



## OPEN ACCESS

## EDITED BY

David Pérez-Jorge,  
University of La Laguna, Spain

## REVIEWED BY

Joseline Santos,  
Bulacan State University, Philippines  
Kathy Sylva,  
University of Oxford, United Kingdom

## \*CORRESPONDENCE

Claire E. Cameron  
✉ cecamero@buffalo.edu

RECEIVED 07 April 2023

ACCEPTED 18 December 2023

PUBLISHED 11 January 2024

## CITATION

Cameron CE, McClelland MM, Kwan T,  
Starke K and Lewis-Jones T (2024) HTKS-Kids:  
A tablet-based self-regulation measure to  
equitably assess young children's school  
readiness. *Front. Psychol.* 14:1202239.  
doi: 10.3389/fpsyg.2023.1202239

## COPYRIGHT

© 2024 Cameron, McClelland, Kwan, Starke  
and Lewis-Jones. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC  
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# HTKS-Kids: A tablet-based self-regulation measure to equitably assess young children's school readiness

Claire E. Cameron<sup>1\*</sup>, Megan M. McClelland<sup>2</sup>, Tammy Kwan<sup>3</sup>,  
Krystal Starke<sup>1</sup> and Tanya Lewis-Jones<sup>1</sup>

<sup>1</sup>Department of Learning and Instruction, University at Buffalo, Buffalo, NY, United States, <sup>2</sup>School of Human Development and Family Sciences, Oregon State University, Corvallis, OR, United States, <sup>3</sup>Cognitive ToyBox, Inc., New York, NY, United States

**Background:** Technology advances make it increasingly possible to adapt direct behavioral assessments for classroom use. This study examined children's scores on HTKS-Kids, a new, largely child-led version of the established individual research assessment of self-regulation, Head-Toes-Knees-Shoulders-Revised task (HTKS-R). For the HTKS-Kids tablet-based assessment, which was facilitated by children's preschool teachers, we examined (1) preliminary reliability and validity; (2) variation in scores predicted by child age and background characteristics; and (3) indication that HTKS-Kids provides different information from teacher ratings of children.

**Method:** Participants included  $n = 79$  4-year-old children from two urban areas in upstate New York, USA. Average parent education was 12.5 years, ranging 3–20. A researcher administered the HTKS-R to individual children, and teachers (eight white, two Latino) were trained to use the HTKS-Kids tablet-based assessment and asked to play once with each study child. Teachers also rated each child on 10 Child Behavior Rating Scale (CBRS) items about classroom self-regulation.

**Results:** We found evidence that (1) the HTKS-Kids captures variation in children's self-regulation and correlates positively with established measures, (2) parent education was the best predictor of HTKS-Kids scores, and (3) teachers rated Black children significantly worse and white children better on the CBRS, with the magnitude of group differences similar to the contribution of parent education. In contrast, Black and white children showed no score differences on HTKS-Kids.

**Implications:** The HTKS-Kids is a promising new tablet-based assessment of self-regulation that could replace or supplement traditional teacher ratings, which are often subject to implicit bias.

## KEYWORDS

early childhood, equity, executive function, school readiness, self-regulation, technology-based assessment

## Introduction

Developmentally appropriate assessment is part of supporting young children's successful transition to school (Hirsh-Pasek et al., 2005). Assessment includes any tool or method that helps teachers or educational systems document children's knowledge and skills (Neuman and Devercelli, 2013; Smith et al., 2015). In early childhood, assessment

has many purposes, such as to inform learning activities, to identify or screen individual children for intervention in an area such as speech or motor skills, or to improve educational programs and curricula through evaluation (National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education, 2003; Gokiert et al., 2013, p. 1). Experts consider holistic assessment in multiple skill domains to be developmentally appropriate for young children and can support their learning by more precisely identifying their needs and strengths (Hirsh-Pasek et al., 2005; National Research Council, 2008).

