Journal of Psychoeducational Assessment
1–21

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0734282916633009
jpa.sagepub.com
SAGE

An Early Years Toolbox for Assessing Early Executive Function, Language, Self-Regulation, and Social Development: Validity, Reliability, and Preliminary Norms

Steven J. Howard[1] and Edward Melhuish[1,2,3]

Abstract

Several methods of assessing executive function (EF), self-regulation, language development, and social development in young children have been developed over previous decades. Yet new technologies make available methods of assessment not previously considered. In resolving conceptual and pragmatic limitations of existing tools, the Early Years Toolbox (EYT) offers substantial advantages for early assessment of language, EF, self-regulation, and social development. In the current study, results of our large-scale administration of this toolbox to 1,764 preschool and early primary school students indicated very good reliability, convergent validity with existing measures, and developmental sensitivity. Results were also suggestive of better capture of children's emerging abilities relative to comparison measures. Preliminary norms are presented, showing a clear developmental trajectory across half-year age groups. The accessibility of the EYT, as well as its advantages over existing measures, offers considerably enhanced opportunities for objective measurement of young children's abilities to enable research and educational applications.

Keywords

self-regulation, executive function, language, social, development, early years

Introduction

In early childhood, the foundations of later personal, social, and cognitive functioning are established. The first 2 years of life are crucial for sensorimotor development. From the first to the fifth year of life, the basic functions of language and cognition are formed. Cognitive and emotional control systems emerge around the age of three and, although these systems continue developing into adulthood, the foundations laid in early childhood exert strong influence in almost every domain of psychological functioning and behavior later in life (Crone & Dahl, 2012; Moffitt et al., 2011). For example, evidence supports the proposition that a cognitively

[1]University of Wollongong, New South Wales, Australia
[2]University of Oxford, UK
[3]Birkbeck, University of London, UK

Corresponding Author:
Steven J. Howard, School of Education, University of Wollongong, New South Wales, 2522, Australia.
Email: stevenh@uow.edu.au

stimulating environment is associated with better cognitive, language, executive function (EF), and achievement outcomes (Adi-Japha & Klein, 2009; Lugo-Gil & Tamis-LeMonda, 2008; NICHD Early Child Care Research Network, 2003, 2005; Rhoades, Greenberg, Lanza, & Blair, 2011). Brain research has also revealed early sensitive periods in which the developing brain is maximally susceptible to environmental influences (Shonkoff & Phillips, 2000; Sirois et al., 2008). The negative side of the brain's susceptibility is that low-supportive and deprived environments in early childhood may have potentially irreversible consequences for development and later functioning (Hackman & Farah, 2009). Moreover, it has become increasingly clear that some aspects of early development are predictive of longer term outcomes.

Aspects of Development That Are Particularly Predictive of Later Outcomes

Within the early years, research has highlighted some aspects of development that appear particularly predictive of later outcomes. For instance, young children vary in the rate of acquisition of words (expressive vocabulary; for example, Fenson et al., 1994) and such differences have an impact on later language (Rowe, Raudenbush, & Goldin-Meadow, 2012), literacy, and educational success (Dickinson & Tabors, 2001; Duncan et al., 2007; Snow, Burns, & Griffin, 1998). A weak vocabulary leaves children with a smaller reserve of sound and word knowledge and increased difficulties with decoding real words when they read (Wellman et al., 2011). Vocabulary is vital for comprehension, writing, and content-area learning, and is an area of weakness for poor readers (Snow, 2011).

It is clear that early experience in the home influences language development. Hart and Risley (1995) have shown that by the age of 4, children from disadvantaged backgrounds may have been exposed to as many as 30 million fewer words than children from advantaged backgrounds. Quality as well as quantity of language experience varies as advantaged parents not only use more words but also show more diverse vocabulary with more complex syntax than disadvantaged parents (Hoff, 2003; Hoff-Ginsberg, 1991; Huttenlocher, Vasilyeva, Waterfall, Vevea, & Hedges, 2007; Pan, Rowe, Singer, & Snow, 2005). Thus, children from disadvantaged backgrounds are more likely to have less extensive language exposure, which may influence their expressive vocabulary and put them at risk of subsequent literacy difficulties. In fact, children's expressive vocabulary on entering school is a good indicator of school readiness, as well as later educational success (e.g., Snow et al., 1998). Thus, knowing the child's early vocabulary improves our ability to understand factors that may help identify children at risk of poor language, literacy, and educational outcomes. It may also provide understanding that can guide early intervention to enable children to enter school at less of a disadvantage.

From birth onwards, children's social development is similarly influenced by their daily interactions with parents, carers, family members, child care, preschool, and school staff, as well as with their peers. Moreover, by actively interacting with others, children affect the ways that adults and their peers relate to them. The transactional nature of social development means that early social development will influence longer term social development, as well as other aspects of development. As substantial knowledge of a child is necessary to judge their social development, measures of children's social development often are based on questionnaire ratings by adults familiar with the child (e.g., parent, carer, educator; Goodman, 1997; Hogan, Scott, & Bauer, 1992). Using such measures, it has been found that parenting behaviors influence children's social development from the early years (Daglar, Melhuish, & Barnes, 2011). Child care and preschool experience are also associated with social development, including sociability, externalizing (antisocial) and internalizing (anxious or depressed) behavior, as well as prosocial behavior (e.g., sharing, showing empathy; Lamb & Ahnert, 2006; Melhuish & Barnes, 2012). Such effects of early experience influence social development not only in the preschool period but also in the long term (Vandell et al., 2010). While social skills blossom in the toddler and

preschool period, they continue to develop along at least partly predictable trajectories through childhood (Hughes, 2011). For example, later social-emotional intelligence is linked to earlier peer relationships (Slaughter, Imuta, Peterson, & Henry, 2015). In addition, social development may also influence other aspects of subsequent development. For example, the quality of preschoolers' interactions with their parents is associated with later language and literacy (Reese, Sparks, & Leyva, 2010). Hence, knowledge of early social development aids understanding not only of future social outcomes but also other aspects of development, as social abilities underpin interactions and communication and their role in later learning and development.

Another aspect of development that appears to be highly predictive of later outcomes is self-regulation, which refers to the ability to control one's thoughts, behaviors, emotional reactions, and social interactions, even when impulses and urges run contrary to proximal or distal goals. The capacity to self-regulate develops rapidly in the first 5 years of life (Blair, 2002; Galinsky, 2010), with far-reaching implications for later development. In a longitudinal study of 3,000 children in the United Kingdom, for instance, self-regulation at age 5 was highly predictive of literacy and numeracy abilities more than 5 years later, even after allowing for the impact of child and family demographics (Melhuish et al., 2007). Moreover, young children who engage in intentional self-regulation tend to learn more and go further in their education (Blair & Diamond, 2008). By contrast, children who cannot effectively regulate (e.g., their emotions) tend to move away from, rather than engage in challenging learning activities. It is therefore unsurprising that early self-regulation is closely related to school engagement and meta-cognition (Boeckaerts & Corno, 2005). Self-regulation is also related to a range of social competencies including prosocial behavior, competence to collaborate, and empathy (Kochanska, Murray, & Harlan, 2000).

Certain cognitive capacities contribute to the ability to self-regulate, allowing the child to hold goals and other task-relevant information in mind (working memory [WM]), resist distraction and impulses in the task of goal attainment (inhibition), and flexibly shift attention if information is no longer goal-relevant (shifting). These cognitive capacities are routinely bundled as "executive functions" (EFs) given their overarching function of cognitive control. Besides being contributors to self-regulatory development (Hofmann, Schmeichel, & Baddeley, 2012), EFs have been found to be an even stronger predictor of school achievement than IQ (Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; St. Clair-Thompson & Gathercole, 2006; Van der Ven, Kroesbergen, Boom, & Leseman, 2012). EFs are predictive of children's school readiness (Blair & Razza, 2007; Welsh, Nix, Blair, Bierman, & Nelson, 2010), early literacy and numeracy skills (Blair & Razza, 2007), and general learning ability (Bull et al., 2008). EFs are also predictive of children's social and emotional development (e.g., social understanding and moral conduct; Riggs, Jahromi, Razza, Dillworth-Bart, & Müeller, 2006).

