om te bepalen welke updatingtest (kleurentest, cijfertest of complexe test) gebruikt zou worden in het daadwerkelijke onderzoek. Toestemming voor deelname van de kleuters aan de pilotstudie werd verkregen via ouders door middel van een informed consent. De pilot is afgenomen onder 44 kleuters. Na het analyseren van de gegevens is uiteindelijk de cijfertest betrokken in het daadwerkelijke onderzoek.

Aan de start van het onderzoek is overleg geweest met de directie en leerkrachten van de scholen, daarbij is de leerkracht om toestemming gevraagd voor participatie in het onderzoek door middel van een informed consent. Ook aan ouders van alle kinderen is toestemming gevraagd voor de participatie van hun kind aan het onderzoek door middel van een informed consent. Er werd toestemming aan de ouders gevraagd om de testen af te nemen bij hun kind (met inbegrip van de SDQ en het delen van de cito-gegevens) en om de resultaten van hun kind door te geven aan de school van het kind. De leerkracht/school mocht de testscores alleen inzien als hiervoor toestemming werd gegeven door ouders. Ook werd in dit informed consent uitgelegd dat alle informatie volledig geanonimiseerd zou worden en dat er op ieder moment gestopt kon worden met deelname aan het onderzoek indien het niet gewenst was verder te gaan.

In het huidige onderzoek waren vier testleiders elk verantwoordelijk voor dataverzameling bij tien kleuters. De kleuters hadden op beide testmomenten dezelfde testleider en werden individueel getest op de eigen school in een aparte ruimte. Elke testsessie duurde ongeveer 30 minuten. De tijd tussen testmoment 1 en testmoment 2 varieerde van minimaal vijf tot maximaal tien dagen. Verschillen hierin zijn ontstaan door

de beschikbaarheid van school/leerkracht en/of door ziekte van een participant. Tijdens elke sessie zijn de instrumenten afgenomen in een vaste volgorde: Corsi-blokkentest, DCCS, updatingtest en HTKS.

Meetinstrumenten

Corsi-blokkentest. In 1972 is door Corsi de Corsi-blokkentest ontwikkeld als instrument voor het werkgeheugen (Kessels et al., 2008). In deze test wordt kinderen gevraagd een serie blokken aan te wijzen die op een onregelmatige manier gerangschikt zijn op een plaat, nadat de testleider het heeft voorgedaan. De Corsi-blokkentest bestaat uit twee delen, bij de Forward Corsi moeten kinderen het patroon nadoen dat de testleider heeft voorgedaan. Bij de Backward Corsi wordt de kinderen gevraagd om het patroon dat de testleider heeft voorgedaan in omgekeerde volgorde na te doen. Prestatie wordt gemeten door het aantal correct uitgevoerde handelingen voor alle trials bij elkaar op te tellen. De betrouwbaarheid van de Corsi-blokkentest is voldoende en ligt tussen .70 en .76 (Orsini, 1994). Huidig onderzoek vond een Cronbach's alpha van .83 (T1) en .84 (T2).

Dimensional Change Card Sort (DCCS). Deze test is ontwikkeld door Zelazo (2006) voor het meten van cognitieve flexibiliteit. Kinderen moeten kaarten sorteren op basis van een bepaalde dimensie (kleur) en daarna op basis van een andere dimensie (vorm). Daarna volgt een meer complexe taak waarbij een derde sorteerregel wordt toegevoegd (wel of geen zwarte rand). In deze fase moeten de participanten van kaart tot kaart switchen tussen de dimensies. Prestatie wordt gemeten door het aantal correct gesorteerde kaarten voor alle trials die zijn afgenomen bij elkaar op te tellen. Er zijn geen gegevens bekend over de validiteit en betrouwbaarheid van deze test. Huidig onderzoek vond een Cronbach's alpha van .51 (T1) en .56 (T2).

Updatingtest. Deze test is recentelijk door Wijnroks ontwikkeld. Bij deze test geldt de volgende procedure: er worden steeds kaartjes laten zien, die worden vervolgens omgekeerd op tafel gelegd, waarbij het kind moet onthouden en benoemen wat er op die kaartjes stond terwijl er kaartjes worden weggehaald en worden toegevoegd. Er zijn nog geen gegevens bekend over de validiteit en betrouwbaarheid van deze test. Huidig onderzoek vond een Cronbach's alpha van .96 (T1) en .95 (T2).

Hoofd-Tenen-Knieën-Schouders (HTKS). Deze test is ontwikkeld door Ponitz et al. (2008) als maat voor voornamelijk inhibitie, maar ook als maat voor werkgeheugen en gerichte aandacht. De kinderen wordt gevraagd het tegenovergestelde te doen van wat de testleider zegt. De testleider vraagt de kinderen om hun hoofd aan te raken (of hun tenen), maar in plaats van de instructie op te volgen, moeten de kinderen het tegengestelde doen en hun tenen aanraken (of hun hoofd). Vervolgens krijgen zij een moeilijker deel waarin de knieën en schouders worden toegevoegd. Prestatie wordt gemeten door het aantal correct uitgevoerde handelingen bij elkaar op te tellen. De

constructvaliditeit van de HTKS is significant gerelateerd aan de drie kerncomponenten van EF: cognitieve flexibiliteit, werkgeheugen en inhibitie (McClelland et al., 2014). Huidig onderzoek vond een Cronbach's alpha van .93 (T1) en .95 (T2).