Holistic child assessment includes academic as well as non-academic skills and is recognized as critical, given the intertwined nature of multiple developmental domains in the early childhood period (McClelland and Cameron, 2018). Experts recommend naturalistic and observation-based assessment approaches but these can pose significant burdens for teachers (Cameron et al., 2023). Ideal assessment practices are not always implemented however, and teachers, who may be short of time, may inadvertently introduce bias or error in their documentation of children's skills (Waterman et al., 2012). Structured direct assessment is one alternative to naturalistic observations where an adult presents an individual child with tasks or questions. Direct assessment is more standardized, which can increase the reliability and validity of assessment (National Research Council, 2008); and allows for the development of reports and recommendations to compare data across children, classrooms and programs (Waterman et al., 2012). In particular, technology-assisted direct assessment has the potential to increase the frequency of assessment, feedback to students and teachers, objectivity and consistency, and administrative efficiency (Bull and McKenna, 2003). In this study, we report preliminary psychometric properties from a tablet-based adaptation of an internationally-used direct research assessment of self-regulation, the Head-Toes-Knees-Shoulders-Revised or HTKS-R (Gonzales et al., 2021; McClelland et al., 2021). The new measure, HTKS-Kids, is a tablet assessment with highly similar regulatory demands to the HTKS-R research instrument. HTKS-Kids is teacher-facilitated but largely child-led, and is designed to be integrated into the regular preschool day.

## Assessment of self-regulation in early childhood

The preschool years are a critical time for children's development, particularly as they learn to self-regulate, or to effectively manage their nervous system, emotions, cognition, and behaviors across contexts (Bailey and Jones, 2019; Blair and Ku, 2022). Self-regulation has its roots in infant attention and reactivity (Rothbart et al., 2006) and changes throughout development (McClelland et al., 2015b). Self-regulatory skills among typically developing children improve dramatically in early childhood as the prefrontal cortex matures to support executive function (McClelland and Cameron, 2012; Bailey and Jones, 2019). Executive function (EF) is an umbrella term that describes the specific cognitive processes associated with developing, planning, and executing goals (Miyake et al., 2000).

As students progress through school, their self-regulatory capacities grow with advances in EF, enabling them to plan and carry out increasingly complex task sequences (Blair and Raver, 2015). In early childhood, where this study focuses, strong EF allows children to intake, process, filter and organize information; discard extraneous information; and make adaptive choices (Diamond, 2016). By the end of early childhood, EF processes that can be measured distinctly include working memory, task-switching, and inhibitory control; which work together to contribute to overall self-regulation of behavior and responses in a given environment such as a classroom (Blair and Ku, 2022). Both constructs predict future academic and personal success, with EF more closely associated with cognitive processes that are consciously applied, and self-regulation encompassing adaptations to environments that may or may not allow for the practice and exercise of EF (Zelazo, 2020; Blair and Ku, 2022). Children's performance on measures of both self-regulation and EF are strong predictors of their school readiness (Blair, 2002; McClelland et al., 2014), as well as math and verbal abilities (Blair and Razza, 2007). EF is linked to abstract thinking and problem-solving, otherwise known as fluid intelligence (Blair, 2006). More adaptive self-regulation at age 4 is directly linked to greater academic achievement starting at 7 years old through adulthood (McClelland et al., 2013). Children with strong self-regulation skills are also less likely to engage in criminal or problematic behavior as a young person or adult (Moffitt et al., 2011). In sum, EF is a set of cognitive skills that facilitates learning as well as adaptive self-regulation, which is a broader term that encompasses cognitive skills but also refers to children's functioning in social contexts (Bailey and Jones, 2019; Blair and Ku, 2022).

Many research measures of self-regulation as well as EF exist for children under 5 years (Lipsey et al., 2017; McClelland et al., 2022). Strong measures need to be developmentally appropriate, ecologically relevant, and demonstrate strong psychometric properties. Historically, EF had been evaluated in children by using adult assessments extended downward for use in children (Gny and Willis, 1991; Delis et al., 2001). Adult-derived tests, however, did not sufficiently evaluate children's skills and the content was not always relevant to children (Anderson, 1998). Furthermore, traditional cognitive measures of EF were developed in highly controlled laboratory-based settings, which often do not reflect the more dynamic and diffuse self-regulatory demands on children in less formal settings (Salthouse et al., 2003). On the other hand, in schools, self-regulation is typically measured with observer-report checklists or surveys (Zelazo et al., 2016). As the literature has expanded emphasizing the importance of both EF and self-regulation for young children, it is increasingly recognized that measures must be appropriate for use in educational or other naturalistic settings (Franzen and Wilhelm, 1996; McClelland and Cameron, 2012). While many new instruments to measure EF and self-regulation as part of social-emotional learning (SEL) have been developed for children at the formal school transition (Carlson, 2005; Denham et al., 2010), few direct assessments are available for non-researchers.