The development of self-regulation, like executive functioning, is influenced by early experience both in the home and outside the home (e.g., child care; Melhuish et al., 2007). Children from disadvantaged backgrounds are more likely to begin school with weaker self-regulation skills than their peers. Improving self-regulation and EFs through appropriate early education is therefore likely to be particularly beneficial for children from disadvantaged backgrounds (Miech, Essex, & Goldsmith, 2001). Hence, parents and early childhood teachers can play an important role in helping young children regulate their thinking and behaviors (Bodrova & Leong, 2007; Galinsky, 2010). This requires adjustment of children's everyday experiences to the child's current level of development, which is more effectively achieved with clear and accurate information on the child's emerging abilities.

Providing environments for developing language and self-regulation in young children is thus a major objective for early childhood education and care (ECEC). For instance, development of cognitive control and emotional self-regulation in early childhood can be promoted by peer interaction in pretend play (Berk, Mann, & Ogan, 2006; Diamond & Lee, 2011; Lillard et al., 2013). High-quality ECEC fosters the development of cognitive control and self-regulation (Diamond

& Lee, 2011; McClelland, Acock, & Morrison, 2006; Melhuish et al., 2007), as well as language development (NICHD Early Child Care Research Network, 2003, 2005; Sammons et al., 2004). Establishing effective environments, experiences, and education that foster these abilities, however, requires accurate assessment of the child's developmental level. While assessment methods already exist, current methods have important limitations and there remains scope for improvement, particularly suited to the characteristics of young (i.e., preschool and early primary school) children.

Current Issues in Early Cognitive Development Measures

Despite the importance of early cognitive development and its assessment for a broad range of subsequent outcomes, there remain both conceptual and pragmatic issues in early measurement of young children's emerging abilities. Conceptually, for instance, often there is a lack of clarity in the construct being measured. To illustrate, in the self-regulation and EF literature measures and definitions are used interchangeably by some (e.g., McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009), yet are seen as distinct by others (e.g., Hofmann et al., 2012). There also remains debate regarding the specific measures, stimuli, timing, and instructions that are required to most effectively capture these abilities in young children (e.g., Blair, Zelazo, & Greenberg, 2005). Similarly, the widespread computerization of many tasks, while effective for standardization and accurate data collection, often introduces response artifacts whenever young children are not familiar with the laptop keyboard they must navigate to respond (Howard & Okely, 2015). Given these issues, there remains little agreement on the optimal measures and their configuration (stimuli, timing, instructions, response method) for capturing important aspects of early cognitive development.

One important advance in this regard has been the introduction of the National Institute of Health (NIH) Toolbox (NIH Toolbox CB, 2013), which is a freely available and widely accessed battery of tasks for assessing key aspects of development from 3 to 15 years of age (Zelazo et al., 2013). The NIH Toolbox is a set of brief measures aiming to assess cognitive (e.g., EF, attention, memory), emotional (e.g., well-being, stress), motor (e.g., locomotion, strength), and sensory abilities (e.g., audition, vision) across the life span. In addition to its applicability to a range of research designs and extensive data norming, the NIH toolbox also overcomes at least some of the issues of accessibility and cost that plague many commercial and task-based measures—yet there remains room for improvement. For instance, the full NIH Toolbox's requirement for a computer, parallel monitor, and continuous Internet connection restricts who and where data can be collected. Many users (e.g., ECEC), such as those in areas of disadvantage or the developing world, often lack the required connectivity for such tools. The recent release of a tablet version for a subset of NIH Toolbox tasks partly addresses this issue; yet, the associated annual subscription fee may constrain users to those with a suitable research background, purpose, and funding. Furthermore, the NIH Toolbox does not appear to address additional limitations often found in existing early years measures. For example, in contrast to the basic stimuli used in the NIH Toolbox, research suggests that children are often more highly engaged (thus yielding better capture of their abilities) by dynamic and immersive stimuli (Howard & Okely, 2015). This research also highlights the potential pitfalls of nonintuitive response methods (e.g., a mouse or spacebar), especially with young children. Finally, there remain questions about how accurately young children's capacities are captured by tasks that are downward extensions of adult measures, without sufficient consideration of the unique constraints of assessing young children (Korkman, 1999; Korkman, Kirk, & Kemp, 2007). This includes young children's comparatively limited capacity, duration, and control of attentional focus (Best, Miller, & Jones, 2009), ability to understand task instructions and communicate a response (Hughes, 1998), knowledge base (Chi, 1978), and increased susceptibility to distraction (Howard, Johnson, & Pascual-Leone, 2014).

As such, there remains an opportunity for the development of measures that are more playful, specifically developed with young children in mind (e.g., brief, engaging), acceptable to children, their parents, and their caregivers, and are easily and freely accessible to all. As previously argued, such tools can be highly beneficial beyond their research applications. For instance, parents and educators may use this data in their efforts to support their child(ren)'s development. Achieving this, however, requires simultaneously addressing conceptual and pragmatic limitations of existing measures while also rethinking and democratizing early developmental measurement.

The Current Study

Given these limitations, we sought to develop, validate, and establish preliminary norms for a toolbox of early self-regulation, EF, language, and social development measures that addressed these issues. Specifically, we developed the Early Years Toolbox (EYT) comprising measures of visual-spatial and phonological WM, shifting, inhibition (routinely bundled as "EFs"), vocabulary, as well as an educator- or parent-report measure of self-regulation and social behavior. To address conceptual issues in existing measures, we sought to ensure that all EYT measures were (a) developmentally appropriate (e.g., in instructions, task requirements, response method); (b) developmentally sensitive; (c) brief (≤5 min per task); (d) engaging, valid, and reliable for use with young children (i.e., accurate capture of young children's abilities, which requires that children be sufficiently engaged in the task); (e) technologically dynamic (i.e., using visual and audio cues to communicate instructions, animating stimuli, the ability to manipulate key variables) without advantaging children with technology experience; and (f) potentially internationally applicable. To circumvent the practical limitations of existing measures, we developed this toolbox for the iPad, which ensured portability, free and easy access without need for prior permission (download from the iTunes App Store), no additional equipment requirements, and the ability to collect data without an Internet connection. To evaluate the toolbox, we combined data from five initial studies using the EYT, yielding a large sample ($N = 1{,}764$) with good representation across the levels of core population demographic characteristics (i.e., sex, socioeconomic status [SES], maternal education, Indigeneity), to examine its reliability, validity, and establish preliminary norms. The accessibility of the EYT, as well as its pragmatic and measurement advantages over existing measures, offers considerably enhanced opportunities for objective measurement of young children's abilities to enable research and educational applications that are constrained by existing methods.

Material and Method

Participants

Participants were 1,764 2½- to 5-year-old preschool and early primary school students recruited from more than 80 sites (i.e., preschool, long-day care, kindergarten) across four Australian states (i.e., NSW, VIC, SA, QLD). In planning participant recruitment, we incorporated the EYT into planned studies to ensure simultaneous collection of data on the toolbox and appropriate convergent validation measures. Samples were recruited across several Australian states in which our Institute was undertaking studies that involved samples spanning the demographic spectrum of those states. Our resultant sample consisted of 946 males and 818 females, which largely reflect the 2011 Australian census data distribution of males and females in this age group (Australian Institute of Health and Welfare [AIHW], 2012), $\chi^2(1) = 3.77$, $p = .05$. From this sample, 150 (8.5%) identified as Indigenous, which suggests that children of Aboriginal and Torres Strait Islander descent were overrepresented in this sample compared with the estimated 4.9% of the

total child population in Australia (AIHW, 2012). Home postcodes were reported for 96.6% of the sample, derived from which were "Socioeconomic Indices for Areas (SEIFA)—Advantage and Disadvantage" indices (a measure of relative socioeconomic advantage and disadvantage within a geographic area, which considers factors such as education levels, household income, and unemployment). The sample was well distributed across all SEIFA deciles, with some bias toward lower SES areas (Table 1). Parental education levels were also reported via parental survey for 66.2% of the sample, which was largely representative of 2011 Australian census data on highest level of education for persons 15 years and over (Australian Bureau of Statistics [ABS], 2012).

