HHS Public Access

Author manuscript

Infant Child Dev. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

Infant Child Dev. 2019 ; 28(1): . doi:10.1002/icd.2116.

Examining the Associations Between Performance Based and Ratings of Focused Attention in Toddlers: Are We Measuring the Same Constructs?

Ibrahim H. Acar, Ph.D.,

Department of Early Childhood Education, College of Education, Istanbul Medipol University, Room C-307, Beykoz, 34810 Istanbul, Turkey, +90 212 444 8544

Scott Frohn, Ph.D.,

Psychometrician and Educational Psychologist, PSI Services LLC, 18000 W 105th St, Olathe, KS 66061, USA

Amanda Prokasky, M.S.,

Department of Child, Youth, and Family Studies, Center for Brain, Biology and Behavior, C80 East Stadium, University of Nebraska-Lincoln, Lincoln, NE 68588-0172, 402-472-8982

Victoria J. Molfese, Ph.D., and

Department of Child, Youth & Family Studies, 133 Mabel Lee Hall, College of Education and Human Sciences, University of Nebraska- Lincoln, Lincoln NE 68588-0236, Phone (402) 472-6399

John E. Bates, Ph.D.

Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St. Bloomington, IN 47405, Phone: (812) 855-8693

Abstract

The study examines the concurrent and longitudinal associations between ratings-based measures (parents, secondary caregivers, observers) and performance-based measures of focused attention in toddlers aged 30- (n = 147), 36- (n =127), and 42-months (n =107). Parents and secondary caregivers rated focused attention behaviors using the Children's Behavior Questionnaire (Rothbart et al., 2001), and observers rated toddlers' focused attention during a series of laboratory tasks using the Leiter-R Examiner Rating Scale (Roid & Miller, 1997). Toddlers' behaviors on three structured tasks (Token Sort, Toy Play, Lock Box) were used to assess their performance based focused attention in a laboratory setting. Correlations show that parent ratings are not related to observer and teacher ratings, or to the performance-based measures at all ages tested. Second, based on confirmatory factor analyses, a single factor explains the common variance between indicators when the parent ratings are not included in the models. The single factor shows measurement invariance between ages 36 and 42 months based on factor structure, relations of indicators to the factor, and factor scale over time. Third, indicators of focused attention at age 30 months do not seem to measure a common, coherent factor. Interpretations of similarities and

Correspondence to: Ibrahim H. Acar.

differences between ratings and performance-based indicators of focused attention and the presence of a focused attention construct are discussed.

Keywords

Toddlers; Ratings of Focused Attention; Performance based Focused Attention; Structured Tasks; Temperament

Attention skills are important for all children as they engage in and respond to events and behaviors in complex social and physical environments. Attention skills are important for problem solving, waiting to take turns in a school setting, and understanding the rules of games played with peers. Previous research has shown that early attention skills are predictive of the development of concurrent and future psychosocial, academic, and cognitive skills (e.g., Cuevas & Bell, 2014; Duncan et al., 2007; Graziano, Calkins, & Keane, 2011; Ruff & Capozzoli, 2003). Researchers have reported evidence of individual differences in attention skills in the behaviors of infants and toddlers along with the emergence of cognitive and self-regulatory skills (Gaertner, Spinrad, & Eisenberg, 2008; Rothbart, Posner, Kieras, 2006; Ruff & Rothbart, 1996).

Attention skills depend on the characteristics of the context, such as task demands and other external stimuli that compete for the child's attention (NICHD Early Child Care Research Network, 2005; Ruff & Lawson, 1990). In studies of attention, researchers have categorized attention behaviors as casual (looking at target but not engaging; only the eyes are on the target), settled (touching or playing with the target with no intention or purpose; e.g., Ruff, Capozzoli, & Weissberg, 1998); and focused attention ("sustained and active engagement with a stimulus or task"; Gaertner et al., 2008, p. 340). Thus, there are different perspectives on the definition and conceptualization of attention in children. In the current study, we specifically explored focused attention, consistent with the definition above, as those behaviors in which both the eyes and the hands are focused on the target with the intention of interacting with the target as instructed by an experimenter. This definition does not include attention when the child is simply playing with toys or looking at toys provided during the task.

The purpose of the study was to examine focused attention in toddlers to determine the stability of the construct when measured across three time-points—30-, 36-, and 42-months. In addition, previous research has shown that there is a difference between performance based behavior/skills (e.g., attention in a structured context) and ratings of behavior/skills that could encompass different contexts (e.g., parents' and caregivers' reports of attention or more natural or everyday context) (Toplak et al., 2013). Thus, the relation between performance based measures of focused attention in toddlers and ratings of focused attention (parent, secondary caregivers, and observers) are examined.

Literature Review

Attention in Young Children

Focused attention measured during play-like tasks has been found to improve with age, with older children able to engage longer with an instructed task and better able to resist distractors once engaged (Ruff & Capozzoli, 2003; Ruff, Capozzoli, & Saltarelli, 1996). However, characteristics of the stimuli (e.g., attractiveness, complexity, demands of the context, familiarity of stimuli, and novelty of context) and individual differences (e.g., temperament and internal motivation) affect children's focused attention (Harman, Posner, Rothbart, & Thomas-Thrapp, 1994; Posner & Rothbart, 2007; Ruff & Lawson, 1990; Ruff & Rothbart, 1996). For example, Ruff and Lawson (1990) conducted two cross-sectional studies with children at ages 1, 2, and 3.5 years in Study 1, and children at ages 2.5, 3.5, and, 4.5 years in Study 2. The researchers examined whether children focused their attention on a certain set of toys over other toys during free play and the length of time children sustained their attention on the task across ages. Results showed children's focused attention skills increased as they got older, and older children could focus their attention significantly longer than younger children.

Assessment of Attention Skills in Early Childhood

Considering the multifaceted structure of attention, researchers have used different methods to measure it, such as parent ratings (Rothbart, Ahadi, Hershey, & Fisher, 2001) and performance on structured laboratory tasks (Graziano, Calkins, & Keane, 2011). Parent reports of attention using rating scales are intended to reflect children's attention capacity in their everyday lives. Performance based measures of attention involve standardized laboratory tasks that target assessment of children's attention, such as Token Sort (colored tokens are mixed together and the child's task is to sort the tokens by color into separate bins), Toy Play (the child is instructed to sit a a table and play with toys on a table while the experimenter does paperwork) and Lock Box (the child is instructed to open a lock to retrieve a toy using a set of keys that do not open the lock). Although it is assumed that both types of measurement approaches assess children's focused attention skills, it is unclear whether ratings and performance based measures are capturing the same construct and whether children's focused attention is uniformly measured by different laboratory tasks.

Ratings of Attention

Ratings of focused attention are based on the perceptions of the informants (e.g., parent, secondary caregiver, or independent observer). Item are written such that a wide range of attention-related behaviors across different environments, such as home, school, or during play, could be used as the basis for the rating. One of the most common rating measures of children's focused attention is the Attentional Focusing subscale of the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fischer, 2001), and the CBQ Short Form (CBQ-SF; Putnam & Rothbart, 2006). For these questionnaires parents use a 7-point Likert-type scale to rate how well a number of statements describe the intensity and duration of their child's attention focus. The original nine-item scale and the shortened six-item version demonstrate adequate internal consistency, with Cronbach's alphas ranging from .67 to .75 (Putnam & Rothbart, 2006; Rothbart et al., 2001). The CBQ also has been used for

secondary caregiver report, with Cronbach's alphas ranging from .78 to .96 (e.g., Gunnar, et al., 1997, 2003).

Performance Based Measures of Focused Attention

Performance based measures of attention use structured tasks in which the child is expected to attend to and actively engage with the task stimuli, and resist other distractions present in the task context (e.g., Gaertner et al., 2008; Ruff & Capozzoli, 2003). Performance based measures are generally administered in settings where stimuli are controlled by an experimenter so that each child experiences the same process. Scores from performance based tasks measure a child's accuracy and response time during the task, which are through to reflect how the child focuses his/her attention during task performance (Ruff & Rothbart, 1996; Toplak et al., 2013).