Overall, improvements in assessment strengthen reliability and validity of an instrument for a specific population or age group. Reliability indicates whether an assessment is consistent, that is,

whether different items measure the same underlying construct; while validity refers to whether the assessment measures the knowledge, skills, or capacities that it is designed to measure (Arizmendi et al., 1981; Hartmann and Pelzel, 2015). There are different ways to demonstrate a measure's validity, including correlations with established instruments that measure the same construct; investigation of demographic or other characteristics known to explain variance in the construct; and correlations with other constructs that are related to, but not the same as, the construct of focus (Cameron Ponitz et al., 2008; Gonzales et al., 2021).

## Importance of direct assessment options for equity in early childhood

While developmentally-appropriate measures have improved accuracy in capturing skill levels, measuring individual differences remains a challenge, especially for children whose regulatory skill development is nascent (Willoughby et al., 2012; Gonzales et al., 2021). Furthermore, recent efforts have focused on measuring these skills in a way that is fair and equitable in a society where systemic oppression limits opportunities for Black, brown, and poor people (Miller-Cotto et al., 2022). For example, Miller-Cotto et al. note that current conceptualizations and measurement of EF and self-regulation are rooted in decades of research on primarily white children from privileged backgrounds. They urge the “repositioning of executive functions as skills developed through task–environment exchanges” (p. 6). Part of this effort means recognizing that how EF develops and even how it is measured is itself a cultural enterprise, which mostly white and privileged researchers have historically overseen.

With equity goals in mind, we draw from prior research showing that EF and self-regulation improve with age, especially in the years from 3 to 5 (Garon et al., 2008). The research base also indicates that children from impoverished communities, and those whose parents have obtained less education, demonstrate lower levels of self-regulation and EF compared with same-age children from more resourced backgrounds, whose parents tend to have higher levels of education (Ursache et al., 2016). These patterns have been linked to opportunities to develop and practice EF, which are more common in high-resource homes (Blair and Raver, 2015). Of note, cultural group membership and socioeconomic status (SES) are closely intertwined given the history of power in the U. S. Historical context must be acknowledged when assessing children's regulatory capacities, and when developing new assessments (Miller-Cotto et al., 2022). Historically and today, people racialized as white retain power, privilege, and resources that lead to greater opportunity, on average, than people who are racialized as non-white. This reality necessitates intentionality in developing assessments that have the potential to uncover assets held by children who have been minoritized. Miller-Cotto et al. urge researchers to bring assessment out of historically white spaces, such as the laboratory, and to “celebrate children's ability to persist through real-world distractions and perform complex, playful actions in rapidly changing environments” (p. 9). In other words, researchers must strive to measure children's skills where they are relevant.

This study introduces a measure of self-regulation drawing on EF that can be used within the school context. We acknowledge that like laboratories, U. S. education settings tend to be white-dominated, where assessing self-regulation and EF is a culturally-embedded activity with serious implications for non-white and poor children (Miller-Cotto et al., 2022). Although differences in self-regulation scores by child characteristics may be expected because children have different experiences, as well as energy or attention levels, researchers caution that differences commonly arise from factors outside the child (Mashburn et al., 2006). Equitable early childhood assessment can minimize the extent to which non-child factors contribute to score differences.

Teacher ratings of children's behavior and self-regulatory skills provide a comprehensive view of a student and are used ubiquitously in schools to identify children for behavioral intervention, disciplinary action, and instructional needs. Teacher ratings can be accurate, especially when identifying students in need of academic intervention (Gresham et al., 1987). There is growing evidence, however, that teacher bias also exists and has a significant impact on long term academic outcomes for students (Reardon et al., 2017). Waterman et al. (2012) report that preschool teachers, as compared to extramural research assessors, appear prone to significant bias when rating children's skills.

Social reproduction theory helps to explain how schools replicate social inequalities, particularly racial inequality (Dixon and Rousseau, 2005). These inequalities are exacerbated by racial, ethnic, and cultural misunderstandings between teachers and students (Boykin, 1986; Delpit, 2006). In their study of 701 prekindergarten students across 11 states, Downer et al. (2016) found that Black students are more often the recipients of escalating disciplinary action by white teachers over time. Black students are more often recommended for special education services by white teachers than by Black teachers (Wiley et al., 2013). Implicit bias is the unconscious opinions or attitudes held against different social groups, and implicit bias influences student outcomes in schools (Glock and Kovacs, 2013). The impact of implicit bias can be significant, with teachers' beliefs about student performance resulting in self-fulfilling prophecies for students of color (Papageorge et al., 2016). Some evidence indicates that compared to white teachers, Black teachers hold higher expectations for Black students, and this contributes to more positive outcomes for all students (Gregory et al., 2011).