Whereas all participants aged 3 to 5 years were administered the EYT measures of visual-spatial WM, inhibition, and shifting, only a subset of this full sample was administered the phonological WM (n = 1,095), vocabulary (n = 1,261, which also included the 2-year-old children) and self-regulation measures (n = 414). Demographic patterns for these subsamples mirrored those of the full sample. A further subsample of those administered all toolbox tasks (n = 86) was selected to also receive a battery of existing measures for convergent validity analyses. This subsample also mirrored the larger sample in terms of age (M = 3.92, SD = 0.61), sex (44.2% female), socioeconomic profile, and maternal education.

Measures

The EYT is a freely available battery of iPad-based EF, language, self-regulation, and social development measures. Each measure was designed to be brief (≤5 min, including instruction and practice), engaging (illustrated by a children's book illustrator), and leverage the possibilities of technology (e.g., animation, audio, ease of responding, accurate capture of responses and response timings) without introducing effects of technological expertise (i.e., methods of responding were designed to be intuitive and mirror noncomputerized analogues of these measures). To standardize administration, tasks provide instructions both visually and auditorily, with supplementary information from the administrator if the child requires further clarification. All instructional and practice trials provide dynamic visual and auditory feedback to participants. Furthermore, an effort was made to minimize the literacy and numeracy demands of these tasks to mitigate effects of prior knowledge and learning on children's performance. Each of these tasks can be downloaded for free from the iTunes App Store by searching "EYT," and its training materials can be accessed from www.eytoolbox.com.au. In addition to the EYT, an additional battery of convergent validity tasks was administered to a subsample of children, drawn from the British Ability Scales (BAS; Elliott, Smith, & McCulloch, 1996), NIH Toolbox Cognition Battery and the Strengths and Difficulties Questionnaire (SDQ).

EYT "Mr. Ant" Task (Visual-Spatial WM)

This task, adapted from Case's (1985) Mr. Cucumber task and following the protocols of Morra (1994), asks participants to remember the spatial locations of "stickers" placed on a cartoon ant and identify these locations after a brief retention interval (for a screen capture, see Figure 1). Test trials increase in difficulty (i.e., WM demand) as the task progresses, with three trials at each level of complexity (progressing from one to eight stickers). All trials progress as follows: (a) Mr. Ant presented with n colored stickers (where n equals the current level of difficulty) for 5 s, (b) presentation of a blank screen for 4 s, then (c) an image of Mr. Ant without stickers—along with an auditory prompt to recall where the stickers were—presented until the participant's response is complete. Participants respond by tapping the spatial locations on Mr. Ant that they judge previously held stickers. The task continues until the earlier of completion (at Level 8, with eight spatial locations to remember) or failure on all three trials at the same level of difficulty.

Table 1. Demographic Characteristics of Sample by Age Group.

	Age groups (years:months)							Full sample	Population proportion
	2:6-2:11	3:0-3:5	3:6-3:11	4:0-4:5	4:6-4:11	5:0-5:5	5:6-5:11		
Sex									
Male	51.6%	52.4%	53.9%	54.8%	54.4%	58.0%	43.2%	53.6%	51.3%
Female	48.4%	47.6%	46.1%	45.2%	45.6%	42.0%	56.8%	46.4%	48.7%
Indigenous status									
Indigenous	6.6%	7.4%	6.8%	8.5%	9.7%	11.5%	14.8%	8.5%	4.9%
Nonindigenous	93.4%	92.6%	93.2%	91.5%	90.3%	88.5%	85.2%	91.5%	95.1%
SEIFA									
1st decile	23.0%	21.4%	21.8%	16.8%	19.8%	24.9%	36.7%	21.9%	10.0%
2nd decile	12.6%	16.0%	11.4%	6.3%	4.3%	3.6%	0.0%	9.5%	10.0%
3rd decile	16.1%	10.4%	10.7%	8.4%	10.6%	13.3%	11.4%	10.9%	10.0%
4th decile	13.8%	13.4%	10.5%	14.3%	18.8%	26.2%	30.4%	15.9%	10.0%
5th decile	3.4%	15.3%	14.3%	8.8%	12.6%	9.3%	6.3%	12.0%	10.0%
6th decile	10.3%	9.9%	10.7%	16.0%	10.6%	7.1%	5.1%	10.5%	10.0%
7th decile	4.6%	5.9%	5.9%	6.3%	5.3%	4.9%	1.3%	5.5%	10.0%
8th decile	3.4%	3.5%	6.4%	10.5%	8.7%	4.9%	3.8%	6.1%	10.0%
9th decile	9.2%	3.8%	6.8%	10.1%	7.7%	4.4%	5.1%	6.3%	10.0%
10th decile	3.4%	0.5%	1.6%	2.5%	1.4%	1.3%	0.0%	1.4%	10.0%
Maternal education									
Up to year 11	17.9%	19.2%	19.4%	21.7%	16.8%	22.6%	32.7%	20.3%	32.9%
Year 12	21.4%	16.6%	15.0%	16.7%	20.2%	20.0%	15.4%	17.0%	19.0%
Certificate	17.9%	20.1%	18.2%	15.2%	19.1%	17.4%	15.4%	18.2%	17.5%
Diploma	11.9%	16.3%	15.5%	10.1%	16.9%	14.8%	11.5%	14.7%	9.2%
Degree	21.4%	18.3%	22.9%	23.9%	15.7%	19.1%	23.1%	20.6%	15.4%
Postgraduation	9.5%	9.5%	9.1%	12.3%	11.2%	6.1%	1.9%	9.2%	6.1%

Note. All proportions have been calculated based on available (valid) data. Totals may not sum to 100% due to rounding. Population proportions for children (i.e., sex, Indigenous status) are based on 2011 Australian census data collapsed across children aged 1 to 4 and 5 to 9 years (AIHW, 2012). Parental education proportions are based on 2011 Australian census data for persons aged 15 years and over ABS (2012). SEIFA = Socioeconomic Indices for Areas.

Instruction and three practice trials serve to familiarize participants with task requirements. WM capacity was indexed by a point score (Morra, 1994) calculated as follows: beginning from Level 1, one point for each consecutive level in which at least two of the three trials were performed accurately, plus 1/3 of a point for all correct trials thereafter.

EYT "Not This" task (phonological WM). This task, loosely based on the Direction Following Task (Im-Bolter, Johnson, & Pascual-Leone, 2006), requires participants to carry out auditory instructions of increasing complexity (Figure 1). Instructions ask participants to point to a stimulus that is *not* of a particular color, shape, or size (or some combination of these). The requirement to find a shape that is *not* of a particular quality is important to minimize the opportunity to chunk these auditorily presented features. The task consists of five trials at each level of complexity (Levels 1-8), the difficulty of which is aligned with the number of stimulus features that must be concurrently activated in mind. For instance, a Level 1 trial may ask the participant to "Find a shape that is not red" (a single feature—red—to hold in mind), whereas Level 3 trial may ask the participant to "Find a shape that is not small, not blue and not a circle" (three features to hold in mind—small, blue, circle). Directions referring to multiple stimuli must be carried out in the specified order. Each trial proceeds as follows: (a) an auditory instruction played against a white screen, (a) a 3-s delay against a white screen, and then (c) a 4 × 5 array of different colored and sized shapes with cartoon faces, presented until a response is made by tapping the shape(s) that the participant believes