Performance based tasks in structured settings tend to have similar characteristics. For example, Gaertner and colleagues (2008) provided mothers and toddlers with a basket of toys and the child were allowed to play with toys for 3 minutes. Focused attention was measured by examining toddlers' attention to the toys, concentration, orientation toward the toys, and intensity of engagement with the toys. Lack of focused attention was evidenced by off-task behaviors (e.g., looking around) and lack of engagement with provided toys. In other tasks, researchers utilized video clips to measure children's focused attention to the videos. Graziano et al. (2011) used 5-minute segments of a video about a puppy exploring a neighborhood and timed the duration of children's attention to the video without being distracted. Overall duration of focused attention was measured as the ratio of time looking at the video to video length.

Ratings of Attention versus Performance Based Measures of Attention

Ratings and performance based measures of focused attention are different from each other in terms of administration and scoring, but there are additional differences that can contribute to differences in focused attention scores across ratings and performance based measures. Parent and secondary caregiver ratings of focused attention have the advantage of being assessed by the person(s) who know the child the best, including how the child responds to a variety of "real world" situations. However, parents and secondary caregivers' ratings may be unintentionally influenced by positive (or negative) perceptions of the child on the part of the parent, introducing social desirability bias in focused attention scores. Secondary caregivers, in contrast to most parents, have the added advantage of having a larger "comparison" group by which to judge the child's focused attention skills via experience with the other children currently or previously in their care. However, secondary caregiver's assessment of children's focused attention is limited to experiences based on structured setting of the early care and education classroom, most often during the daytime hours, which may not generalize to other settings. Performance based measures of focused attention involve measurement based on observations of the child's behavior in a controlled setting by research personnel not familiar with the child. Focused attention performance is judged based on a standardized research protocol for each task are assumed to reflect differences not based on familiarity with the child or differences in focused attention tasks. However, the controlled setting of the laboratory is rarely found in the real world. Thus, a

child's performance on a laboratory focused attention task may not generalize to the child's focused attention skills in other settings, particularly at home or school.

Although there are inconsistencies among findings and paucity of studies examining differences between ratings of focused attention and performance based measures, these differences between these two types of measures of focused attention may be reflected in the findings from previous research that the strength of the statistical association between caregiver or naturalistic observer ratings of focused attention and performance based measures are medium (e.g., $r = .31$; Choudhury & Gorman, 2000) or non-significant (Gaertner et al., 2008; Toplak et al., 2013). Such statistical associations could reflect an inconsistency between the constructs measured by different methods or method-specific measurement error. The inconsistent findings could also pertain to theoretically different perspectives on cognitive ability. Ratings of focused attention and performance based measures may tap different levels and parts of attention. Stanovich (2009) posited that the cognitive mind (ability) is divided in two parts; the algorithmic and the reflective mind. Algorithmic mind refers to cognitive processing mechanisms, such as working memory, long-term memory, and perceptual focusing. Reflective mind refers to broader goals and beliefs of an individual and how individuals select their actions related to their beliefs and goals (Stanovich, 2009; 2011; Toplak et al., 2013). Toplak and colleagues (2013) explained that ratings of executive function (which includes focused attention) cover a broad range of behaviors that are observed in the real environment. However, performance based measures selectively target only certain behaviors, such as behavioral inhibition, in a structured setting, which may provide information about how children process and become efficient in information processing that might be useful for thinking about children's performance in the classroom and other structured learning situations. On the other hand, ratings of attention may give useful information about how children control and use their focused attention skills in every day, out of school settings (Toplak et al., 2013). Both ratings and performance based measures of attention skill represent important sources of information regarding children's focused attention. Thus, comparing both types of measurement provides the broader information needed to understand how children's focused attention skills emerge in different contexts. In addition, there are several ways that utilizing both ratings and performance based focused attention as well as comparing these two methods of data collection could contribute the understanding of focused attention in children. First, discrepancies between ratings and performance based focused attention could reflect true variation in children's focused attention across different settings and circumstances. Second, utilizing information from both ratings and performances may inform practice as both methods reflect differences and/or similarities in focused attention reflecting contextual characteristics. Third, comparing these two methods of assessing children's focused attention may inform research design. Researchers could choose caregiver ratings versus performance based measures depending on the purpose of the study. Considering these ways of utilizing both ratings and performance based measures of focused attention of children, in the current study, we examined the information gained from ratings and performance based measures of focused attention in toddlers, a developmental period in which changes in self-regulation generally, and attentional skills specifically, are occurring (Rothbart, Posner & Kieras, 2006). Our study investigated consistencies in measures across three ages and relations between types of

measures. To do so, we used ratings of attention (parents, secondary caregivers, and observer) and performance based measures of attention in a laboratory setting using three different types of structured tasks.

The Current Study

The purpose of the current study was to investigate the relations between different types of measures of toddler's focused attention, including both performance based tasks (Token Sort, Lock Box, and Toy Play) and ratings of focused attention provided by a parent, secondary caregiver, and based on reports of observers), over the course of one year. The current study is directed by two research questions. The first research question is: *"Do ratings and performance based measures of focused attention reflect the same construct?"* Given some inconsistencies between ratings and performance based tasks, we explored the construct of focused attention at each age to examine whether a single factor model best fitted the data. Alternatively, based on Stanovich's (2009) framework, it is plausible that performance based tasks measure cognitive processing mechanisms (i.e., algorithmic abilities), whereas parent, secondary caregiver-, and observer ratings measure broader-based aspects of focused attention, including also reflective abilities. Therefore, we investigated whether these different types of measures conform to a single- or two-factor model.

Once we determined whether a single- or two-factor model of focused attention best represents the relations among measures of focused attention, we addressed our second research question: *"Do the different measures of focused attention measure the same construct over time?"* We hypothesized that over the course of one year—from age 30-months to age 42-months—the relation between these measures and the construct of focused attention would be stable over time. In other words, we expected the measurement models to demonstrate longitudinal invariance across the three time points.

Method

Participants

The participants in the present study are part of a larger investigation examining the relations between measures of sleep and self-regulation in toddlers 30- to 42- months of age. The study was approved by the Institutional Human Research Protection Committee. English speaking families with typically developing toddlers aged 30-months were recruited for participation; children with diagnosed physical or developmental disabilities were excluded. Recruitment was conducted at local day care programs and out of home day care sites, and through personal contacts, phone calls, distribution of flyers describing the study, and by asking participants to refer other potential participants to the study. Families with toddlers who fit within the age range, lived within a 1.5-hour drive of the research site, and who expected to be available for the one-year testing period of the study were eligible for participation. Data were gathered when toddlers were 30, 36, and 42 months.

The longitudinal sample included 147 toddlers aged 30 to 31 months (M = 30.14 months, SD = 0.35; 50.7% males) at the first time point; 127 toddlers 36 to 37 months (M = 36.09 months, SD = 0.28; 52.8% males) at the second time point; and 107 toddlers 42 to 43

months (M = 42.15 months, SD = 0.35; 53.3% males) at the third time point. The majority of toddlers were White, Non-Hispanic (82.8%), with the remaining African American (1.7%), Hispanic (3.4%) or other (12%). The ethnic breakdown of the sample reflected the broader demographics of the community from which it was drawn. Parents' ages ranged from 23 to 46 years (M = 32.09, SD = 4.52) when the toddlers were 30 months. Parent education levels were college degrees (84.8%), some college (12.2%) or high school diploma (2.2%). Family income was reported in blocks of $5,000 and ranged from less than $10,000 to more than $125,000, with a mode of $75,000 to $80,000.

Secondary caregivers were either home-based caregivers (where children go to the caregivers' homes or caregivers come to children's homes), out of home daycare providers, and teachers in center-based early childhood education programs. Some parents did not use secondary caregivers for their children. Therefore, we did not have data from these secondary caregivers. At 30 months we had 75 secondary caregivers (2 males; 12.1 % home-based caregivers) who reported contact hours ranging from 4 to 50 (M = 30.94, SD = 13.72) per week. At 36 months, we had 60 secondary caregivers (2 males; 17.3% home-based caregivers) who reported contact hours ranging from 12.5 to 50 (M = 32.83, SD = 9.51) per week. At 42 months, we had 63 secondary caregivers (2 males; 13.2% home-based caregivers) who reported contact hours ranging from 0.5 to 45 hours (M = 28.78, SD = 14.06) per week. Overall, most of the children with secondary caregivers were in center-based early childhood education programs.