As another example, gender differences are common in teacher ratings, more so than in direct assessments. Differences usually favor girls (Matthews et al., 2009; Wanless et al., 2013), which some experts attribute to the greater alignment of girls' behavior in classrooms with teacher expectations (Entwisle et al., 2007). In addition to identifying as white or Caucasian, most early childhood teachers are also women. Teachers from multiple countries including the U. S. tend to report that girls have higher classroom self-regulation (Wanless et al., 2011b).

If group-based differences appear in teacher ratings but not direct assessments, that gap raises questions about differences in the assessment contexts and in the individuals responsible for the assessment. Given the pervasiveness of teacher bias in educational systems, multiple modes of assessment including ratings and direct measures can be employed to provide a more holistic, and potentially more equitable, evaluation of children.

## Technology as part of early childhood assessment

In this study we explored an alternative to teacher ratings by adapting a child-friendly research-based measure of self-regulation, through digital technology. Tablet use by children, particularly the use of tablets in all facets of education, has grown over the past few decades (Fletcher et al., 2014). Furthermore, touchscreens have eased technology use for young children (Christakis, 2014; Spawls and Wilson, 2017) and as a result preschool children use technology frequently (Vandewater et al., 2007; Rideout and Katz, 2016). Furthermore, recent studies show that instruction using tablets can be beneficial to enhancing early childhood learning alphabet awareness (De Jong and Bus, 2004; Xie et al., 2018; Griffith et al., 2020) and numeracy support (Outhwaite et al., 2017). Overall, experts recommend that digital technology can be effectively incorporated as part of instructional programming that is of high quality and age appropriate.

The adoption of digital technology in preschool classrooms can only occur with preparedness and engagement of educators. Before COVID-19, teachers were expected to use digital technology as at least a supplement to traditional classroom instruction (Collier et al., 2004; Hernández-Ramos, 2005). Large-scale studies highlight the increase in technology use in classrooms (Barron et al., 2003; Carson et al., 2014; Denham et al., 2020). Teachers are finding technology-based testing easy to administer and there is research to support the benefits of technology-based testing over traditional tests (Tymms, 2001; Martin, 2008). Young students typically enjoy using tablets in particular, and are able to negotiate the use of digital technology with relative ease (Jones and Liu, 1997). Technology-based assessment reduces the time and effort required to administer and score the assessment, as well as to train testing examiners (Denham et al., 2020).

## Rationale for the present study

This study introduces the HTKS-Kids tablet-based assessment of self-regulation, requiring EF. In early childhood education settings, observation-based assessment and behavioral rating scales remain the most common approach to assess school readiness, including EF and self-regulation (Schilder and Carolan, 2014; Isaacs et al., 2015). Direct measures of self-regulation have blossomed in research settings but are still not widely available for preschool programs. Research suggests that both teacher rating scales and direct assessments can predict children's outcomes measured longitudinally (Schmitt et al., 2014). Given that all assessment, including technology-based assessment, is culturally-embedded with potential for equity or bias, it is imperative to understand how a new teacher-facilitated, largely child-led technology-based assessment of self-regulation captures children's skills in a diverse sample. HTKS-Kids is child-friendly and the concept of touching the opposite (head vs. toes) is based on a game designed for naturalistic settings (McCabe et al., 2004), as opposed to laboratory tasks that prioritize standardized administration. Thus, HTKS-Kids may have advantages over other tablet-based EF

measures with origins in the laboratory, which may have more rigid administration requirements (Carlson and Zelazo, 2014).

In this study, we compared a new tablet-based direct assessment of self-regulation with the original research task and teacher-rated classroom self-regulation, keeping implications for equity in mind. We examined preliminary psychometric properties, including measure variability, reliability, and validity; potential sources of difference in children's scores; and finally, evidence that the tablet-based measure provides different information from teacher ratings. We posed the following research questions:

1. Does HTKS-Kids show preliminary validity and reliability as a measure of self-regulation among low-income 4-year-old preschoolers, when compared with the established HTKS-R research measure of self-regulation?
2. How much variation in children's HTKS-Kids scores is due to key background and sociocultural characteristics, including their age, gender, parent education, first language, and ethnicity identified by parents (Black, Latino/a, or white)?
3. What is the association between children's HTKS-Kids scores and teacher ratings of their classroom self-regulation, and do these two measures provide different information, focusing on key child characteristics (gender and ethnicity)?