Wetenschappelijke en maatschappelijke context

Er is weinig informatie bekend over de test-hertestbetrouwbaarheid van EF-instrumenten voor kleuters. Kennis over de test-hertestbetrouwbaarheid is van belang omdat dit aangeeft of een instrument consistent hetzelfde construct meet over tijd. Het geeft daarmee aan in hoeverre de prestaties een weergave zijn van de huidige capaciteiten op dat gebied (Willoughby & Blair, 2011). Test-hertestbetrouwbaarheid biedt daarom belangrijke informatie met betrekking tot de ontwikkeling van EF-instrumenten voor kleuters.

Met kwalitatief goede psychometrische instrumenten voor kleuters is het mogelijk tijdig problemen in de EF te signaleren om in te kunnen grijpen op jonge leeftijd. Dit is van belang vanwege de grote invloed die EF hebben op de verschillende ontwikkelingsdomeinen van kinderen, en de samenhang van verminderde executieve functies met bepaalde ontwikkelingsstoornissen (Hughes, 2011; Zelazo & Müller, 2010).

Data-analyse

Scores op testmoment 1 (T1) en testmoment 2 (T2) zijn de afhankelijke variabelen en zijn van ratio meetniveau. De samenhang tussen T1 en T2 van elk afzonderlijk instrument is bepaald aan de hand van de Pearson product-moment correlatieanalyse. Voor het meten van de test-hertestbetrouwbaarheid geldt dat er tweezijdig is getoetst, omdat de nadruk lag op het verschil in samenhang tussen de scores op T1 en T2, waaraan geen eenduidige richtinggevende hypothese ten grondslag lag.

De data van de Corsi-blokkentest voldeed aan de assumptie van normaliteit op T1 (skewness = .28, kurtosis = -.21) en T2 (skewness = .59, kurtosis = -.46). De data van de DCCS voldeed aan de assumptie van normaliteit op T1 (skewness = .90, kurtosis = .02) en T2 (skewness = .79, kurtosis = -.82). De data van de updatingtest voldeed aan de assumptie van normaliteit op T1 (skewness = -.04, kurtosis = -.35) en T2 (skewness = -.16, kurtosis = -.63). De data van de HTKS voldeed niet aan de assumptie van normaliteit op T1 (skewness = -1.91, kurtosis = 3.83) en T2 (skewness = -3.21, kurtosis = 12.58). De data zijn echter wel geïnterpreteerd als zijnde normaal verdeeld omdat volgens de centrale limietstelling de steekproef van voldoende grootte is om deze bij benadering als normaal verdeeld te behandelen (Field, 2009). Er zijn enkele uitschieters gevonden op de twee testmomenten van de HTKS, echter zijn deze wel meegenomen in de analyse vanwege de relatief kleine steekproefgrootte. Alleen de data van de DCCS voldeed niet aan de assumptie van lineariteit, deze is echter wel als dusdanig behandeld

vanwege de voldoende steekproefgrootte (Field, 2009). Alle data van de vier instrumenten voldeden aan de assumptie van homoscedasticiteit.

Om meer informatie te verkrijgen over de test-hertestbetrouwbaarheid is onderzocht of leeftijd van invloed is op de test-hertestbetrouwbaarheid door het uitvoeren van een partiële-correlatieanalyse waarbij gecontroleerd is voor leeftijd. Om te onderzoeken of er bij de vier testen sprake was leereffecten is een gepaarde t-test uitgevoerd. Deze is voor alle leeftijdscategorieën samen uitgevoerd en daarna uitgesplitst om eventuele leereffecten per leeftijdscategorie te detecteren.

Resultaten

Beschrijvende statistieken

In tabel 1 zijn de beschrijvende statistieken voor de variabelen uit de analyses gepresenteerd.

Tabel 1.

Beschrijvende Statistieken op testmoment 1 (T1) en testmoment 2 (T2)