Data Collection Procedure

After recruitment, a home visit was scheduled with families to inform them about the study and its procedures. After obtaining consent to participate, the parent, usually the mother, was given a packet of questionnaires. One week later, the parent and toddler came to the lab to participate in a series of 17 behavioral tasks lasting approximately 1.5 hours. Efforts were made to reduce fatigue, hunger, sleepiness, etc. to ensure the toddlers were able to perform on the tasks. First, lab visits were scheduled at a time that parents reported would not conflict with the toddler's regularly scheduled meals or naps. Second, the experimenter or the parent or child indicated that the child needed a break (e.g. to use restroom), the task was paused and resumed when the child was ready. Finally, half way through the testing protocol, a break was taken, and the child was provided with a drink, snack, and opportunity to use the restroom. The target tasks (Toy Play, Lock Box, and Token Sort) were administered immediately followed this break, and were the 10th, 11th and 14th tasks in the fixed series of 17 tasks.

Parents provided information needed to contact their child's secondary caregiver, who completed the demographic form about their contact with target child and the Children's Behavior Questionnaire - Short Form (Putnam & Rothbart, 2006). The parent, toddler, and secondary caregivers were provided with compensation for participating.

Research assistants used video recordings or live observations of the laboratory tasks to score the Leiter-R Examiner Rating Scale (Leiter-R; Roid & Miller, 1997) based on behaviors observed across all 17 of the laboratory tasks. Scores from the attention subscale of Leiter-R Examiner Rating Scale were used in the analyses described below.

Measures

Demographics.—Parents reported on their toddlers' and their own sex, ethnicity, and age, parental relation to the child, and marital status. Parents reported their own education from six pre-defined categories (8th Grade or Less, Some High School, GED, High School Diploma, Some College, and College Degree). Annual family income was reported on a 25-point scale, ranging from low ("Less than $10,000") to high ("More than $125,000") in intervals of $5,000.

Ratings of Attention

Parent and Secondary Caregiver Report.: We used the short-form of the Children's Behavior Questionnaire (CBQ-SF; Putnam & Rothbart, 2006) to obtain parent- and secondary caregiver-ratings of children's focused attention. The CBQ was created as a parent report measure of children's temperament, but it has also been used for teacher/secondary caregiver report (e.g., Eisenberg et al., 2011; Rudasill et al., 2014). Parents and secondary caregivers completed the entire CBQ-SF (94 items on 15 subscales), responding to statements using a 7-point scale (1 = *"Extremely untrue of your child"* to 7 = *"Extremely true of your child"*; Rothbart et al., 2001). For the current study, we used scores from the six-item Attentional Focusing subscale, which measures the reporter's perception of a child's ability to focus on a given activity (e.g., "When building or putting something together, becomes very involved in what s/he is doing, and works for long periods"). Internal consistency for the subscale was adequate for parents (α = .66 at 30 months, α = .69 at 36 months, and α = .67 at 42 months), and secondary caregivers (α = .78 at 30 months, α = .82 at 36 months, and α = .75 at 42 months). These reliability estimates are similar to those reported by Putnam and Rothbart (2006).

Observer Report.: We used the 10-item attention subscale of the Leiter-R Examiner Rating Scale from the International Performance Scale – Revised (Leiter-R; Roid & Miller, 1997) to obtain observer ratings of children's focused attention during the laboratory visit. The Leiter-R has been widely used in research to examine children's cognitive and emotional abilities (Roid, Pomplun, & Martin, 2009). The Examiner Rating Scale is typically scored after the administration of the Leiter-R to report children's behaviors during the assessment. Items are rated on a 4-point Likert-like scale (0 = *"rarely/never"* to 3 = *"usually/always"*) and measures overall level of focused attention (e.g., *"pays attention to details within tasks"*). We did not implement the Leiter-R tasks in the current study, but scored the attention subscale items based on behaviors observed across all administered laboratory tasks. At each measurement time point, research assistants used video recordings or live observations of the laboratory tasks to score the Leiter-R Examiner Rating Scale. Observers individually scored the rating scale, and then item-by-item scores were compared across each pair of observers to reach consensus ratings of the child's attention subscale scores. Because we utilized the Leiter Examiner Rating Scale without conducting the Leiter-R, we needed to establish that the focused attention ratings produced scale scores that met criteria for inclusion in a confirmatory factor analysis (i.e., sufficiently reliable and unidimensional). To establish unidimensionality of the attention scale, the items were subjected to a principal axis factor analysis, with varimax rotation, resulting in a one-factor solution at 30 months (Eigenvalue = 6.63, variance = 66.34%, α =.94) and 42 months (Eigenvalue = 6.21, variance

= 62.08%, α = .93), and a two-factor solution at 36 months (first factor: Eigenvalue = 6.50, variance = 65.01% and second factor: Eigenvalue = 1.18, variance = 11.8% at 36 months). However, given the small percent of variance in the second factor at 36 months, and because a scale using the one factor solution had higher internal reliability (α = .93), we chose to retain the one factor solution for all time-points. At each age, we computed an unweighted composite average score based on the 10 attention subscale items.

Performance Based Measures of Attention.

Token Sort.: This task was adapted from the bead sorting task of the Laboratory Temperament Assessment Battery (LAB-TAB; Goldsmith & Rothbart, 1999). For Token Sort, the experimenter brought a box with three colors of tokens mixed together and asked the child to sort them, by color, into separate bins. The duration of the Token Sort task was three minutes, but sometimes was stopped early under special circumstances (e.g., the child became extremely upset and the parent requested the research assistants to end the task). We coded the duration of the child's engagement in actively sorting the colored tokens into the three boxes as focused attention. If the child engaged in off-task behaviors, such as looking around, wandering, going to the door, or playing with tokens instead of sorting them, the duration of these behaviors was not scored as focused attention. Thus, we attempted to capture the child's persistence, engagement, instruction following, and interest with the assigned task. The final score was the proportion of total time the child spent on the task as instructed out the duration of the task (up to 3 minutes).

Toy Play.: This task was adapted from the sustained attention task used by Ruff, Capozzoli, and Weissberg (1998). In this task, the child was asked to play with a set of toys on a table. The experimenter sat in the room across from the child but did not interact with the child. The task lasted up to five minutes, unless the task was stopped upon the parent's request. Within the five-minute period, the child's engagement, focus, and on-task behaviors (i.e., keeping eyes and hands on the toys) were timed. It was considered inattention if the child looked around but still kept the toys in hand because both hands and eyes were not on the toy. This task provided information about children's focused attention and engagement with the toys when the experimenter was present in the room as a potential distractor. The final score is the proportion of total time the child spent on the task as instructed (i.e., with eyes and hands on the toys) out the task duration (up to five minutes).

Lock Box.: This task was adapted from Calkin's (2002) task. For this task, the child was shown how to open a lock and then was allowed to pick out a toy to take home with them at the end of the lab visit. The toy was then put into a clear box and secured with a lock. The child was given a ring of keys—none of which worked—and told to try to get their toy out of the box while the experimenter left the room to get the next game ready. The task lasted up to four minutes, unless stopped at the parent's request. The child's focused attention was measured as duration of engagement with the keys or box with the intent to open the box. The final score is the proportion of total time the child spent on the instructed task out of the task duration (up to four minutes).

As indicated earlier, although nature of each task was different, the main purpose was to capture duration of attention given to the task by the child as instructed by the experimenter. Therefore, we used duration (i.e., proportion of time spent on the given task) as the representative composite score for focused attention.