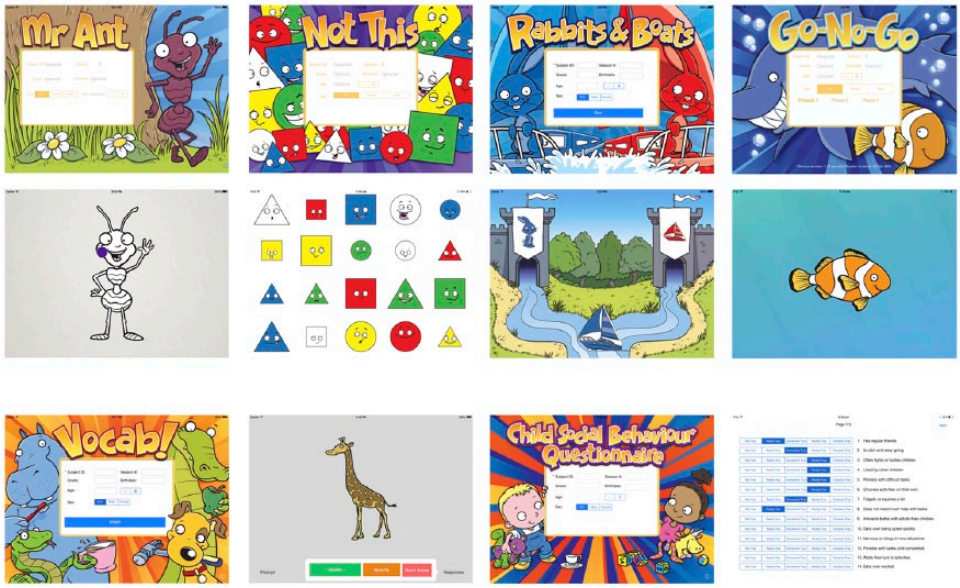

Figure 1. Screen captures of Early Years Toolbox Apps (title screen and test item screen captures).
Note. Top row, from left to right: Mr. Ant (visual-spatial working memory), Not This (phonological working memory), Card Sorting (shifting), Go/No-Go (inhibition). Bottom row, from left to right: Expressive Vocabulary, Child Self-Regulation and Behavior Questionnaire (self-regulation and social development).

correspond to the auditory instruction. The task continues until the earlier of completion (at Level 8, eight characteristics to remember) or failure to accurately complete at least three of the five trials within a level. Like the Mr. Ant task, performance was indexed by a point score calculated as follows: beginning from Level 1, one point for each consecutive level in which at least three of the five trials were performed accurately, plus 1/5 of a point for all correct trials thereafter.

EYT "Go/No-Go" task (inhibition). This task, following previously established protocols (Howard & Okely, 2015; Wiebe, Sheffield, & Espy, 2012), requires participants to tap the screen on "go" trials ("catch the fish") and not tap the screen on "no-go" trials ("avoid catching sharks"; see Figure 1). As the majority of stimuli are go trials (80% fish), this generates a prepotent tendency to respond, requiring participants to inhibit this response on no-go trials (20% sharks). Prior to commencing, participants are given instructions and practice as follows: go instructions, followed by five practice "go" trials; no-go instructions, followed by five practice "no-go" trials; combined go/no-go instructions, followed by a mixed block of 10 practice trials (80% go trials); and a recap of instructions. Feedback in the form of auditory tones is provided on all practice trials. The task proceeds with 75 stimuli divided evenly in three test blocks (each separated by a short break and a reiteration of instructions). Stimuli are presented in pseudo-random order, such that a block never begins with a no-go stimulus and no more than two successive trials are no-go stimuli. Each trial involves presentation of an animated stimulus (i.e., fish or shark) for 1,500 ms, separated by a 1,000 ms interstimulus interval. Inhibition was indexed by an impulse control score that is the product of proportional "go" (to account for the strength of the prepotent response generated) and "no-go" accuracy (to index a participant's ability to overcome this prepotent response).

EYT "Card Sorting" task (shifting). This task, based on the protocols of Zelazo (2006), requires children to sort cards (i.e., red rabbits, blue boats) by a sorting dimension (i.e., color or shape) into one of two locations (identified by a blue rabbit or a red boat), and then switch to the

alternate sorting rule (see Figure 1). After a demonstration trial and two practice trials, children begin sorting by one dimension for six trials. In the subsequent postswitch phase, children are required to sort cards by the other sorting dimension, as prompted by auditory instructions preceding postswitch test trials. In all conditions, each trial begins by reiterating the relevant sorting rule and then presenting a stimulus for sorting. If the participant correctly sorts at least five of the six pre and postswitch stimuli, they proceed to a border phase of the task. In this phase, children are required to sort by color if the card has a black border or sort by shape if the card has no black border. After a demonstration trial and two practice trials, this sorting rule is reiterated prior to presenting the six stimuli for sorting (consisting of three bordered stimuli and three nonbordered stimuli). For all conditions, cards are ordered such that a particular stimulus is never presented more than twice in a row. Scores represent the number of correct sorts after the preswitch phase.

EYT "Expressive Vocabulary" task (language development). This 45-item measure of a child's expressive vocabulary development requires children to verbally produce the correct label for each depicted stimulus (Figure 1). Participants respond verbally and a data collector records this response within the app. In cases of an incorrect label initially being produced, the data collector prompts participants by asking "what else might this be called" until either a correct production or some indication that the child is unable to produce the required word. Vocabulary items were selected from pilot testing (n = 1,319) of a preliminary set of 68 words, from which items were removed if there was evidence of multiple common labels, insufficient factor loadings in exploratory factor analyses, misfit in Rasch modeling and/or a poor discrimination index. A six-item stop rule was implemented to minimize administration time. This "stop score" correlated highly (r = .99) with the full-scale score in pilot data. An overall accuracy score indexed expressive vocabulary performance.

EYT "Child Self-Regulation and Behavior Questionnaire" (CSBQ; self-regulation and social development). Initially, 49 candidate items were administered to 114 children 3 to 6 years of age. Subsequent data reduction through factor analysis resulted in the CSBQ as a 33-item educator-report (or parent-report) questionnaire that yields subscales of Cognitive Self-Regulation, Behavioral Self-Regulation, and Emotional Self-Regulation, as well as Sociability, Prosocial Behavior, Externalizing Problems, and Internalizing Problems. Each item asks the respondent to evaluate the general frequency of target behaviors on a scale from 1 (*not true*) to 5 (*certainly true*). All subscales contain at least five items. Pilot testing of this scale (n = 114) indicated that all items loaded well on constituent factors in exploratory factor analysis and all subscales were reliable (all Cronbach's αs >.80). Some items contributed to more than one subscale reflecting the correlated nature of the subscales. Average subscale ratings were used for subsequent analyses and preliminary norm derivations.

Convergent Validity Measures. To evaluate convergent validity of the toolbox tasks, several existing measures were also administered to a subsample of participants. Measures of WM (List Sorting), inhibition (Flanker), and shifting (Dimensional Change Card Sorting [DCCS]) were drawn from the NIH Toolbox's Cognition Battery (for full description, see Tulsky et al., 2013; Zelazo et al., 2013). NIH List Sorting required participants to view and then recall, in size-sequential order, an increasing series of items. The NIH Flanker task required children to indicate the direction of a central stimulus flanked by congruent or incongruent flankers. The NIH DCCS task required children to sort stimuli (i.e., boats, rabbits) first by one dimension (i.e., color or shape), then by the other dimension and, if performance was above threshold, then flexibly switch between these sorting rules. In all cases, raw scores—generated by the NIH's central database extraction process—indexed performance. Participants also completed the BAS-2 Expressive Vocabulary subtest (for details, administration protocols, and scoring, see Elliott et al., 1996). Educators also completed the SDQ (Goodman, Meltzer, & Bailey, 1998) for each child as a comparison for the CSBQ.