	N	T1		T2	
		M (SD)	Range	M (SD)	Range
HTKS					
4-jarigen	11	34.27 (14.88)	0 – 48	39.45 (15.06)	0 – 50
5-jarigen	18	41.61 (10.54)	8 – 51	45.44 (6.31)	28 – 52
6-jarigen	11	46.55 (6.68)	33 – 52	49.09 (1.70)	47 – 52
Alle leeftijden	40	40.95 (11.69)	0 – 52	44.80 (9.47)	0 – 52
DCCS					
4-jarigen	11	17.91 (1.04)	16 – 19	20.09 (4.93)	17 – 30
5-jarigen	18	19.61 (2.45)	17 – 24	22.44 (5.34)	16 – 30
6-jarigen	11	19.36 (2.77)	15 – 24	21.45 (4.01)	15 – 29
Alle leeftijden	40	19.08 (2.33)	15 – 24	21.53 (4.87)	15 – 30
Corsi-blokken					
4-jarigen	11	6.65 (2.16)	2 – 11	7.18 (1.33)	5 – 9
5-jarigen	18	9.50 (2.31)	6 – 14	9.44 (2.73)	5 – 14
6-jarigen	11	11.73 (3.10)	6 – 16	12.45 (2.98)	9 – 17
Alle leeftijden	40	9.30 (3.13)	2 – 16	9.65 (3.15)	5 – 17
Updating					
4-jarigen	11	24.36 (10.84)	6 – 38	23.00 (9.10)	5 – 35
5-jarigen	18	22.78 (10.23)	4 – 39	26.83 (8.77)	10 – 43
6-jarigen	11	27.18 (9.28)	16 – 48	27.64 (10.64)	15 – 43
Alle leeftijden	40	24.43 (10.06)	4 – 48	26.00 (9.35)	5 – 43
Testbatterij	40	93.75 (10.69)	24 – 128	101.98 (18.81)	39 – 133

Test-hertestbetrouwbaarheid

De test-hertestbetrouwbaarheid is gemeten, waarbij de samenhang tussen T1 en T2 per taak en voor de testbatterij is onderzocht. Samenhang tussen T1 en T2 is onderzocht door de bivariate Pearson product-moment correlatieanalyse. Hierbij is gekeken naar de samenhang tussen scores op T1 en scores op T2. Voor alle analyses geldt dat getoetst wordt bij $\alpha = .01$. In tabel 2 zijn de correlaties tussen T1 en T2 gepresenteerd van de HTKS, DCCS, Corsi-blokkentest, updatingtest en voor de testbatterij.

Tabel 2.

Correlatiecoëfficiënten (r) en partiële-correlatiecoëfficiënten (ρ)

	r	ρ
HTKS	.77***	.73***
DCCS	.19	.17
Corsi-blokken	.85***	.76***
Updating	.65***	.65***
Testbatterij	.82***	.79***

Noot. N = 40. *** $p < .001$

Uit de analyse is een significante positieve correlatie gevonden tussen T1 en T2 op de HTKS, $r = .77$, $p < .001$. Dit betekent dat de score op T2 sterk te herleiden was uit de score op T1 via een lineaire relatie. De proportie verklaarde variantie was groot, $r^2 = .59$. Dit betekent dat 59.3% van de variantie in de score op T2 kon worden verklaard door de variantie in de score op T1. Er is geen significante correlatie gevonden tussen T1 en T2 op de DCCS, $r = .19$, $p = .24$. Dit betekent dat er sprake was van geen tot een zeer zwakke positieve samenhang tussen T1 en T2. De score op T2 was niet te herleiden uit de score op T1 via een lineaire relatie. De proportie verklaarde variantie was zeer klein, $r^2 = .04$. Dit betekent dat slechts 3.61% van de variantie in de score op T2 kon worden verklaard door de variantie in de score op T1. Tussen T1 en T2 van de Corsi-blokkentest is een significante positieve correlatie gevonden, $r = .85$, $p < .001$. Dit betekent dat de score op T2 sterk te herleiden was uit de score op T1 via een lineaire relatie. De proportie verklaarde variantie was groot, $r^2 = .72$. Dit betekent dat 72,3% van de variantie in op T2 kan worden verklaard door de variantie op T1. Uit de analyse is een significante positieve correlatie tussen T1 en T2 op de updatingtest gevonden, $r = .65$, $p < .001$. De proportie verklaarde variantie was matig, $r^2 = .42$. Dit betekent dat 42,3% van de variantie op T2 kan worden verklaard door de variantie op T1. Voor de testbatterij is een significante positieve correlatie tussen T1 en T2 gevonden, $r = .82$, $p < .001$. Dit betekent dat de score op T2 sterk te herleiden was uit de score op T1 via een lineaire relatie. De proportie verklaarde variantie was groot, $r^2 = .67$. Dit betekent dat 67.3%

van de variantie in de score op T2 kon worden verklaard door de variantie in de score op T1.

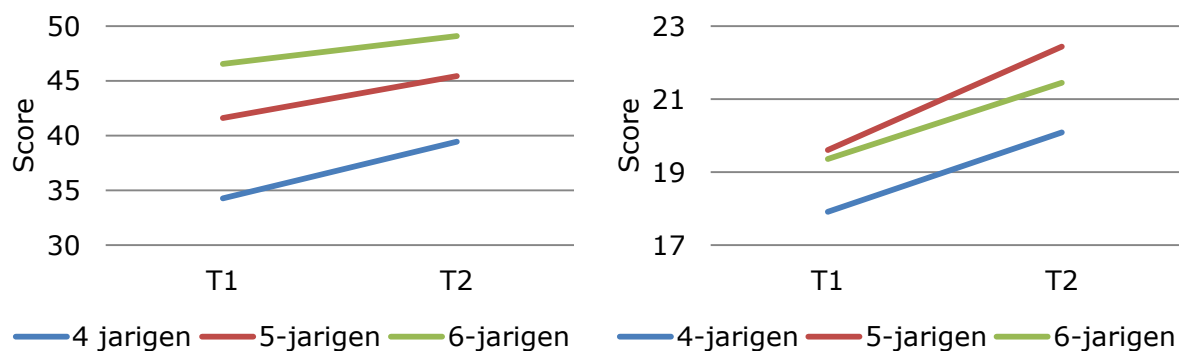
Invloed leeftijd op de test-hertestbetrouwbaarheid

Er is een partiële-correlatieanalyse uitgevoerd om te controleren voor leeftijd. In tabel 2 zijn de correlatiecoëfficiënten van de Pearson product-moment correlatieanalyse en van de partiële-correlatieanalyse naast elkaar weergegeven ter vergelijking.