Duration of focused attention was measured using Datavyu Software (Datavyu Team, 2014) by watching videos of the task while simultaneously timing the duration of focused attentions. Two coders timed the pre-defined behaviors based on instructions given to the child for each task and used proportion scores to reflect the proportion of time in which the child was attending to the task as instructed during the total task time. Approximately 20% of data were coded by the two coders. Intraclass correlation ranged from .88 to .99 across the three ages and three tasks, indicating good reliability.

Analytical Approach

To address our first research question—do ratings and performance based measures of focused attention (FA) load on a single construct—we estimated two measurement models in a confirmatory factor analysis (CFA) framework at each age. The hypothesized one-factor models were estimated with the three task measures (Token Sort, Lock Box, Toy Play) and three ratings (Parent-Report CBQ FA, Secondary Caregiver-Report CBQ FA, Observer-Report Leiter FA) loading freely on a single latent factor. Alternative two-factor models were estimated with the three task measures loading on a latent Task factor, and the three ratings loading on a latent Rating factor, with a correlation estimated between the two latent factors. Model indicators—the six measures of focused attention—were evaluated for significant loadings. Any variable with non-significant loadings at all timepoints for either the one-factor or two-factor models was dropped, and subsequent modeling was performed to evaluate the factor structure of FA without the variable.

Absolute model fit was evaluated using model χ^2, Root Mean Square Error of Approximation (RMSEA), and the Comparative Fit Index (CFI). Non-significant χ^2 values ($p > .05$) indicate good fit, RMSEA values below .06 indicate close fit and values less than .08 indicate acceptable fit, and CFI values above .95 indicate good fit and values above .90 indicate acceptable fit (Brown, 2006; Hu & Bentler, 1998). One- and two-factor models were compared using the relative fit indices Akaike information criterion (AIC) and Bayesian information criterion (BIC), where lower values indicate a model better represents the data (Burnham & Anderson, 2004). Because both one-factor and a two-factor models were based on the same covariance matrices, and the two-factor models have an additional estimated parameter (correlation between the factors), we conducted a −2*log-likelihood (−2LL) difference test with one degree of freedom to empirically compare fit between the one- and two-factor models. A one-factor model with a significantly ($p < .05$) greater log likelihood value than the two-factor model would be considered a poorer model and therefore the two-factor model would be considered more parsimonious (Brown, 2006). Essentially, this process tests *whether* the measurement models at each age are the equivalent or invariant, and *to what degree* the measurement models are the equivalent by placing a series of increasingly restrictive constraints on the models. The most basic level of equivalence is “configural” invariance, meaning that the factor structures are the same.

Configural invariance establishes that all of the indicators significantly relate to the latent factor at each time point. In statistical terms, each measurement model must have matching sets of indicators, all with significant factor loadings. Model parameters (i.e., loadings, intercepts, and residual variances) are freely estimated.

The next level of equivalence is “metric” invariance, meaning that the indicators relate to the factor in the same way at each age (i.e., the factor loadings are identical over time). Metric invariance establishes that the same construct is being measured at each time point. The loadings of each measurement model are constrained at each age, so that, for example, the loadings for Lock Box are fixed to be equal at each age, the loadings for Toy Play are fixed, and so on. The metric model is then compared to the configural model, and if the scaling-corrected difference in log-likelihoods (−2LL) is not significant, the metric model is considered to be “not worse” than the configural.

The strongest level of equivalence we tested for is “scalar” invariance, meaning that the latent constructs measured at each time point are on the same scale (i.e., the intercepts vary consistently relative to the factor over time). Scalar invariance is necessary for mean comparisons across age, and if scalar invariance is not established, comparisons of latent factor means are not meaningful (e.g., the constructs could not be used to determine whether older children have greater focused attention than younger children). In addition to constraining factor loadings at each age as with the metric model, for the scalar model the indicator intercepts are also constrained. The scalar model is then compared to the metric model via a −2LL test.

Results

Descriptive statistics of all study variables are presented in Table 1. Table 2 shows the correlations between parent and secondary caregiver reported ratings of focused attention with observer ratings based on the Leiter Examiner Rating Scale and performance based measures of attention during the laboratory tasks. We examined correlations between reporters’ ratings of attention across ages. There were significant, moderate correlations within the three sources of reporters’ ratings from 30 to 36 months, from 36 to 42 months, and from 30 to 42 months. However, correlations between parent-reports and reports by secondary caregivers and observer reports were weak and significant for only 2 of 18 comparisons. Correlations between secondary caregivers and observer reports were significant and moderate to strong in all comparisons.

We also examined correlations between performance based attention scores across ages. There were significant, moderate correlations within the three measures at 30 to 36 months, from 36 to 42 months, but correlations between 30 and 42 months were low within all three measures as was the correlation between Toy Play at 36 and 42 months. However, correlations between performance based scores were low and with only 6 comparisons reaching significance.

Comparison of One- and Two-Factor Models

To facilitate interpretation, all indicator variables were standardized at their 30-month values (Mean = 0, Variance = 1). CFA modeling was conducted using *Mplus version 7.11* (Muthén & Muthén, 1998–2013). One- and two-factor CFA models are presented in Table 3. For the one-factor model at 30 months, four of six indicators had non-significant factor loadings (Lock Box, Toy Play, Parent-Report CBQ FA and Secondary Caregiver-Report FA). In the two-factor model at age 30 months, all three indicators of the Task factor had non-significant loadings, and one indicator of the Rating factor had non-significant loadings (Parent-Report CBQ FA). For the one-factor models at 36 and 42 months, the loading for Parent-Report CBQ FA was non-significant, as was the loading for Lock Box at 36 months. For the two-factor models at age 36 and 42 months, the loadings for Parent-Report CBQ FA on the Rating factors were non-significant. Comparisons of model fit are presented in Table 4

Because the loading for Parent-Report CBQ FA were non-significant at all time-points for both one-factor and two-factor models, a second set of CFAs—herein referred to as the "revised" models—were estimated excluding this set of variables. For the revised CFAs, the patterns of significant loadings were similar to what was found with the original set of CFAs that included Parent Report CBQ FA. Three of five loadings for the one-factor model at 30 months were non-significant (Lock Box, Toy Play, and Teacher Report CBQ FA). The task factor loadings in the two-factor model at 30 months were non-significant. For the one-factor 36 month model, only one indicator, Lock Box, had a non-significant loading. All other loadings were significant for one- and two-factor models at each timepoint. These revised models are presented in Table 5 and fit comparisons are presented in Table 6

Both revised one- and two-factor models at 30 months did not fit well: single-factor fit, $\chi 2(5) = 15.42$, $p = .009$, RMSEA = .119, CFI = .721; two-factor fit, $\chi 2(4) = 9.94$, $p = .059$, RMSEA = .093, CFI = .864. However, revised models at 36 and 42 months did fit well, with all χ^2 values non-significant, RMSEA = .000, and CFI = 1.000. The single-factor models at 36 and 42 months were more parsimonious (i.e. they did not fit worse) than the two-factor models, as indicated by the non-significant p-values from the $\Delta -2LL$ tests, and lower AIC and BIC values for the one-factor models. Focused attention was deemed to be most concisely represented by a single-factor—rather than two-factors—and therefore the revised one-factor models were retained for further measurement invariance testing. Because the revised 30month models fit so poorly, each with three non-significant loadings, invariance testing was conducted using only 36-month and 42-month data.

Measurement Invariance

A configural model was estimated by including the two measurement (CFA) models in a single model with an estimated covariance between the two Focused Attention latent factors at 36 and 42 months. Loadings, intercepts, and residual variances for each indicator were freely estimated. Covaried residuals between tasks across age were also freely estimated (e.g., residuals of Token Sort at 36-months were covaried with residuals of Token Sort at 42 months). These covaried residuals were estimated to account for identical measurement instruments at different time points. To scale the latent factors in the configural model, the loadings for the Leiter indicator at each age were set to 1.0, as it had the strongest loading in

the individual CFAs. The Leiter intercepts at each age were set to 0.0. The configural model fit very well, $\chi 2(29) = 28.45$, $p = .494$, RMSEA = .000, CFI = 1.000.