Procedure

For EYT tasks, to minimize fatigue and maximize attention, each child participated in two separate assessment sessions. This consisted of one WM assessment per session paired with one or two additional toolbox tasks. Each session lasted no more than 15 min and took place in a quiet space in the child's ECEC setting. Within this data collection period, educators familiar with the child completed the CSBQ (and, for the convergent validity subsample, the SDQ). As these settings rarely had a broadband wireless Internet connection available, EYT tasks were administered offline via an iPad and data were remotely sent to a secure database at the end of each day. For NIH Toolbox tasks, a mobile broadband modem (to ensure continuous Internet connection), 17″ touchscreen laptop and a same-sized monitor (to enable the required tester display) were required. Convergent validity task administration paralleled EYT assessment sessions (i.e., split into two testing sessions in the same setting as EYT measures were completed) 3 weeks after EYT task administration. Child assessors trained in the use of these measures administered all tasks.

Results

Data Screening

To ensure only valid responses were included in analyses, data and tester logs were first screened to identify potential grounds for exclusion of individual data points. For the EYT Go/No-Go task, data were removed in cases of extremely rapid responding (individual trials were removed if response time < 300 ms, because responding was unlikely to have been in response to the stimulus), nonresponsiveness (blocks were removed if go accuracy fell below 20% and no-go accuracy exceeded 80%), and indiscriminant responding (blocks were removed if go accuracy exceeded 80% and no-go accuracy fell below 20%). This screening resulted in the removal of 2.9% of participants' Go/No-Go data. Data were also unavailable in cases of early withdrawal, which accounted for 2.0% data loss. For the convergent validity subsample, 12 data points (4.7%) were missing for the NIH toolbox due to early withdrawal from the tasks (compared with three data points, or 0.9%, for EYT).

EYT Reliability

Internal consistency analyses were conducted for the Go/No-Go, Expressive Vocabulary, and CSBQ measures of the EYT as an indicator of the reliability of each measure. Go/No-Go had good reliability for both go (Cronbach's $\alpha = .95$) and no-go trials (Cronbach's $\alpha = .84$). Reliability of the Expressive Vocabulary task was excellent (Cronbach's $\alpha = .92$). Cronbach's alphas for the CSBQ subscales ranged from acceptable to very good as follows: Sociability = .74, Internalizing = .78, Emotional Self-Regulation = .83, Cognitive Self-Regulation = .87, Externalizing = .88, Prosocial = .89, and Behavioral Self-Regulation = .89. While test–retest reliability would be the more appropriate index of reliability for the other EYT tasks due to their comparatively few trials at each level of complexity, these data were not available for analyses and thus reliability estimates for these tasks could not be accurately computed.

EYT Convergent Validity

Convergent validity was assessed in a subsample of participants by correlations of EYT measures with existing measures of these same constructs (i.e., NIH Toolbox, BAS, SDQ) that have been used extensively in this age group. Results indicated that EYT measures correlated well with

Table 2. Inter-Task Correlations for EYT and Convergent Validity Measures (EFs and Language).

		1	2	3	4	5	6	7	8	9
1	EYT VWM	—	.28*	.35*	.12	.43*	*.46**	.38*	.26*	.20
2	EYT PWM		—	.34*	.23*	.57*	*.42**	.28*	.50*	.42*
3	EYT Inhibition			—	.12	.38*	.37*	*.40**	.31*	.21*
4	EYT Shifting				—	.30*	.27*	.05	*.45**	.13
5	EYT Vocabulary					—	.51*	.36*	.50*	*.60**
6	NIH WM						—	.34*	.44*	.54*
7	NIH Inhibition							—	.27*	.28*
8	NIH Shifting								—	.37*
9	BAS Vocabulary									—

Note. Inhibition was measured by a Go/No-Go (EYT) and Flanker task (NIH). Shifting was measured by a card sorting task in both EYT and NIH. EYT = Early Years Toolbox; EF = executive function; VWM = visual working memory (as measured by the Mr. Ant task); PWM = phonological working memory (as measured by the Not This task); NIH = National Institute of Health; Vocab = expressive vocabulary measures; BAS = British Ability Scales.
*$p < .05$.

their comparison measures, as follows: language, $r(84) = .60$, $p < .001$; WM, $r(79) = .46$, $p < .001$ (with visual-spatial) and $r(79) = .42$, $p < .001$ (with phonological); Inhibition, $r(80) = .40$, $p < .001$; and Shifting, $r(80) = .45$, $p < .001$ (see Table 2, for a full inter-task correlation matrix). Consistently significant and moderately strong correlations between EYT measures and the corresponding convergent validation measures indicated good convergent validity with established measures. However, that these correlations did not approach a perfect relationship indicated the different nature of the measures (e.g., differential capture of the latent construct, or error variance). Regarding this point, differences in participants' ability to perform EYT tasks (2.3% of data points at floor) versus NIH Toolbox tasks (10.6% of data points at floor) are notable. Differential data loss due to participants' early withdrawal (0.9% for EYT tasks vs. 4.7% for NIH tasks) also supports the advantages of the EYT.

Correlations between analogous CSBQ and SDQ subscales were as follows: Externalizing, $r(84) = .91$, $p < .001$; Internalizing, $r(84) = .78$, $p < .001$; and Prosocial, $r(84) = .85$, $p < .001$. For all other CSBQ subscales, correlations were compared with the nearest-comparison SDQ subscale given there were no direct analogues. Correlations between these subscales were as follows: Sociability, $r(84) = .48$, $p < .001$ (with Prosocial); Behavioral Self-Regulation, $r(84) = -.81$, $p < .001$; Emotional Self-Regulation, $r(84) = -.66$, $p < .001$; and Cognitive Self-Regulation, $r(84) = -.70$, $p < .001$ (with Hyperactivity). These correlations were consistently strong (see Table 3, for a full inter-correlation matrix), suggesting that the subscales were tapping into similar constructs. Although CSBQ Self-Regulation subscales were correlated with a single SDQ subscale (Prosocial), factor analysis of the CSBQ yielding seven separable, yet related subscales provided evidence for diversity in the constructs being captured by each subscale. In fact, the CSBQ Prosocial subscale correlated similarly well, but not perfectly, with its behavioral Self-Regulation ($r = .69$), Cognitive Self-Regulation ($r = .78$), and Emotional Self-Regulation subscales ($r = .70$). This suggests that the CSBQ Self-Regulation subscales were not mere variants of an underlying prosocial behavior scale.

Age Effects and Preliminary Language and EF Norms

Descriptive statistics and preliminary norms are presented in Tables 4 and 5. With the expressive vocabulary task, an analysis of variance indicated a significant effect of Age, $F(6, 1260) = 82.39$, $p < .001$, $\eta^2 = .28$. Post hoc Ryan-Einot-Gabriel-Welsh Q test (REGWQ) analyses indicated that

Table 3. Inter-Task Correlations for EYT and Convergent Validity Measures (CSBQ and SDQ).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	CSBQ Sociab	—	.11	−.28*	.63*	.18	.44*	.27*	−.23*	−.66*	.48*	−.36*	−.08	−.27*	−.86*
2	CSBQ External		—	.02	−.52*	−.83*	−.36*	−.71*	.91*	.09	−.45*	.66*	.97*	−.05	.24*
3	CSBQ Internal			—	−.05	−.02	−.17	−.25*	−.01	.78*	.01	.03	−.04	.92*	.29*
4	CSBQ Prosoc				—	.69*	.78*	.70*	−.67*	−.32*	.85*	−.71*	−.53*	−.01	−.57*
5	CSBQ BehSR					—	.63*	.80*	−.87*	−.07	.56*	−.78*	−.81*	.08	−.22*
6	CSBQ CogSR						—	.71*	−.53*	−.26*	.66*	−.65*	−.34*	−.12	−.33*
7	CSBQ EmoSR							—	−.69*	−.28*	.61*	−.60*	−.66*	−.19	−.27*
8	SDQ External								—	.10	−.62*	.90*	.93*	−.10	.31*
9	SDQ Internal									—	−.20	.17	.03	.85*	.77*
10	SDQ Prosoc										—	−.69*	−.47*	.08	−.47*
11	SDQ Hyperact											—	.67*	−.06	.39*
12	SDQ Conduct												—	−.12	.20
13	SDQ Emo													—	.33*
14	SDQ Peer														—

Note. EYT = Early Years Toolbox; CSBQ = Child Self-Regulation and Behavior Questionnaire; SDQ = Strengths and Difficulties Questionnaire; Sociab = Sociability subscale; External = Externalizing subscale; Internal = Internalizing subscale; Prosoc = Prosocial subscale; BehSR = Behavioral Self-Regulation subscale; CogSR = Cognitive Self-Regulation subscale; EmoSR = Emotional Self-Regulation subscale; Hyperact = Hyperactivity subscale; Conduct = Conduct Problems subscale; Peer = Peer Problems subscale.
*$p < .05$.