Uit de analyse is gebleken dat, na het controleren voor leeftijd, de samenhang tussen T1 en T2 van de HTKS significant, positief en sterk is gebleven, $r = .73$, $p < .001$. Het is gebleken dat leeftijd geen invloed heeft op de mate van samenhang tussen T1 en T2 van de HTKS. Uit de analyse is gebleken dat, na het controleren voor leeftijd, de positieve samenhang tussen T1 en T2 van de DCCS niet significant is gebleven, $r = .17$, $p = .30$. De samenhang tussen T1 en T2 is, na controle voor leeftijd, zeer zwak gebleven. Het is gebleken dat leeftijd geen invloed heeft op de mate van samenhang tussen T1 en T2 van de DCCS. Uit de analyse is gebleken dat, na het controleren voor leeftijd, de samenhang tussen T1 en T2 van de Corsi-blokkentest significant, positief en sterk is gebleven, $r = .76$, $p < .001$. Het is gebleken dat leeftijd geen invloed heeft op de mate van samenhang tussen T1 en T2 van de Corsi-blokkentest. Uit de analyse is gebleken dat, na het controleren voor leeftijd, de samenhang tussen T1 en T2 van de updatingtest significant, positief en middelmatig is gebleven, $r = .65$, $p < .001$. Het is gebleken dat leeftijd geen invloed heeft op de mate van samenhang tussen T1 en T2 van de updatingtest. Voor de testbatterij is gebleken dat, na controle voor leeftijd, de samenhang tussen T1 en T2 significant, positief en sterk blijft, $r = .79$, $p < .001$. Het is gebleken dat leeftijd geen invloed heeft op de mate van samenhang tussen T1 en T2 van de testbatterij.

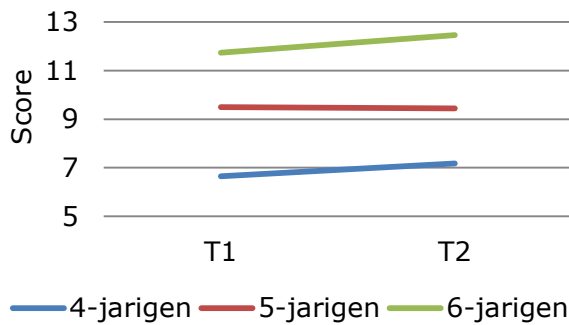
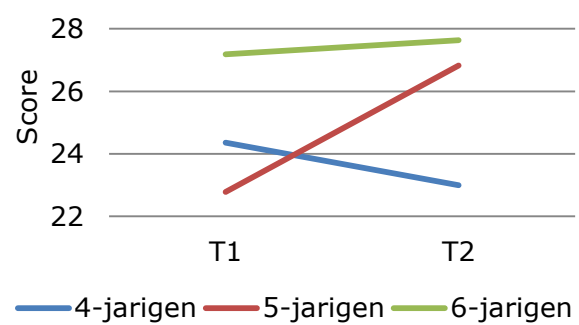
Leereffecten

De mogelijke aanwezigheid van leereffecten is onderzocht aan de hand van een gepaarde t-test. Figuur 1 tot en met 4 geven de gemiddelde scores uitgesplitst naar leeftijd op T1 en T2 weer.



Figuur 1. Scores T1 en T2 HTKS

Figuur 2. Scores T1 en T2 DCCS

**Figuur 3.** Scores T1 en T2 Corsi**Figuur 4.** Scores T1 en T2 Updating

Uit de analyse is gebleken dat kleuters op de HTKS significant hoger scoorden op T2 ($M = 44.80$, $SD = 9.47$) dan op T1 ($M = 40.95$, $SD = 11.76$), $t(39) = -3.24$, $p < .01$, $d = .36$. De effectgrootte was klein. Er was sprake van een leereffect op de HTKS. In figuur 1 zijn de gemiddelde scores weergegeven op de HTKS op T1 en T2 uitgesplitst naar leeftijd in jaren. Wanneer apart naar elke leeftijdsgroep gekeken werd, bleek dat vierjarige kleuters gemiddeld hoger scoorden op T2 ($M = 39.45$, $SD = 15.06$) dan op T1 ($M = 34.27$, $SD = 14.88$). Dit verschil was niet significant: $t(10) = -1.55$, $p = .15$, $d = .35$. Dit was een klein effect. Bij vierjarige kleuters was geen sprake van een leereffect op de HTKS. Vijfjarige kleuters bleken gemiddeld hoger te scoren op T2 ($M = 45.44$, $SD = 6.31$) dan op T1 ($M = 41.61$, $SD = 10.54$). Dit verschil was significant: $t(17) = -2.06$, $p < .01$, $d = .44$. De effectgrootte was klein. Bij vijfjarige kleuters was sprake van een leereffect op de HTKS. Zesjarige kleuters bleken gemiddeld hoger te scoren op T2 ($M = 49.09$, $SD = 1.70$) dan op T1 ($M = 46.55$, $SD = 6.68$). Dit verschil was niet significant: $t(10) = -1.32$, $p = .22$, $d = .52$. De effectgrootte was gemiddeld. Er was bij zesjarige kleuters geen sprake van een leereffect op de HTKS.