A metric model was then estimated with indicator loadings constrained across age. To scale the latent factors in the metric model, the variance of the 36 month latent factor was set to 1.0, and the Leiter intercepts were set to 0.0. The metric model fit well, $\chi 2(33) = 35.76$, $p = .340$, RMSEA = .025, CFI = .982. The metric model did not fit worse than the configural model, $\chi^2(4) = 6.780$, $p = .148$, and all loadings were significant.

A scalar model was then estimated by constraining both indicator and intercept loadings across age. To scale the latent factors in the scalar model, the factor mean and variance at 36months were set to 0.0 and 1.0, respectively. The scalar model fit well, $\chi 2(37) = 43.403$, $p = .217$, RMSEA = .036, CFI = .957. The scalar model did not fit worse than the metric model, $\chi^2(4) = 7.466$, $p = .113$, indicating that the measurement of focused attention using Token Sort, Lock Box, Toy Play, Secondary Caregiver-Report CBQ FA, and Observer-Report Leiter scores were invariant from 36 months to 42 months.

Fit comparisons for invariance models are presented in Table 7. The final scalar model is presented in Figure 1. Note that in the final scalar model, the covaried residuals across age were all non-significant ($p > .05$), and are not included. Focused attention significantly increased with age (Factor Mean at 36 months fixed at 0.000; Factor mean at 42 months = 0.629, $p < .001$). The latent factors of focused attention were also stable over this time period, $r = .738$, $p < .001$, indicating that children with relatively high focused attention at 36 months were likely to have higher focused attention at 42 months.

Discussion

The purpose of the current study was to better understand the development of focused attention in toddlers in light of differences in definitions of focused attention used in past research (e.g., Ruff, Capozzoli, & Weissberg, 1998) and different measurement approaches. In this study, we investigate focused attention behaviors of toddlers using ratings of attention by parent and secondary caregivers and ratings by observers based on behaviors during a series of 17 laboratory tasks. These rating measures are compared with performance on laboratory tasks that were scored based on instances in which both eyes and hands were on the task materials and behaviors were consistent with task instructions (e.g., with eyes and hands on the toys). The research questions focused on whether performance based measures of focused attention provide the same information as that based on more global ratings and whether there is stability of focused attention across measurement times.

There are four main findings. First, parent-reported Child Behavior Questionnaire (CBQ) scores were not strongly related to either the two other ratings of attention (secondary caregivers and observers) or the three performance based measures (Token Sort, Lock Box and Toy Play) across the three measurement ages. Second, when the parent-report variable was excluded from analyses, a single factor model explained the common variance between indicators of focused attention better than a two-factor solution. Third, by ages 36 and 42months but not at 30 months, the indicators of focused attention do measure a common,

coherent single factor. Finally, the focused attention factor omitting the parent-report measure shows measurement invariance between ages 36 and 42 months, which indicates the factor structure, relation of indicators to the factor, and factor scale are consistent over time. These findings are each discussed below.

The absence of strong correlations between the parent-report scores with the other ratings based and performance based measures of focused attention is not completely unexpected. Indeed, other researchers have reported that correlations between parent reports and those of secondary caregivers on different measures of temperament, including the CBQ, in toddler and preschool samples are low to moderate (Goldsmith, Rieser-Danner & Briggs, 1991; Northam, Prior, Sanson & Oberklaid, 1987; Rudasill et al, 2014). However, it has been thought that parent reports might be more reflective of a range of toddlers' focused attention behaviors due to those behaviors being observed in "real world" settings that are more varied in context and structure (Anderson et al., 2002; Brocki & Bohlin, 2006; Lee, 2009; Toplak et al., 2013). In contrast, performance based measures of focused attention in structured laboratory settings measure specific and probably more limited aspects of focused attention, as do ratings provided by the secondary caregivers based on observations in babysitting, day care, or school-like settings and the ratings based on observations of behaviors during a series of lab tasks. From this "settings" point of view, highly demanding or behavior restricting settings such as the laboratory or school may require a different set of focused attention skills than other settings, such as the home. Rothbart and Bates (2006) noted that observed differences in parent versus other ratings of temperament may be due to context as well as reference group; teachers have experience with a large number of current and former students with which to compare a single child's behavior, whereas parents likely have only a few. Therefore, teachers, secondary caregivers, or researchers interacting with many children in the same age range may be giving more normative ratings. In this vein, our study supports extant research, with parent ratings providing a demonstrably different measurement of focused attention, while other ratings and performance based measures providing more similar measures of focused attention. Another explanation for the lack of association between parent and caregiver reported focused attention may relate to the scales themselves. Internal consistency of parent-reported focused attention was below .70 for each age, whereas the internal consistency of secondary caregiver reports were a more cohesive measure of focused attention (α range from .75 to .82). Putnam and Rothbart (2006) report similar reliability among a sample of 3-year olds ($\alpha = .70$), so such modest estimates of internal consistency may be expected from this measure. However, lower reliability reflects greater measurement error, and therefore correlations with less reliable scales tend to be smaller than more internally consistent measures (Goodwin & Leech, 2006).

The difference between parent ratings of focused attention and the other measures is reflected in the results of the correlations as well as in the results of the factor analyses. The parent ratings do not load strongly on a latent factor. However, when the parent reports were excluded from the model, the other measurement sources (secondary caregiver ratings, observer ratings, and the three behavioral tasks) all loaded well on a single factor. Interestingly, the model fit statistics at 30 months did not indicate a common, coherent single factor, whereas by age 36 and 42 months the indicators of a single factor were stronger, as reflected in the measurement invariant indicators. Thus, stability in the latent focused

attention factor was present only from age of 36 to 42 months. This finding can be interpreted in light of previous empirical findings of developmental changes in effortful control between three and five years of age. Older children at older ages are theorized to be better able to resolve attention conflicts and perform better on attention tasks (Rothbart et al., 2008). The stability of the latent focused attention factor at the two older ages in the present study may be reflecting greater consistency in toddlers' skills. Differences in performance at 30 months compared 36 and 42 months may be related to maturational changes in brain structure and function. Brain processing during cognitive tasks involves neural networks distributed across different brain regions that must be recruited to process the different components of tasks, for example the visual, language, problem solving, and inhibition components. Differences in the specific brain regions recruited across time during task performance and in the level of processing effort have been found to differentiate between levels of task performance (Molfese et al., 2010). The single factor solution found to best fit the data for the ages tested in the present study could change if we extended the age range tested to 48- or 60-months. Such as extension could allow for an investigation into whether there is a continuation of a single factor model underlying the data from children at these older ages or whether multi-factor models are present. Such multiple factors might be linked to Stanovich's (2009) framework posting the presence of algorithmic abilities and reflective abilities as underlying children's focused attention.

The models suggest that there is a meaningful dimension of focused attention among laboratory task measures, especially at 36 and 42 months. However, many of the loadings on the factors are lower than would be seen in more conventional applications of factor analysis, say using a set of questionnaire measures. The bivariate correlations among the measures are correspondingly small. Evidently, a child's performance from task to task was likely to be rather inconsistent, despite the presence of a latent trait of attentional focus across the tasks. This could reflect both developmental stage—children's performance levels being inconsistent from moment to moment in toddlerhood, as well as task differences—such as whether the experimenter was present during the task or not (present during Toy Play, not for the other tasks), or whether there was a visible, material incentive (as in the Lock Box, but not for the other tasks). Further research is needed to explore how differences in attention task demands influence performance scores of toddlers.

Limitations and Future Directions

The present study provides evidence of the relations between ratings of focused attention and performance based measures useful for understanding the development of focused attention in toddlers. Some limitations in the current study must be acknowledged. First, our sample is mainly representative of middle and higher income level White families, thus limiting generalizability. Future research should examine whether the relation between measures of focused attention differs in lower income or more diverse populations. Second, although most toddlers were from intact families and we encouraged fathers/partners to complete the CBQ, we did not have many fathers/partners who provided data. Future research also would benefit from getting ratings from more fathers, to complement the ratings from mothers. (Mayfield et al., 2016).