## Method

The present research questions were posed in the context of a short-term, cross-sectional study lasting from October 2021 to February 2022. We report data collected with children and families including teacher ratings of children (Teddlie and Tashakkori, 2009).

## Teacher sample

At study enrollment, teachers completed a demographic survey on Qualtrics. Two identified as male and nine as female; nine reported their primary/only ethnic group as white with two selecting Latino/a. All but one teacher reported their age range as 26–39 years, with one reporting 40–49 years. Seven teachers had 6–9 years of experience teaching preschool, and the other 4 had 10 or more years. They taught in three different programs in upstate New York, with one person being the only participant from their program.

## Child sample

Children ( $n = 79$ ) were on average 4.4 years old on November 1st, ranging from 3.9 to 4.9 years. The sample was 54% female and 78% of families reported qualifying for the WIC subsidy. The average years of parent education was 12.5, or just over a high school degree, ranging from 3 to 20. Families were asked to identify the child's ethnicity and could endorse as many groups as they liked. Of 78 families reporting this variable, the sample included  $n = 34$  or 39.5% of families who endorsed Black or multiethnic Black,  $n = 21$  or 26.9% Latino (non-Black),  $n = 17$  or 21.8% White only;



with other groups representing 5% or less of the sample including American Indian, Asian, and Middle Eastern. English was reported as the child's first language for 70 or 81.4% of participants; other first languages included Spanish ( $n = 8$ ), Arabic ( $n = 3$ ), Burmese ( $n = 1$ ), and Nepali ( $n = 1$ ).

## Procedures

Participating teachers sent backpack mail and electronic flyers with the demographic questionnaire home to families, resulting in 86 children enrolled in the study. A total of seven children dropped from their preschool programs before the study was completed. Sample reported in this paper range from 71 to 79 children depending on how many study measures were available. The majority of data were collected in late November and December, with data collection complete by mid-February.

Either the PI or research assistant administered the research assessment, HTKS-R (Gonzales et al., 2021; McClelland et al., 2021) to individual children in a quiet hallway or office near the child's classroom. There were no experimenter differences in mean score obtained ( $t_{72} = 1.00$ ,  $p = 0.3$ ). Teachers facilitated the HTKS-Kids measure with all their study children. About half of teachers left the classroom to work with individual children, to facilitate their engagement, and the other half administered HTKS-Kids to individual children in the classroom, during center time. About half (46%) of the sample were given the HTKS-R first; the remaining 54% of children took the traditional research version of the HTKS-R after they had played the HTKS-Kids version on the tablet with their teachers. Finally, teachers rated each children's classroom self-regulation. Most teachers completed the CBRS on paper, though it was also available electronically.

## Measures

We collected several instruments one time on each child.

### Child demographics

Children's primary caregivers completed a demographic survey for their child upon entry into the study. These were done on paper or electronically. We obtained demographic information for 85 children.

### Traditional HTKS-R assessment

In the first part, called Opposites, children were told to say "head" if the examiner says "toes" and vice versa. Then children were asked to "touch your head" if told to touch their toes and vice versa. Children who did well on the "touch" commands advanced and were taught to touch knees when told to touch shoulders and vice versa, with one of four commands being given (head, toes, knees, shoulders). Finally, if they did well on that part, the rules switched and they were taught to touch their head when told to touch their knees, and touch their toes when told to touch their shoulders. HTKS-R items were scored 0 (incorrect), 1 (self-correct), or 2 (correct); with 1 indicating the child made an initial movement

to the wrong body part but then self-corrected to the correct body part. The HTKS-R has been shown to demonstrate strong reliability and validity in diverse samples of young children (Gonzales et al., 2021; McClelland et al., 2021).