Op de DCCS bleken kleuters gemiddeld hoger te scoren op T2 ($M = 21.53$, $SD = 4.87$) dan op T1 ($M = 19.08$, $SD = 2.33$). Dit verschil was significant: $t(39) = -3.11$, $p < .01$, $d = .64$. De effectgrootte was gemiddeld. Er was sprake van een leereffect op de DCCS. In figuur 2 zijn de gemiddelde scores weergegeven op de DCCS op T1 en T2 uitgesplitst naar leeftijd in jaren. Wanneer apart naar elke leeftijdsgroep gekeken werd, bleek dat vierjarige kleuters gemiddeld hoger scoorden op T2 ($M = 20.09$, $SD = 4.93$) dan op T1 ($M = 17.91$, $SD = 1.04$). Dit verschil was niet significant: $t(10) = -1.63$, $p > .05$, $d = .61$. De effectgrootte was gemiddeld. Er was bij vierjarige kleuters geen sprake van een leereffect op de DCCS. Vijfjarige kleuters scoorden gemiddeld hoger op T2 ($M = 22.44$, $SD = 5.34$) dan op T1 ($M = 19.61$, $SD = 2.45$). Dit verschil was niet significant, $t(17) = -2.06$, $p = .06$, $d = .68$. De effectgrootte was gemiddeld. Bij vijfjarige kleuters was geen sprake van een leereffect op de DCCS. Zesjarige kleuters bleken gemiddeld hoger te scoren op T2 ($M = 21.45$, $SD = 4.01$) dan op T1 ($M = 19.36$, $SD = 2.77$). Dit

verschil was niet significant: $t(10) = -1.60$, $p = .14$, $d = .61$. De effectgrootte was gemiddeld. Er was bij zesjarige kleuters geen sprake van een leereffect op de DCCS.

Op de Corsi-blokkentest bleken kleuters gemiddeld hoger te scoren op T2 ($M = 9.65$, $SD = 3.15$) dan op T1 ($M = 9.30$, $SD = 3.13$). Dit verschil was niet significant: $t(39) = -1.29$, $p = .21$, $d = .11$. De effectgrootte was klein. Er was geen sprake van een leereffect op de Corsi-blokkentest. In figuur 3 zijn de gemiddelde scores weergegeven op de Corsi-blokkentest op T1 en T2 uitgesplitst naar leeftijd in jaren. Wanneer apart naar elke leeftijdsgroep gekeken werd, bleek dat vierjarige kleuters gemiddeld hoger scoorden op T2 ($M = 7.18$, $SD = 1.33$) dan op T1 ($M = 6.65$, $SD = 2.16$). Dit verschil was niet significant: $t(10) = -1.14$, $p = .28$, $d = .30$. De effectgrootte was klein. Er was geen sprake van een leereffect voor vierjarige kleuters op de Corsi-blokkentest. Vijfjarige kleuters scoorden gemiddeld lager op T2 ($M = 9.44$, $SD = 2.73$) dan op T1 ($M = 9.50$, $SD = 2.31$). Dit verschil was niet significant: $t(17) = .16$, $p = .88$, $d = -.02$. Er was geen sprake van een leereffect voor vijfjarige kleuters op de Corsi-blokkentest. Zesjarige kleuters scoorden gemiddeld hoger op T2 ($M = 12.45$, $SD = 2.98$) dan op T1 ($M = 11.73$, $SD = 3.10$). Dit verschil was niet significant: $t(10) = -1.27$, $p = .23$, $d = .24$. De effectgrootte was klein. Er was geen sprake van een leereffect voor zesjarige kleuters op de Corsi-blokkentest.