Conclusions

The current study contributes the existing literature in several important ways. Parent-reports of focused attention, while frequently used sources of temperament information, do not align with the other sources at the ages measured. This finding should encourage researchers to use multiple methods for obtaining data on focused attention. Not only do single measures have inherent risks of measurement error, but some measures that have been commonly used in studies of focused attention and other self-regulation skills such as parent-report measure did not load on the same construct as the other five measures. Thus, the parent report measure used in this study lacked evidence of the criterion-related validity needed to support its use by itself as an indicator of focused attention. This study demonstrates, based on the measurement indicators for other measures used in the study, that focused attention as a construct is coherent and consistent in its structure from 36-months and 42-months but not at 30-months. Although focused attention is likely present at 30 months and improves across toddlerhood, additional explorations of measurement approaches are needed to establish the presence of the construct at 30 months.

Funding Information:

The research reported in this publication was supported by The Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number 1R01HD073202–01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Anderson V (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. Neuropsychological Rehabilitation, 8, 319–349.

Allan NP, Hume LE, Allan DM, Farrington AL, & Lonigan CJ (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. Developmental psychology, 50(10), 2368–2379. [PubMed: 25069051]

Anderson VA, Anderson P, Northam E, Jacobs R, & Mikiewicz O (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. Child Neuropsychology, 8, 231–240. [PubMed: 12759820]

Brocki KC, & Bohlin G (2006). Developmental change in the relation between executive functions and symptoms of ADHD and co-occurring behaviour problems. Infant and Child Development, 15, 19–40.

Brown TA (2006). Confirmatory factor analysis for applied research. New York: Guilford Press

Browne MW, & Cudeck R (1992). Alternative ways of assessing model fit. Sociological Methods and Research, 21, 230–258. doi:10.1177/0049124192021002005

Burnham KP & Anderson DR (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33, 261–304.

Calkins SD (2002), Does aversive behavior during toddlerhood matter? The effects of difficult temperament on maternal perceptions and behavior. Infant Mental Health Journal, 23, 381–402. doi: 10.1002/imhj.10024

Choudhury N & Gorman KS (2000). The relationship between sustained attention and cognitive performance in 17–24-month old toddlers. Infant and Child Development, 9, 127–146.

Cuevas K, & Bell MA (2014). Infant Attention and Early Childhood Executive Function. Child Development, 85(2), 397–404. doi:10.1111/cdev.12126 [PubMed: 23711103]

Downer JT, Booren LM, Lima OK, Luckner AE, & Pianta RC (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for

observing preschoolers' competence in classroom interactions. Early Childhood Research Quarterly, 25, 1–16. [PubMed: 23175598]

Eisenberg N, Cumberland A, Spinrad TL, Fabes RA, Shepard SA, Reiser M, et al. (2001 The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. Child Development, 72, 1112–1134. [PubMed: 11480937]

Else-Quest NM, Hyde JS, Goldsmith HH, & Van Hulle CA (2006). Gender differences in temperament: A meta-analysis. Psychological Bulletin, 132, 33–72. [PubMed: 16435957]

Datavyu Team (2014). Datavyu: A Video Coding Tool Databrary Project, New York University URL http://datavyu.org.

Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, … Japel C (2007). School readiness and later achievement. Developmental Psychology, 43(6), 1428–1446. [PubMed: 18020822]

Bronson MB (2000). Self-Regulation in Early Childhood: Nature and Nurture. New York: The Guilford Press.

Fletcher JM (1998) Attention in Children: Conceptual and Methodological Issues. Child Neuropsychology, 4, 81–86.

Gaertner BM, Spinrad TL & Eisenberg N (2008). Focused attention in toddlers: measurement, stability, and relations to negative emotion and parenting. Infant and Child Development, 17, 339–363. doi:10.1002/icd.580 [PubMed: 19112517]

Goldsmith HH, Rieser-Danner LA, & Briggs S (1991). Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. Developmental Psychology, 27(4), 566–579.

Graziano PA, Calkins SD, & Keane SP (2011). Sustained attention development during the toddlerhood to preschool period: associations with toddlers' emotion regulation strategies and maternal behaviour. Infant & Child Development, 20(6), 389–408. doi:10.1002/icd.731 [PubMed: 22121338]

Goldsmith HH, & Rothbart MK (1999). The Laboratory Temperament Assessment Battery (Locomotor Version Edition, 3.1) Madison, WI: University of Wisconsin-Madison

Gunnar MR, Sebanc AM, Tout K, Donzella B and van Dulmen MMH (2003), Peer rejection, temperament, and cortisol activity in preschoolers. Developmental Psychobiology, 43, 346–368. doi:10.1002/dev.10144 [PubMed: 15027418]

Goodwin L, & Leech N (2006). Understanding Correlation: Factors That Affect the Size of r. The Journal of Experimental Education, 74(3), 251–266.

Gunnar MR, Tout K, de Haan M, Pierce S, & Stansbury K (1997). Temperament, social competence, and adrenocortical activity in preschoolers. Developmental Psychobiology, 31, 65–85. [PubMed: 9222117]

Harman C, Posner MI, Rothbart MK, & Thomas-Thrapp L (1994). Development of orienting to objects and locations in human infants. Canadian Journal of Experimental Psychology, 48(2), 301–318. [PubMed: 8069287]

Kochanska G & Knaack A (2003), Effortful Control as a Personality Characteristic of Young Children: Antecedents, Correlates, and Consequences. Journal of Personality, 71, 1087–1112. doi: 10.1111/1467-6494.7106008 [PubMed: 14633059]

Lee E-Y. (2009). An Exploratory Analysis of the Ecological Validity of a Performance-Based Assessment of Attention. Unpublished Dissertation, University of South Florida Retrieved on March 2017 from http://scholarcommons.usf.edu/etd/3687

Lemery KS, Goldsmith HH, Klinnert MD, & Mrazek DA (1999). Developmental models of infant and childhood temperament. Developmental Psychology, 35(1), 189–204. [PubMed: 9923474]

MacCallum RC, Browne MW, & Sugawara HM (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130–149.

Matthews G, Warm JS, Reinerman LE, Langheim LK, & Saxby DJ (2010). Task engagement, attention and executive control In Gruszka A, Matthews G, & Szymura B (Eds.), Handbook of individual differences in cognition: Attention, memory, and executive control (pp. 205–230). New York, NY: Springer.

Mayfield AR, Parke KA, … Allen DN, (2016). Equivalence of mother and father ratings of ADHD in children. Child Neuropsychology, 1–18, Published Online. Doi: 10.1080/09297049.2016.1236186

Molfese V, Molfese P, Molfese D, Rudasill K, Armstrong N, & Starkey G (2010). Executive function skills of 6 to 8 year olds: Brain and behavioral evidence and implications for school achievement. Contemporary Educational Psychology: Special Issue on Brain and Academic Development, 30, 116–153

Northam E, Prior M, Sanson A, & Oberklaid F (1987). Toddler temperament as perceived by mothers versus day-care-givers. Merrill-Palmer Quarterly, 33, 213–229.

NICHD Early Child Care Research Network (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. Developmental Psychology, 41(1), 99–114. doi:10.1037/0012-1649.41.1.99 [PubMed: 15656741]

Posner MI, & Rothbart MK (2007). Research on attentional networks as a model for the integration of psychological science. Annual Review of Psychology, 58, 1–23

Putnam SP, & Rothbart MK (2006). Development of short and very short forms of the Children's Behavior Questionnaire. Journal of Personality Assessment, 87(1), 103–113. doi:10.1207/s15327752jpa8701_09.

Putnam SP, Rothbart MK and Gartstein MA (2008), Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. Infant and Child Development, 17, 387–405. doi:10.1002/icd.582

Reck SG, & Hund AM (2011). Sustained attention and age predict inhibitory control during early childhood. Journal of Experimental Child Psychology, 108, 504–512. [PubMed: 20801457]

Roid GH & Miller LJ (1997). Leiter International Performance Scale-Revised. Wood Dale, IL: Stoelting Col.

Roid GH, Pomplun M, & Martin JJ (2009). Nonverbal intellectual and cognitive assessment with the Leiter International Performance Scale-Revised (Leiter-R) Practitioner's Guide to Assessing Intelligence and Achievement. Naglieri JA & Goldstein S

Rothbart MK (2011). Becoming who we are: Temperament and personality in development. New York, NY: The Guilford Press.