### HTKS-Kids tablet assessment

The HTKS-Kids tablet-based version of HTKS-R includes two formats where the teacher is first more involved, and then less involved. In the first part, Opposites, the child sat next to the teacher who was holding the tablet, and the child listened to the tablet app instructions ("If I say head, you say toes"), and then stated their answer verbally. Teachers entered on the tablet whether the child's response was head or toes. After these items, the teacher handed the tablet to the child and listened to instructions that were analogous to HTKS-R items, interacting with an animated panda on the touchscreen instead of their own body (i.e., the tablet would say, "If I say tap panda's head, you tap panda's toes"). The teacher remained next to the child to facilitate engagement. HTKS-Kids included four sections with the same rules as in the HTKS-R: Opposites (spoken) and Parts 1, 2, and 3 (child touches panda's body parts instead of their own). HTKS-Kids items were scored 0 (incorrect), 1 (self-correct), 2 (correct) in the Opposites section that was teacher-mediated; and 0 (incorrect) or 2 (correct) in the other (panda) sections.

We created a short HTKS-Kids training video and handout and discussed it with teachers. They were asked to play the HTKS-Kids tablet assessment once with each study child, entering children's first names and last initial. Teachers or the study RA exported the data to a secure folder that only the research team could access. Two teachers accidentally gave the HTKS-Kids assessment to 12 children more than once (from 2 to 5 times). Teachers played HTKS-Kids with 79 children, including 75 children who also took the HTKS-R.

### Key differences in HTKS-R and HTKS-Kids

We note several key differences between HTKS-R and HTKS-Kids. First, HTKS-Kids removed the gross motor component that is a defining feature of the HTKS-R. Thus, the HTKS-Kids requires children to apply their EF while self-regulating to sit, speak, and hold and touch the tablet on this revised version of the task. Second, we eliminated from 1 to 3 items in each of the four HTKS-Kids sections to reduce overall assessment time and increase engagement; this means there are fewer total HTKS-Kids items (38) than HTKS-R items (59). Third, self-correct scores were not an option for children in the HTKS-Kids panda sections (Parts 1, 2, and 3), because training children on how to change their answer on the tablet was too complicated. Finally, HTKS-R was given by an assessor previously unknown to the child in a quiet area outside the classroom, whereas HTKS-Kids was given by the child's regular teacher in different settings determined by the teacher, including within the classroom while other activity was happening around them. These differences mean that the tasks pose varying EF and self-self-regulatory demands and make a study comparing HTKS-Kids and HTKS-R scores important.

## Child behavior rating scale

Teachers rated each study child using a 5-pt Likert-style scale on 10 items from the Child Behavior Rating Scale (CBRS; Bronson, 1994) that represent children's ability to demonstrate self-regulation in the complex context of the classroom. Example items from the classroom self-regulation subscale include "observes rules and follows directions without reminders," and "returns to unfinished task after interruption." The CBRS classroom self-regulation composite has been shown to be reliable and valid in diverse groups of children (Matthews et al., 2009; Wanless et al., 2011a). Previously reported correlations between CBRS classroom self-regulation and earlier versions of HTKS-R vary; in a preschool sample of 247 children with similar characteristics, the correlation was  $r = 0.35$  (Schmitt et al., 2014). The reliability for CBRS items in this study was high at  $\alpha = 0.95$ . We obtained CBRS ratings for 80 children and calculated a mean score composite from the 10 items for use in analyses.

## Analytic approach

We used EpiData for paper data entry including double entry of HTKS-R forms and CBRS rating scales, Excel for data management and preparation; and SPSS 27 (IBM Corp, 2021) and Mplus 8.0 (Muthén and Muthén, 1998) with MCAR estimator.

For RQ1: We analyzed only those items that were the same in HTKS-R and HTKS-Kids, with each tasks' score maximum therefore being 76 points for 38 total items. Practice items and test items counted toward this total of 38 items, and all items were included when creating composites. We analyzed both raw HTKS-R scores, and rescaled HTKS-R where for Parts 1, 2, and 3, self-correct scores of 1 were recoded as 2 to match the scale on the HTKS-Kids panda sections. Prior analyses have shown that a self-correct score on HTKS is statistically similar to a score of 2 (Bowles et al., n.d.).<sup>1</sup> For all analyses comparing HTKS-R and HTKS-Kids task items or composites, we used the rescaled HTKS-R scores. Finally, because separate task sections include different numbers of items, we calculated sum scores but also mean scores to facilitate task and composite comparisons.

For RQ2: To understand sources of variability in HTKS-Kids scores, we performed stepwise regressions where HTKS-Kids sum score was regressed on age, then we added parent education, then we added first language other than English, then we added whether the child was female, and finally we added whether the child was Black or Latino.