Table 4. Preliminary Expressive Vocabulary Norms, by Age Group—Quintiles or Corrected Score by SEIFA.

	Age groups						
	2:6-2:11	3:0-3:5	3:6-3:11	4:0-4:5	4:6-4:11	5:0-5:5	5:6-5:11
25th percentile	6.00	12.00	16.00	19.00	22.00	23.00	26.00
50th percentile	11.00	17.00	21.00	24.00	28.00	29.00	29.00
75th percentile	18.00	22.00	25.00	29.00	32.00	32.75	33.00
M	11.19	16.68	20.35	23.13	26.90	27.55	28.66
SD	6.78	7.20	7.28	8.04	7.25	7.36	6.21

Note. Age groups are presented as years:months.

Table 5. Preliminary Executive Function Norms by Age Group.

	Age groups (years:months)					
	3:0-3:5	3:6-3:11	4:0-4:5	4:6-4:11	5:0-5:5	5:6-5:11
Mr. Ant (visual-spatial working memory)						
25th percentile	0.00	0.33	1.00	1.33	1.33	2.00
50th percentile	1.00	1.00	2.00	2.00	2.00	2.33
75th percentile	1.33	2.00	2.33	2.33	2.33	2.75
M	0.85	1.14	1.57	1.74	1.98	2.31
SD	0.76	0.86	0.89	0.87	0.82	0.80
Not This (phonological working memory)						
25th percentile	0.40	1.00	1.00	1.40	1.40	1.40
50th percentile	1.20	1.40	1.40	2.00	2.00	2.20
75th percentile	1.60	2.00	2.20	2.40	2.40	2.80
M	1.14	1.38	1.52	1.88	1.99	2.11
SD	0.72	0.73	0.81	0.77	0.76	0.79
Go/No-Go (Inhibition)						
25th percentile	0.22	0.26	0.39	0.48	0.50	0.63
50th percentile	0.32	0.43	0.54	0.67	0.68	0.73
75th percentile	0.46	0.61	0.71	0.77	0.81	0.86
M	0.36	0.44	0.54	0.63	0.65	0.71
SD	0.19	0.22	0.21	0.20	0.21	0.20
CS (shifting)						
25th percentile	0.00	0.33	0.00	0.00	0.00	1.00
50th percentile	1.00	1.00	3.00	4.00	7.00	9.00
75th percentile	3.00	4.00	8.00	9.00	9.00	10.00
M	2.07	2.59	3.87	4.89	5.54	6.33
SD	2.98	3.48	3.97	4.34	4.35	4.25

Note. Age groups are presented as years:months. CS = Card Sorting task.

performance improved with increasing age from 2.5 years to 4.5 years, with nonsignificant improvements in performance at 5 and 5.5 years of age. Furthermore, preliminary norms for the 25th, 50th, and 75th percentile on the expressive vocabulary task display clear improvement in performance with half-yearly increases in age across these percentiles.

For the WM tasks, analyses of variance again indicated a significant effect of Age: Mr. Ant, $F(6, 1628) = 58.55$, $p < .001$, $\eta^2 = .24$; and Not This, $F(6, 1087) = 26.64$, $p < .001$, $\eta^2 = .13$. Post

hoc REGWQ analyses of the Mr. Ant task indicated that performance increased across the age groups, although with no significant difference between the 4- and 4.5-year-old age groups, or between the 5- and 5.5-year-old age groups. Post hoc REGWQ analyses were similar for the Not This task, such that (a) 4.5- to 5.5-year-old groups, which did not significantly differ, outperformed all other age groups; and (b) the 3.5- to 4-year-old groups, which did not significantly differ, outperformed the 3-year-old group. Of particular note is that for both tasks, despite differing in their phonological or visual-spatial nature, means were largely in line with a priori theoretical predictions regarding the development of mental-attentional capacity (positioned as the causal mechanism underlying developmental growth of WM; Pascual-Leone & Johnson, 2005, 2011). That is, the 3-year-old group had a mean score of 1.00 ($SD = 0.83$) on Mr. Ant and 1.26 ($SD = 0.74$) on Not This, which is highly consistent with the predicted (and empirically demonstrated) mental-attentional capacity of one unit of information (i.e., scheme) that can be concurrently activated in mind at this age. Similarly, in line with predictions that this increases by one unit every other year until around 15 years of age, the 5-year-old group had a mean score of 2.06 ($SD = 0.83$) on Mr. Ant and 1.98 ($SD = 0.74$) on Not This. This provides additional theoretical grounds to suggest the appropriateness of these tasks as measures of WM (or, perhaps more accurately, mental-attentional) capacity.

For the other EF tasks, there were similar effects of age: Go/No-Go, $F(6, 1592) = 96.83$, $p < .001$, $\eta^2 = .27$, and Card Sorting, $F(6, 1607) = 37.81$, $p < .001$, $\eta^2 = .12$. Post hoc REGWQ analyses indicated that for Go/No-Go, performance increased across age groups until 4.5 years, with nonsignificant improvements in performance in the 5- and 5.5-year-age groups. For Card Sorting, performance was highest for the 4.5- to 5.5-year-old age groups, which did not significantly differ. The 4-year-old group also displayed significantly better performance than the 3- or 3.5-year-old groups, which did not significantly differ. Once again, preliminary norms for the 25th, 50th, and 75th percentiles across the EF tasks displayed clear improvement in performance with half-yearly increases in age across these percentiles. Mean performance did not approach ceiling on any EYT measure for any of the age groups. This suggests potential for these tasks to be used even further into the early primary years.

Preliminary CSBQ Norms

Given that there were not expected to be linear increases in educator-rated self-regulation abilities with increasing age, and consistent with norm presentation for the SDQ, 80th, 90th, and 95th percentile scores were computed for negatively framed subscales and 5th, 10th, and 20th percentile scores for positively framed subscales (see Table 6). This was used to derive slightly high/low, high/low and very high/low categorization norms, consistent with the SDQ method of norm presentation.

Discussion

Given the limitations in existing early cognitive development measures, we sought to develop, validate, and establish preliminary norms for an EYT of early language, EF, self-regulation, and social behavior measures that improve upon these limitations. Specifically, to address conceptual limitations, we sought to ensure age-appropriate stimuli, timings, instructions, and response methods (e.g., intuitive, engaging, and appropriately challenging), while seeking to minimize the influence of prior learning and experience that might conflate performance such as literacy and numeracy demands. In response to pragmatic limitations, we sought to develop a free, mobile, and easy to administer toolbox that does not require constant Internet access or equipment beyond the tablet device (currently iPad, but android versions are under development). Results from our large-scale administration of this toolbox indicated very good reliability, convergent validity

Table 6. Preliminary CSBQ Norms.