Rothbart MK, Ahadi SA, Hershey KL, & Fisher P (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. Child Development, 72(5), 1394–1408. [PubMed: 11699677]

Rothbart MK, Posner MI, & Kieras J (2006). Temperament, attention, and the development of self-regulation In McCartney K & Phillips D (Eds.), Blackwell handbook of early childhood development (pp. 338–357). Malden, MA: Blackwell Publishing.

Rudasill KM, Prokasky A, Tu X, Frohn S, Sirota K, & Molfese VJ (2014). Parent vs. teacher ratings of children's shyness as predictors of early language and attention skills. Learning and Individual Differences, 34, 57–62. doi: 10.1016/j.lindif.2014.05.008

Ruff HA, & Capozzoli MC (2003). Development of attention and distractibility in the first 4 years of life. Developmental Psychology, 39, 877–890. [PubMed: 12952400]

Ruff HA, Capozzoli M, Saltarelli LM (1996). Focused visual attention and distractibility in 10-month-old infants. Infant Behavior & Development, 19, 281–293.

Ruff HA, Capozzoli M, & Weissberg R (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. Developmental Psychology, 34, 454–64. [PubMed: 9597356]

Ruff HA, & Lawson KR (1990). Development of sustained, focused attention in young children during free play. Developmental Psychology, 26(1), 85–93.

Ruff HA, & Rothbart MK (1996). Attention in early development: Themes and variations. New York: Oxford University Press.

Selig JP, & Little TD, (2012 Autoregressive and cross-lagged panel analysis for longitudinal data In Laursen B, Little TD, & Card NA (Eds). Handbook of Developmental Research Methods. New York, NY: Guilford.

Spaulding TJ, Plante E, & Vance R (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. Journal of Speech, Language, and Hearing Research, 51(1), 16–34. DOI: 10.1044/1092-4388(2008/002)

Stanovich KE (2009). Distinguishing the reflective, algorithmic, and autonomous minds: It is time for a tri-process theory? In Evans J.St. B.T. & Frankish K (Eds.), In two minds: Dual processes and beyond (pp. 55–88). New York: Oxford University Press.

Stanovich KE (2011). Rationality and the reflective mind. New York: Oxford University Press.

Teglasi H, Schussler L, Gifford K, Annotti LA, Sanders C, & Liu H (2015). Child behavior questionnaire–Short form for teachers: Informant correspondences and divergences. Assessment, 22(6), 730–748. [PubMed: 25573857]

Toplak ME, West RF & Stanovich KE (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? Journal of Child Psychology and Psychiatry, 54, 131–143. doi:10.1111/jcpp.12001 [PubMed: 23057693]

Twardosz S (2012). Effects of Experience on the Brain: The Role of Neuroscience in Early Development and Education. Early Education and Development, 23(1), 96–119.

Warner-Rogers J, Taylor A, Taylor E, & Sandberg S (2000). Inattentive behavior in childhood: Epidemiology and implications for development. Journal of Learning Disabilities, 33, 520–536. [PubMed: 15495395]

Wolf EJ, Harrington KM, Clark SL, & Miller MW (2013). Sample size requirements for structural equation models: An valuation of power, bias, and solution propriety. Educational & Psychological Measurement, 73(6), 913–934. doi:10.1177/0013164413495237

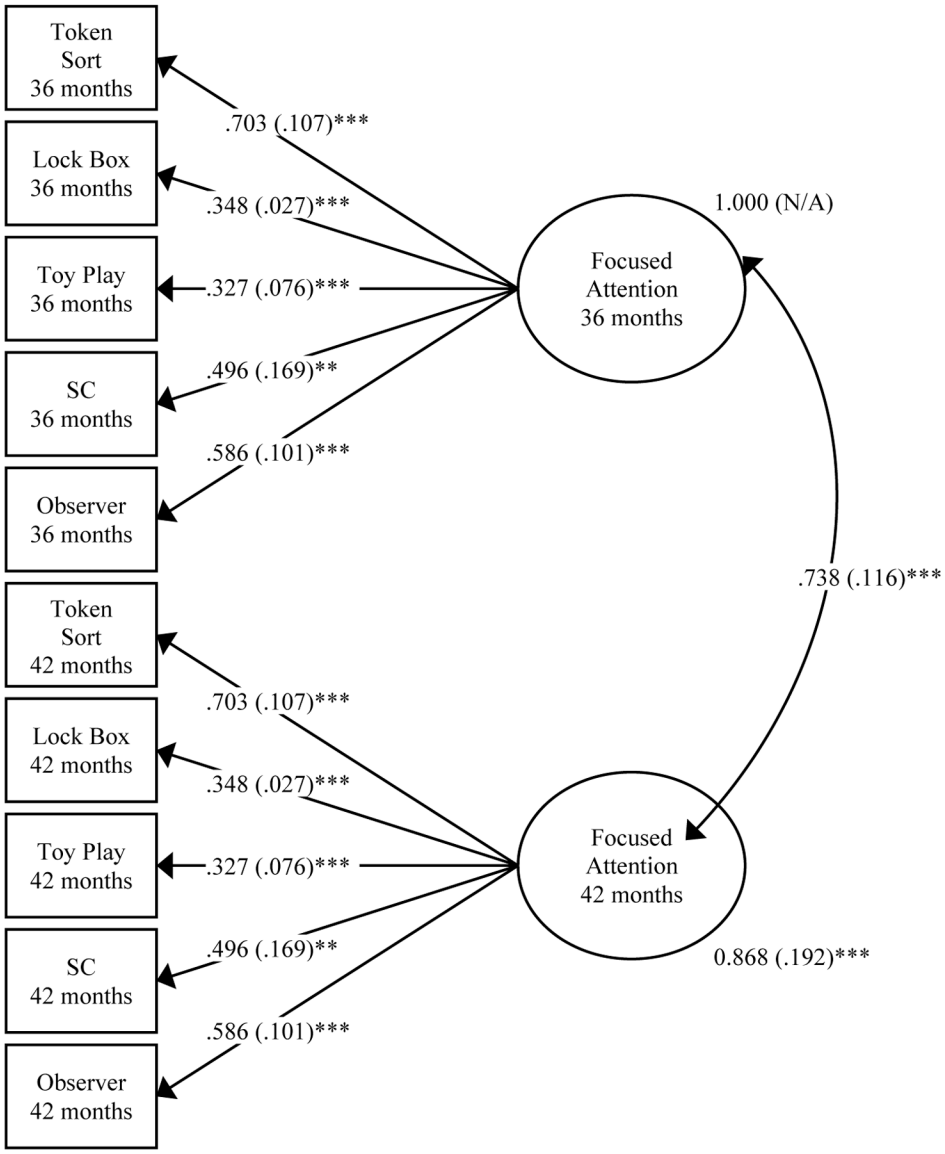

Figure 1.

One-factor scalar invariance model at 36 and 42 months (unstandardized parameters).

** $p < .01$, *** $p < .001$. SC= Secondary Caregiver

Table 1

Descriptive Statistics of Study Variables

Variable	30 months			36 months			42 months			
	M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	Difference[τ]
Ratings of Attention										
CBQ Parent	4.72	.90	2.33–6.67	4.70	.84	2.67–6.83	4.81	.84	2.50–6.83	NS
CBQ-Secondary Caregiver	4.60	1.03	2.17–6.50	4.71	1.26	1.83–7	4.75	1.11	2.50–7	NS
Leiter-R	1.94	.70	.10–3	2.32	.62	.40–3	2.52	.55	1–3	30M < 36M[*] 36M < 42M[*] 30M < 42M[*]
Performance Based Attention										
Token Sort	.23	.32	0–1	.59	.39	0–1	.77	.32	0–1	30M < 36M[*] 36M < 42M[*] 30M < 42M[*]
Toy Play	.64	.26	0–1	.76	.22	0–1	.83	.22	0–1	30M < 36M[*] 36M < 42M[+] 30M < 42M[*]
Lock Box	.51	.25	0–1	.58	.24	04–1	.65	.22	06–1	30M < 36M[+] 36M < 42M[+] 30M < 42M[*]

Note. NS= Not significant; M= Months; CBQ= Children's Behavior Questionnaire;

[τ]Difference was calculated using paired sample t-tests.