Op de updatingtest bleken kleuters gemiddeld hoger te scoren op T2 ($M = 26.00$, $SD = 9.35$) dan op T1 ($M = 24.43$, $SD = 10.84$). Dit verschil was niet significant: $t(39) = -1.23$, $p = .23$, $d = .16$. De effectgrootte was klein. Er was geen sprake van een leereffect op de updatingtest. In figuur 4 zijn de gemiddelde scores weergegeven op de updatingtest op T1 en T2 uitgesplitst naar leeftijd in jaren. Wanneer apart naar elke leeftijdsgroep gekeken werd, bleek dat vierjarige kleuters gemiddeld lager scoorden op T2 ($M = 23.00$, $SD = 9.10$) dan op T1 ($M = 24.36$, $SD = 10.84$). Dit verschil was niet significant: $t(10) = .42$, $p = .69$, $d = .14$. De effectgrootte was klein. Er was geen sprake van een leereffect voor vierjarige kleuters op de updatingtest. Vijfjarige kleuters scoorden gemiddeld hoger op T2 ($M = 26.83$, $SD = 8.77$) dan op T1 ($M = 22.78$, $SD = 10.23$). Dit verschil was significant: $t(17) = -2.47$, $p < .05$, $d = .43$. De effectgrootte was klein. Er was sprake van een leereffect voor vijfjarige kleuters op de updatingtest. Zesjarige kleuters scoorden gemiddeld hoger op T2 ($M = 27.64$, $SD = 10.64$) dan op T1 ($M = 27.18$, $SD = 9.28$). Dit verschil was niet significant: $t(10) = -.25$, $p = .81$, $d = .05$. De effectgrootte was klein. Er was geen sprake van een leereffect voor zesjarige kleuters op de updatingtest.

Conclusie en Discussie

De huidige studie heeft de test-hertestbetrouwbaarheid van vier EF-instrumenten bij kleuters in de leeftijdscategorie van vier tot zeven jaar onderzocht. Daarnaast is de

test-hertestbetrouwbaarheid gemeten wanneer de vier instrumenten geïnterpreteerd werden als testbatterij en is de mogelijke aanwezigheid van leereffecten onderzocht.

Test-hertestbetrouwbaarheid

Uit de resultaten blijkt dat de vier testen verschillen betreffende de mate van test-hertestbetrouwbaarheid. Verschillen in test-hertestbetrouwbaarheid kunnen zijn ontstaan omdat de testen diverse aspecten van EF meten. Leeftijd blijkt niet van invloed te zijn op de test-hertestbetrouwbaarheid van de vier EF-instrumenten.

De HTKS is gebruikt als maat voor inhibitiecontrole. Uit de resultaten blijkt dat de test-hertestbetrouwbaarheid van de HTKS goed is. Dit is overeenkomstig met de verwachting dat de test-hertestbetrouwbaarheid van de HTKS minimaal voldoende zou zijn (Floyd & Kirby, 2001; Müller et al., 2012; Nampijja et al., 2010). Echter is voorzichtigheid geboden bij het interpreteren en generaliseren van deze resultaten aangezien de literatuur met betrekking tot inhibitietesten beperkt is. Tot dusver is huidig onderzoek het enige onderzoek naar de test-hertestbetrouwbaarheid van de HTKS. Er is meer onderzoek nodig naar deze specifieke test als aanvulling op de gevonden resultaten van de huidige studie.

De DCCS wordt wereldwijd breed ingezet in onderzoek naar de ontwikkeling van cognitieve flexibiliteit in de vroege kindertijd (Frye et al., 1995; Zelazo, 2006). In overeenstemming met het onderzoek van Müller en collega's (2012) en in tegenstelling tot het onderzoek van Beck en collega's (2011), blijkt uit de resultaten van de huidige studie dat de test-hertestbetrouwbaarheid van de DCCS onvoldoende is. Dit resultaat suggereert dat de DCCS gevoelig is voor dagelijkse fluctuaties in prestatie (Müller et al., 2012). De DCCS lijkt over onvoldoende test-hertestbetrouwbaarheid te beschikken om ingezet te kunnen worden als instrument om de ontwikkeling van EF en/of cognitieve flexibiliteit te meten.

Binnen huidig onderzoek zijn de Corsi-blokkentest en updatingtest ingezet om het werkgeheugen te meten. Uit de resultaten blijkt dat de test-hertestbetrouwbaarheid van de Corsi-blokkentest goed is. Dit is in overeenstemming met de verwachting dat werkgeheugen instrumenten over een goede test-hertestbetrouwbaarheid beschikken (Nampijja et al., 2010; Müller et al., 2012). Voorzichtigheid is echter geboden bij het interpreteren van deze resultaten, aangezien de literatuur met betrekking tot werkgeheugen instrumenten beperkt is. Daarnaast is dit onderzoek, tot dusver, het enige onderzoek naar de test-hertestbetrouwbaarheid van de Corsi-blokkentest. Binnen huidig onderzoek is geen onderscheid gemaakt tussen de verschillende aspecten van het werkgeheugen volgens het model van Baddeley (1996), waardoor het mogelijk is dat de Corsi-blokkentest andere aspecten van werkgeheugen meet dan de testen uit het literatuuronderzoek. Uit de resultaten blijkt dat de test-hertestbetrouwbaarheid van de updatingtest middelmatig is. Dit is niet in overeenstemming met de literatuur over test-

hertestbetrouwbaarheid van werkgeheugen instrumenten. Er is echter nog geen informatie bekend over testen die specifiek trachten de updating functie van het werkgeheugen te meten. Meer onderzoek is daarom nodig naar de betrouwbaarheid en validiteit van de nieuw ontwikkelde updatingtest.