Negative	*M* (*SD*)	Close to average	Slightly high, 80th percentile	High, 90th percentile	Very high, 95th percentile
Externalizing	1.91 (0.98)	1.00-2.59	2.60-3.39	3.40-3.79	3.80-5.00
Internalizing	1.85 (0.73)	1.00-2.39	2.40-2.99	3.00-3.39	3.40-5.00
Positive	***M* (*SD*)**	**Close to average**	**Slightly low, 20th percentile**	**Low, 10th percentile**	**Very low, 5th percentile**
Prosocial	3.54 (0.84)	2.86-5.00	2.29-2.85	2.00-2.28	1.00-1.99
Cognitive SR	3.41 (0.84)	2.83-5.00	2.27-2.82	1.83-2.26	1.00-1.82
Behavioral SR	3.54 (0.90)	2.75-5.00	2.33-2.74	2.00-2.32	1.00-1.99
Emotional SR	3.48 (0.67)	2.86-5.00	2.57-2.85	2.29-2.56	1.00-2.28
Sociability	3.62 (0.80)	3.00-5.00	2.40-2.99	2.20-2.39	1.00-2.19

Note. CSBQ = Child Self-Regulation and Behavior Questionnaire; SR = self-regulation.

with high-profile existing measures, and developmental sensitivity across half-year age groups. Furthermore, patterns of results were suggestive of potentially better capture of children's performance relative to comparison measures, insofar as there was comparatively less withdrawal and performance at floor (without concomitant ceiling effects). The EYT thus appeared to achieve its stated aims of developmental appropriateness, sensitivity, brevity, engagement, validity and reliability, as well as being technologically dynamic and with the potential for international application (although further research, such as predictive validity, test–retest reliability and cross-cultural comparison studies would provide further evidence of these achievements).

Overall, validation analyses indicated good to excellent reliability for all of the toolbox measures evaluated and good convergent validity with existing measures. However, that the correlations between EYT and existing measures were typically not near perfect supports the view that these measures were tapping into the same underlying construct(s), but that there was differentiation between the measures, which is to be expected if the new measures reflect improvements. While some of this variability is likely related to different task demands (e.g., inhibiting a prepotent screen touch in the EYT Go/No-Go task and suppressing interference from flanking distractors in the NIH Flanker task), it is notable that even for the most highly comparable tasks the correlation remained modest. Between the EYT and NIH card sorting, as an example, the correlation was $r = .45$ despite highly similar stimuli and task demands. While further research is required to identify the sources of differentiation between these measures, patterns of performance suggest that at least one factor might be differences in children's abilities to demonstrate their emerging competencies on these tasks. That is, more than 10% of the NIH data were at floor compared with only 2.3% for EYT. Early withdrawal while the children performed the tasks resulted in a further 4.7% of the data being lost for NIH tasks, yet only 0.9% for EYT tasks. As such, it is possible that at least one source of differentiation between the tasks are differing levels of engagement and ability to understand and carry out task requirements. As such, an important area for further study is an examination of the task characteristics that optimally permit children to demonstrate their emerging cognitive competencies. Furthermore, given that these results could also be explained by limited test–retest reliability, which would constrain inter-task correlations within and between toolboxes' tasks, it is also important to investigate the extent to which these measures provide sufficient test–retest reliability.

While the inconsistency in developmental sensitivity of early cognitive measures is well documented (Carlson, 2005; Zelazo et al., 2013), current results with the EYT displayed clear cross-sectional trends across all quartiles at half-year age bands. Although it is very common for EF measures

to show age-related change, especially in the preschool years in which these abilities are rapidly developing (Best & Miller, 2010), few studies have sought to establish developmental norms for these measures (the NIH Toolbox being a notable exception). This serves to limit capacity for comparisons across studies, demographic and developmental subgroups, and geographic areas. In contrast, we have provided preliminary norms for each of the EYT measures to support use and discussion by fellow researchers. We term these norms *preliminary* due to their creation from an exclusively Australian sample, which was slightly biased toward disadvantaged participants, and these norms will be updated as additional data become available. Nevertheless, the norms provide a base for further development, research, use, and refinement both nationally and internationally.

In addition to the exclusively Australian sample, and low-SES bias within that sample, two other limitations are apparent in the current constitution of the toolbox. First, as with any age-constrained cognitive measure, the necessity of using separate measures in early and later childhood makes the mapping of life span developmental trajectories difficult. For this reason, efforts such as the NIH Toolbox (Weintraub et al., 2013) have sought to create measures that span a much broader age range. However, at least anecdotally, there appears to be growing recognition that measures that are equally appropriate for young children, older childhood, and adolescents may not be ideal for accurately capturing the abilities of these highly distinct age groups. This is at least partly a consequence of young children's comparatively limited knowledge base, experience, cognitive control (e.g., ability to sustain their attention), as well as their comparatively limited ability to understand instructions and communicate a response (Howard & Okely, 2015). This has led to development of distinct measures for early (e.g., Weschler Preschool and Primary Scale of Intelligence, for children aged 2-7 years) and later cognitive development (e.g., Weschler Intelligence Scale for Children, for children aged 6-16 years). In fact, even the NIH Toolbox has an early years cognition battery that is at least partially distinct from their full cognition battery. There is also some evidence of the problems associated with the downward extension of measures that were developed for older children or adults (Howard & Okely, 2015), such that in the present study, the children had a comparatively more-limited ability to perform NIH Toolbox tasks than the analogous EYT tasks.

A further limitation that applies most saliently to EF measurement is an inability with the EYT to conduct latent variable analyses with at least three indicators for each latent factor (although this is also the case for other task batteries). That is, given the "task impurity" issue common in EF measures, a latent variable approach to data analysis (e.g., confirmatory factor analysis, structural equation modeling) is commonly undertaken to minimize the likelihood of conclusions that are influenced by variance unassociated with the constructs of interest (Friedman & Miyake, 2004; Howard et al., 2014; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). With only a single measure of inhibition and shifting, and two measures of WM, a latent variable approach is only possible by using complementary measures from outside the toolbox (further suggesting the utility of a toolbox that complements existing measures and toolboxes). While further development of measures is planned to remedy this, at present, this means that the typical practice of generating latent variables from highly discrepant measures (and the consequent measure-specific influence on results) will continue at least in the short term. It is thus vitally important that research in the area of age- and sample-appropriate task development, selection, and combination continue. Given inter-task correlations within the toolbox ranging from .24 to .48 in the current study (compared with inter-task correlations more often between .09 and .35 in other EF studies in this age group; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012; St. Clair-Thompson & Gathercole, 2006), there is evidence that the EYT represents an early step in this regard.

Conclusion

The current study thus documents both methodological and practical advances from which to inform future task development, selection, and comparisons. Specifically, this study serves to

establish the validity, reliability, and developmental sensitivity of the EYT for use with preschool and early primary school students. As a testament to this usability, following word-of-mouth discussion among stakeholders, the EYT has been downloaded by more than 7,000 users and is in use by researchers in Australia, Canada, the United Kingdom, and South Africa and is currently being translated to additional languages (e.g., Afrikaans) and platforms (i.e., Android). On the basis of this broad use and adaptation, future studies will seek to develop norms in other countries and cultures. We are also aware of its use in the education system in Australia to supplement schools' academic data collection processes, suggesting its potential relevance to not only research applications but also brief screening and educational data collection and decision making. The current study further serves to highlight at least some of the design characteristics that have successfully yielded cognitive measurements in young children, such as the need to minimize the impact of prior learning and create tasks that are sufficiently engaging to young children. Comparison with additional task batteries common in other domains (e.g., the Developmental Neuropsychological Assessment toolbox [NEPSY]; Korkman et al., 2007) could yield further insight into this issue. In a more practical vein, the EYT provides stakeholders and researchers with a freely and easily accessible means of brief assessment of young children's development. Creation of preliminary norms further supports use across this wide range of interests and expertise. Much remains unclear regarding early development, including the changes in the nature and structure of cognition (e.g., EFs) from early childhood to adulthood and whether attempts to support and enhance early cognitive development have consequent effects on outcomes often associated with these early abilities (e.g., school readiness, literacy and numeracy advantage, social and emotional competence; Blair & Razza, 2007; Bull et al., 2008; Welsh et al., 2010). However, a first step in clarifying these issues requires accurate, consistent, and early measurement of these emerging abilities. The EYT represents an important advance in this regard.