[*] $p < .001$

[+] $p < .05$

Table 2

Intercorrelations Among Study Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.CBQP30	-																	
2.CBQP36	.65**	-																
3.CBQP42	.61**	.73**	-															
4.CBQ30S	.19	.35**	.48**	-														
5.CBQ36S	.13	.26	.43**	.55**	-													
6.CBQ42S	.05	.37**	.26	.53**	.44**	-												
7.LO30	.14	.21*	.18	.28*	.53**	.28*	-											
8.LO36	.05	.09	.16	.35**	.48**	.28*	.58**	-										
9.LO42	−.04	.08	.13	.38**	.50**	.28*	.58**	.63**	-									
10.TS30	.15	−.02	.04	.01	.17	.24	.29**	.24**	.11	-								
11.TS36	.01	−.01	.02	.08	.43**	.29*	.24**	.37**	.32**	.26**	-							
12.TS42	−.03	−.03	−.07	.16	.12	.14	.26**	.25**	.45**	.18	.37**	-						
13.TP30	−.02	.10	−.01	.38**	.03	.09	.29**	.13	.26**	−.02	.06	.30**	-					
14.TP36	.09	.06	.03	.16	.09	.17	.15	.22*	.18	.15	.20*	.19*	.32**	-				
15.TP42	−.01	.11	.03	−.03	.12	.09	.32**	.23*	.29**	.08	.13	.09	.10	.14	-			
16.LB30	−.04	−.04	−.12	.07	−.01	−.14	.18*	.13	.06	−.02	.01	.11	.11	−.03	.09	-		
17.LB36	−.01	−.04	.03	.06	.05	.15	.09	.14	.16	.07	.14	.06	.15	.13	−.13	.19*	-	
18.LB42	.16	.15	.18	.21	.19	.22	.19*	.23*	.35**	.17	.13	.19*	.13	.24*	.25**	.09	.24*	-

Note.

*p < .05,

**p < .01 (two-tailed). CBQ= Children's Behavior Questionnaire; P= Parent; S= Secondary Caregiver; LO= Leiter-R Observer; TS= Token Sort; TP= Toy Play; LB= Lock Box. Gender (1= female, 0= male

Table 3

Single- And Two-Factor CFA Models for Focused Attention Variables

	Single Factor Models								
	30 months			36 months			42 months		
Indicator	**λ**	**SE**	***p***	**λ**	**SE**	***p***	**λ**	**SE**	***p***
Token Sort	.296	.084	<.001	.714	.150	<.001	.509	.109	<.001
Lock Box	.189	.098	.053	.204	.109	.061	.382	.113	.001
Toy Play	.291	.164	.076	.254	.094	.007	.293	.116	.012
CBQ Focused Attention - Parent	.144	.107	.181	.109	.113	.335	.164	.123	.182
CBQ Focused Attention – SC	.276	.190	.147	.822	.178	<.001	.371	.184	.044
Leiter Focused Attention - Observer	.981	.339	.004	.600	.108	<.001	.675	.124	<.001
	Two Factor Models								
	30 months			36 months			42 months		
Factor / Indicator	**λ**	**SE**	***p***	**λ**	**SE**	***p***	**λ**	**SE**	***p***
Task Factor									
Token Sort	.117	.323	.717	.824	.233	<.001	.497	.121	<.001
Lock Box	.060	.192	.756	.224	.114	.049	.375	.114	.001
Toy Play	.121	.364	.739	.273	.100	.006	.286	.120	.017
Rating Factor									
CBQ Focused Attention - Parent	.164	.127	.197	.118	.110	.285	.166	.125	.184
CBQ Focused Attention – SC	.409	.139	.003	.844	.195	<.001	.386	.190	.042
Leiter Focused Attention - Observer	.697	.177	<.001	.629	.107	<.001	.645	.121	<.001

Table 4

Fit Comparison of One- And Two-Factor CFA Models

	Absolute Fit					Relative Fit				
Model	**χ^2**	**df**	**p**	**RMSEA**	**CFI**	**AIC**	**BIC**	**−2LL**	**Δ−2LLχ^2**	**p**
30 months two-factor	14.70	8	.065	.075	.828	2115.2	2172.2	2077.2		
30 months one-factor	18.22	9	.033	.083	.764	2116.4	2170.4	2080.4	3.57	.059
36 months two-factor	5.31	8	.724	<.001	1.000	1859.1	1914.0	1821.1		
36 months one-factor	6.03	9	.737	<.001	1.000	1858.2	1910.2	1822.2	0.70	.403
42 months two-factor	11.94	8	.154	.065	.914	1566.7	1619.4	1528.7		
42 months one-factor	12.06	9	.210	.054	.933	1564.9	1614.7	1528.9	0.12	.731

Table 5

Revised Single- and Two-Factor CFA Models for Focused Attention Variables (Without Parent Report CBQ FA)

	Single Factor Models					
	30 months		**36 months**		**42 months**	
Factor Loadings	**β**	***p***	**β**	***p***	**β**	***p***
Token Sort Engagement	.290	.001	.731	.000	.515	.000
Lock Box Engagement	.188	.118	.208	.055	.374	.001
Toy Play Engagement	.281	.190	.254	.007	.292	.013
CBQ Focused Attention - SC	.265	.237	.820	.000	.357	.035
Leiter Focused Attention - Observer	1.012	.036	.592	.000	.678	.000
	Two Factor Models					
	30 months		**36 months**		**42 months**	
Factor Loadings	**β**	***p***	**β**	***p***	**β**	***p***
Task Factor						
Token Sort Engagement	.129	.602	.830	.000	.504	.000
Lock Box Engagement	.076	.674	.224	.047	.366	.001
Toy Play Engagement	.154	.639	.269	.006	.284	.017
Rating Factor						
CBQ Focused Attention - SC	.422	.001	.845	.000	.373	.031
Leiter Focused Attention - Observer	.629	.000	.623	.000	.628	.000

Note. SC= Secondary Caregiver. CBQ= Children's Behavior Questionnaire. FA= Focused Attention

Table 6

Fit Comparison of Revised Single- And Two-Factor CFA Models (Without Parent Report CBQ FA)

	30 months		*36 months*		42 months	
Absolute Fit	**Single Factor**	**Two Factor**	***Single Factor***	**Two Factor**	**Single Factor**	**Two Factor**
Chi-Square	15.424*	9.094*	*3.443**	2.895*	3.469*	3.202*
df	5	4	*5*	4	5	4
p	.009	.059	*.632*	.576	.628	.525
RMSEA	.119	.093	*.000*	.000	.000	.000
CFI	.721	.864	*1.000*	1.000	1.000	1.000
Relative Fit						
−2LL	1698.17	1694.28	*1518.74*	1517.79	1257.49	1257.31
Δ−2LL	11.93		*0.59*		0.20	
p	.001		*.441*		.654	
AIC	1728.17	1726.29	*1548.74*	1549.79	1287.49	1289.32
BIC	1773.12	1774.24	*1591.87*	1595.80	1328.67	1333.23

Note. CBQ= Children's Behavior Questionnaire. FA= Focused Attention

Table 7

Longitudinal Measurement Invariance Testing Models Without Parent Ratings At 36 And 42 Months

	Absolute Fit					Relative Fit					
Model	**χ^2**	**df**	***p***	**RMSEA**	**CFI**	**AIC**	**BIC**	**−2LL**	**Δ−2LL χ^2**	**df**	***p***
Configural (all parameters free)	28.454	29	.494	.000	1.000	2768.9	2873.3	2696.9	-	-	-
Metric (loadings constrained)	35.756	33	.340	.025	.982	2769.3	2862.1	2705.3	6.780	4	.148
Scalar (loadings + intercepts constrained)	43.403	37	.217	.036	.957	2769.7	2850.8	2713.7	7.466	4	.113