In overeenstemming met de verwachting is er sprake van een goede test-hertestbetrouwbaarheid van de testbatterij (Willoughby & Blair, 2010). De test-hertestbetrouwbaarheid van de testbatterij is hoger dan drie van de vier individuele instrumenten (met uitzondering van de Corsi-blokkentest). Een testbatterij lijkt de werkelijke EF-capaciteiten meer nauwkeurig te kunnen meten wat een bevestiging is voor de veronderstelling dat EF ongedifferentieerd is in de kleutertijd (Espy et al., 2010; Wiebe et al., 2008).

Leereffecten

Uit de resultaten blijkt een significant leereffect voor de gehele steekproef op de HTKS. Dit is in tegenspraak met eerder onderzoek waarin geen significante leereffecten zijn gevonden (Nampijja et al., 2010; Müller et al., 2012). Deze discrepantie komt mogelijk doordat de taken in eerder onderzoek in dusdanige mate verschillen van de HTKS, bijvoorbeeld doordat Müller en collega's (2012) in hun onderzoek gebruik hebben gemaakt van een gecomputeriseerde taak. Ook kan er verschil zitten in verbale en non-verbale taken. Wanneer echter per leeftijdscategorie gekeken wordt, blijkt dat alleen voor vijfjarige kleuters sprake is van een significant leereffect. De groep vijfjarige kleuters is in de huidige steekproef de grootste groep ($n = 18$) waardoor deze groep een groter aandeel kan hebben gehad in de bevindingen.

In tegenstelling tot de bevindingen van Müller en collega's (2012) is in huidig onderzoek wel een leereffect gevonden op de DCCS. Uit de resultaten blijkt echter dat er geen leereffect is voor de afzonderlijke leeftijdscategorieën. De resultaten betreffende de aanwezigheid van leereffecten op de DCCS zijn dusdanig inconsistent dat hier geen conclusies uit getrokken kunnen worden. Een verklaring voor de inconsistentie ligt mogelijk in de moeilijkheidsgraad van de DCCS. De standaardtaak is wellicht te makkelijk, waardoor een plafondeffect optreedt en de complexe DCCS is wellicht te moeilijk waardoor deze taak niet voldoende differentieert tussen de verschillende leeftijdscategorieën (Doebel & Zelazo, 2015).

In tegenspraak met de verwachting dat werkgeheugen instrumenten gevoelig kunnen zijn voor leereffecten (Müller et al., 2012), zijn geen significante leereffecten gevonden op de Corsi-blokkentest. De discrepantie tussen deze resultaten kan worden verklaard doordat de werkgeheugen instrumenten in ander onderzoek andere aspecten van het werkgeheugen meten dan de Corsi-blokkentest. Het ontbreken van leereffecten op de Corsi-blokkentest wijst erop dat de prestatie op deze test een accurate weergave is van de huidige capaciteiten. Voor de updatingtest geldt dat er alleen sprake is van een

significant leereffect voor vijfjarige kleuters. Een interessante bevinding van huidig onderzoek is dat vierjarige kleuters op testmoment 2 gemiddeld lager hebben gescoord dan op testmoment 1. Voor zesjarige kleuters is wel een stabiel patroon te zien in scores. Een verklaring voor de inconsistente resultaten is dat de updatingtest niet leeftijdsadequaat is voor vierjarige en vijfjarige kleuters.

Limitaties en aanbevelingen vervolgonderzoek

Het onderzoeken van de psychometrische kwaliteiten van EF-instrumenten voor kleuters is een onderbelicht gebied binnen wetenschappelijk onderzoek (Müller et al., 2012). Een sterke kant van huidig onderzoek is dat hier aandacht aan besteed is door het onderzoeken van de test-hertestbetrouwbaarheid van vier EF-instrumenten voor kleuters. In huidig onderzoek waren vier testleiders elk verantwoordelijk voor de afname bij tien kleuters waardoor er sprake kan zijn van eventuele verschillen in afname. Deze verschillen zijn getracht te beperken door middel van een vaste afnamevolgorde van de testen en door gebruik te maken van handleidingen met expliciete instructies. Een beperking van huidig onderzoek betreft de selecte, doelgerichte en relatief kleine steekproef (N = 40) waardoor voorzichtigheid geboden is bij het generaliseren van de onderzoeksresultaten naar de populatie. Voor vervolgonderzoek is het van belang om een grotere en aselechte steekproef te trekken. Daarnaast is er sprake geweest van een tijdsinterval van vijf tot tien dagen tussen testmoment 1 en testmoment 2. Dit kan invloed hebben gehad op de resultaten van de test-hertestbetrouwbaarheid en het al dan niet optreden van leereffecten. Vervolgonderzoek zou een vast tijdsinterval moeten hanteren tussen de twee testmomenten om de consistentie te vergroten. Een ander punt met betrekking tot de tijdsintervallen is de keuze van huidig onderzoek voor een relatief klein tijdsinterval tussen beide testmomenten (vijf tot tien dagen) om te kunnen waarborgen dat de gevonden verschillen in scores geen reflectie zijn van veranderingen die ten grondslag liggen aan ontwikkeling van de kleuters. Voor het vaststellen van de test-hertestbetrouwbaarheid is het echter van belang om verder onderzoek te doen met meer systematische variatie in test-hertest intervallen (McArdle & Woodcock, 1997). Verder zou vervolgonderzoek zich kunnen richten op het ontwikkelen van een testbatterij voor EF bij kleuters, die de ongedifferentieerde structuur van EF volledig ondervangt.