Acknowledgment

We are grateful to the participating schools and preschools and the children who participated in the project.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by Internal Funding from the University of Wollongong and Early Start Research Institute.

References

Adi-Japha, E., & Klein, P. S. (2009). Relations between parenting quality and cognitive performance of children experiencing various amounts of childcare. *Child Development*, *80*, 803-906.

Australian Bureau of Statistics. (2012). *Census Fact Sheet* [Fact Sheet]. Retrieved from http://www.abs.gov.au/census

Australian Institute of Health and Welfare. (2012). *A picture of Australia's children 2012*. Retrieved from http://www.aihw.gov.au

Berk, L. E., Mann, T. D., & Ogan, A. T. (2006). Make-believe play: Wellspring for development of self–regulation. In D. G. Singer, R. M. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = Learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 74-100). New York, NY: Oxford University Press.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*, 1641-1660.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*, 180-200.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111-127.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, *20*, 899-911.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging literacy ability in kindergarten. *Child Development*, *78*, 647-663.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, *28*, 561-571.

Bodrova, E., & Leong, D. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

Boeckaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, *54*, 199-231.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205-228.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 595-616.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York, NY: Academic Press.

Chi, M. T. H. (1978). Knowledge students and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ: Lawrence Erlbaum.

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews: Neuroscience*, *13*, 636-650.

Daglar, M., Melhuish, E., & Barnes, J. (2011). Parenting and pre-school child behaviour amongst Turkish immigrant, migrant and non-migrant families. *European Journal of Developmental Psychology*, *8*, 261-279.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959-964.

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore, MD: Brookes.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428-1446.

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales second edition (BAS II): Administration and scoring manual*. London, England: Nelson.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs for the Society for Research in Child Development, 59*(5), 1-185.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101-135.

Galinsky, E. (2010). *Mind in the making*. New York, NY: Harper Collins.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581-586.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, *7*, 125-130.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*, 65-73.

Hart, B., & Risley, R. T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368-1378.

Hoff-Ginsberg, E. (1991). Mother–child conversation in different social classes and communicative settings. *Child Development*, *62*, 782-796.

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*, 174-180.

Hogan, A. E., Scott, K. G., & Bauer, C. R. (1992). The Adaptive Social Behavior Inventory (ASBI): A new assessment of social competence in high-risk three-year-olds. *Journal of Psychoeducational Assessment*, *10*, 230-239.

Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, *31*, 1-21.

Howard, S. J., & Okely, A. D. (2015). Catching fish and avoiding sharks: Investigating factors that influence developmentally appropriate measurement of preschoolers' inhibitory control. *Journal of Psychoeducational Assessment*, *33*(6), 585-596.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233-253.

Hughes, C. (2011). *Social understanding and social lives*. Hove, UK: Psychology Press.

Huttenlocher, J., Vasilyeva, M., Waterfall, H., Vevea, J., & Hedges, L. (2007). The varieties of speech to young children. *Developmental Psychology*, *43*, 1062-1083.

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development*, *77*, 1822-1841.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*, 220-232.

Korkman, M. (1999). Applying Luria's diagnostic principles in the neuropsychological assessment of children. *Neuropsychological Review*, *9*, 89-105.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Clinical and interpretive manual*. San Antonio, TX: The Psychological Corporation.

Lamb, M. E., & Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigek (Eds.), *Handbook of child psychology: Volume 4: Child psychology in practice* (5th ed.). Hoboken, NJ: Wiley.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59-80.

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, *139*, 1-34.

Lugo-Gil, J., & Tamis-LeMonda, C. S. (2008). Family resources and parenting quality: Links to children's cognitive development across the first three years. *Child Development*, *79*, 1065-1085.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471-490.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C., Jewkes, A., & Morrison, F. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*, 947-959.

Melhuish, E., & Barnes, J. (2012). Preschool programs for the general population. In R. E. Tremblay, M. Boivin, & R. V. Peters (Eds.), *Encyclopedia on early childhood development*. Montreal, Quebec, Canada: Centre of Excellence for Early Childhood Development and Strategic Knowledge Cluster on Early Child Development. Retrieved from http://www.child-encyclopedia.com/Pages/PDF/Melhuish-BarnesANGxp1.pdf

Melhuish, E., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., & Phan, M. (2007). The home learning environment, attainment and resilience in Effective Pre-school and Primary Education 3-11 (EPPE 3-11). In Effective Pre-school and Primary Education 3-11 Team (Eds.), *Promoting equality in the early years: Report to the equalities review* (pp. 50-71). London, England: Cabinet Office.

Miech, R., Essex, M. J., & Goldsmith, H. H. (2001). Socioeconomic status and the adjustment to school: The role of self-regulation during early childhood. *Sociology of Education*, *74*(2), 102-120.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*, 395-423.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and the contributions to complete "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

Moffitt, T. E., Arseneault, L., Belsky, J., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*, 2693-2698.

Morra, S. (1994). Issues in working memory measurement: Testing for M capacity. *International Journal of Behavioral Development*, *17*, 143-159.

National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB). (2013). *Monographs of the Society for Research in Child Development*, *78*(4), 1-172.

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socio-emotional adjustment during the transition to kindergarten? *Child Development*, *74*, 976-1005.

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*, 99-114.

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*, 763-782.

Pascual-Leone, J., & Johnson, J. (2005). A dialectical constructivist view of developmental intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 177-201). Thousand Oaks, CA: Sage.

Pascual-Leone, J., & Johnson, J. (2011). A developmental theory of mental attention: Its application to measurement and task analysis. In P. Barrouillet & V. Gaillard (Eds.), *Cognitive development and working memory: A dialogue between neo-Piagetian theories and cognitive approaches* (pp. 13-46). New York, NY: Psychology Press.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioural self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*, 605-619.

Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool children's language and emergent literacy. *Journal of Early Childhood Literacy*, *10*, 97-117.

Rhoades, B. L., Greenberg, M. T., Lanza, S. T., & Blair, C. (2011). Demographic and familial predictors of early executive function development: Contribution of a person-centered perspective. *Journal of Experimental Child Psychology*, *108*, 638-662.

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Müeller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology*, *27*, 300-309.

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*, 508-525.

Sammons, P., Elliot, K., Sylva, K., Melhuish, E. C., Siraj-Blatchford, I., & Taggart, B. (2004). The impact of pre-school on young children's cognitive attainments at entry to reception. *British Educational Research Journal*, *30*, 691-712.

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods*. Washington, DC: National Academy Press.

Sirois, S., Spratling, M., Thomas, M. S. C., Westermann, G., Mareschal, D., & Johnson, M. H. (2008). Précis of neuroconstructivism: How the brain constructs cognition. *Behavioral and Brain Sciences*, *31*, 321-331.

Slaughter, V., Imuta, K., Peterson, C., & Henry, J. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development*, *86*, 1159-1174.

Snow, C. E. (2011). *The potential of discussion to improve reading comprehension*. New York, NY: Lecture IDA.

Snow, C. E., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59*, 745-759.

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). V. NIH Toolbox Cognition Battery (CB): Measuring working memory. *Monographs for the Society for Research in Child Development, 78*(4), 70-87.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., &NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. *Child Development*, *81*, 737-756.

Van der Ven, S. H. G., Kroesbergen, E. K., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology*, *82*, 100-119.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., . . . Gershon, R. C. (2013). I. NIH Toolbox Cognition Battery (CB): Introduction and pediatric data. *Monographs for the Society for Research in Child Development, 78*(4), 1-15.

Wellman, R. L., Lewis, B. A., Freebairn, L. A., Avrich, A. A., Hansen, A. J., & Stein, C. M. (2011). Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech, and Hearing Services in Schools*, *42*, 561-579.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*, 43-53.

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development*, *83*, 1245-1261.

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

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs for the Society for Research in Child Development, 78*(4), 16-33.