Referenties

- Anastasi, A., & Urbina, S. (1997). Reliability, validity & norms. In A. Anastasi & S. Urbina (Eds.) *Psychological testing* (pp. 84-105). Upper Saddle River, NJ: Prentice Hall
- Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences, 93*, 13468-13472. doi:10.1073/pnas.93.24.13468
- Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York, NY: Guilford Press
- Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test-retest reliability. *Journal of Cognition and Development, 12*, 169-193. doi:10.1080/15248372.2011.563485
- Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180-200. doi:10.1016/j.dr.2009.05.002
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616. doi:10.1207/s15326942dn2802_3
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.00
- Davis, H., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25-31. doi:10.1080/00049539508258765
- Diamond, A. (2001). A model system for studying the role of dopamine in the prefrontal cortex during early development in humans: Early and continuously treated phenylketonuria. In C. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 433-472). Cambridge, MA: MIT Press.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168. doi:10.1146/annurev-psych-113011-143750
- Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review, 38*, 241-268. doi:10.1016/j.dr.2015.09.001
- Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology, 20*, 607-613. doi:10.1037/0894-4105.20.5.607
- Espy, K. A., Sheffield, T. D., Wiebe, S., Clark, C. A. C., & Moehr, M. (2010). Executive control and dimensions of problem behaviors in preschool children. *Journal of*

- Child Psychology and Psychiatry*, 52, 33–46. doi:10.1111/j.1469-7610.2010.02265.x
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Floyd, R. G., & Kirby, E. A. (2001). Psychometric properties of measures of behavioral inhibition with preschool-age children: Implications for assessment of children at risk for ADHD. *Journal of Attention Disorders*, 5, 79-91. doi:10.1177/108705470100500202
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527. doi:10.1016/0885-2014(95)90024-1s.
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, 134, 31-60. doi:10.1037/0033-2909.134.1.31
- Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, 36, 129-151. doi:10.1016/S0021-9924(03)00003-0
- Hughes, C. (1998). Executive functions in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233-253. doi:10.1111/j.2044-835x.1998.tb00921.x
- Hughes, C. (2011). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development*, 20, 251-271. doi:10.1002/icd.736
- Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12, 9-15. doi:10.1111/j.1467-7687.2009.00848.x
- Kessels, R. P., Van den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-tapping task and its association with the WAIS-III digit span. *Assessment*, 15, 426-434. doi:10.1177/1073191108315611
- Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. (2001). Test–retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology*, 19, 339–348. doi:10.1348/026151001166137
- McArdle, J. J., & Woodcock, R. W. (1997). Expanding test-retest designs to include developmental time-lag components. *Psychological Methods*, 2, 403-435. doi:10.1037//1082-989x.2.4.403
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: the

- head-toes-knees-shoulders task. *Frontiers in Psychology*, 5.
doi:10.3389/fpsyg.2014.00599
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex Function. *Annual Review of Neuroscience*, 24, 167–202.
doi:10.1146/annurev.neuro.24.1.167
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100. doi:10.1006/cogp.1999.0734
- Müller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist*, 26, 271-287. doi:10.1080/13854046.2011.645558
- Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliott, A. M., et al. (2010). Adaptation of Western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *British Journal of Educational Psychology*, 80, 15–30.
doi:10.1348/000709909x460600
- Neuman, W.L. (2011). *Understanding Research*. New York City, NY: Pearson.
- Orsini, A. (1994) Corsi's Block Tapping Test: Standardization and concurrent validity with WISC-R for children aged 11 to 16. *Perceptual and Motor Skills*, 79, 1547-1554.
<http://dx.doi.org/10.2466/pms.1994.79.3f.1547>
- Ponitz, C.C., McClelland, M.M., Jewkes, A.M., Connor, C.M., Farris, C.L., & Morrison, F.J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, 23, 141-158.
doi:10.1016/j.ecresq.2007.01.004
- Rothbart, M., & Posner, M. (2001). Mechanism and variation in the development of attentional networks. In C. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 353–363). Cambridge, MA: MIT Press.
- Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., ... & Taylor, E. (2001). Mapping motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, 13, 250-261.
doi:10.1006/nimg.2000.0685
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44, 575-587. doi:10.1037/0012-1649.44.2.575
- Willoughby, M., & Blair, C. (2011). Test-retest reliability of a new executive function battery for use in early childhood. *Child Neuropsychology*, 17, 564-579.
doi:10.1080/09297049.2011.554390

Zelazo, P.D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297-301.

doi:10.1038/nprot.2006.46

Zelazo, P. D., & Müller, U. (2010). Executive functioning in typical and atypical children. In U. Goswami (Ed.), *Handbook of cognitive development* (2nd ed., pp. 574–603). Oxford, UK: Blackwell.