For RQ3: To understand whether HTKS-Kids provided different information from teacher-rated classroom self-regulation, we first examined correlations between and mean differences in each measure. We then ran simple  $t$ -tests for four different groups (female vs. male, Black or not Black, Latino vs. non-Latino, and white vs. non-White) for the normally-distributed CBRS scores. We used the Mann-Whitney  $U$  statistic for the non-normal HTKS-Kids sum score. Finally, to assess whether any simple mean differences remained statistically significant after adjusting for key background characteristics, we conducted linear regressions in Mplus using the

MLR estimator controlling for important background variables and utilizing all available data.

## Results

### RQ1: HTKS-Kids preliminary reliability and validity

We found that item scores were highly similar between the rescaled HTKS-R items and HTKS-Kids assessment. We ran selected pairwise comparison  $t$ -tests to see whether differences of magnitude 0.20 or above were statistically significant. Children scored the same on analogous task items, except scores were significantly lower on the first three practice items on Part 1 on HTKS-Kids, compared to the same items on the HTKS-R.

### Range and distribution

Overall, composite and total scores were highly similar between the HTKS-R and HTKS-Kids versions (see Table 1). A small number of children scored at floor (<6% of the sample) on each task version. Of note, HTKS-Kids achieved an important objective for score distribution; only 6% or 5 children scored at floor. Recent work on the HTKS-R with a low-income sample of children in Oregon indicated that 3% of 4-year-old children scored at floor (Gonzales et al., 2021). In this study, both tasks showed a positive skew, meaning that fewer children achieved higher scores and the bulk of the sample scored below the mean. The HTKS-Kids distribution was bimodal, with no scores falling between 25 and 35. While bimodal scores are typical of this assessment (Cameron Ponitz et al., 2008), this pattern of distribution was pronounced with the HTKS-Kids and led to our using non-parametric analyses for subsequent analyses. We employed Mann-Whitney  $U$  tests and the MLR estimator in Mplus which is appropriate for non-normally distributed data.

### Inter-item and test-retest reliability for HTKS-Kids

Our study was not designed to assess reliability of HTKS-Kids, but we calculated alpha values for HTKS-Kids items, and also examined test-retest reliability for the handful of children whose teachers mistakenly gave them HTKS-Kids more than once.

Inter-item reliability for the 38 HTKS-Kids items was excellent, at  $\alpha = 0.95$ . Twelve children played HTKS-Kids 2, 3, 4, or 5 times over a 2-week period. Within children, test-retest reliability for the first and second occasions was excellent with Cronbach's alpha of 0.89, and inter-item correlation of 0.80. For these 12 children, the average duration between the 1st and 2nd occasion was 3 days, with a range from 1 to 6 days. Additionally, and for descriptive purposes only given that small numbers of children took HTKS-Kids 3, 4, or 5 times, Table 2 shows every sum score for each child across up to five occasions (T1–T5). With the exception of Child H, whose score improved dramatically after T1, most children scored within a fairly narrow range after repeated attempts.

### Correlations between HTKS-R and HTKS-Kids

HTKS-R and HTKS-Kids scores were positively correlated at  $r = 0.60$ ;  $r = 0.59$  controlling for age. This magnitude is higher

<sup>1</sup> Bowles, R. P., McClelland, M. M., Cameron, C. E., Acock, A. C., Montroy, J. J., and Duncan, R. J. (n. d.). *A Rasch Analysis of the Head-Toes-Knees-Shoulders (HTKS) Measure of Behavioral Self-Regulation*. (Unpublished).

TABLE 1 Composite descriptives by task and task section.

	Sum		Mean		SD		Max	
	Kids	HTKR	Kids	HTKR	Kids	HTKR	Kids	HTKR
Total	21.7	21.8	0.54	0.54	0.47	0.49	1.70	1.90
Opposites	9.9	10.5	1.41	1.51	0.71	0.65	2.00	2.00
Part 1	6.3	6.6	0.52	0.55	0.58	0.65	1.83	2.00
Part 2	3.9	3.1	0.36	0.28	0.61	0.59	2.00	2.00
Part 3	1.6	1.6	0.16	0.16	0.35	0.43	1.40	2.00
	Skewness		Kurtosis		% (#) Floor		% (#) Ceiling	
	Kids	HTKR	Kids	HTKR	Kids	HTKR	Kids	HTKR
Total	0.94	1.35	-0.35	0.88	6 (5)	8 (6)	0	0
Opposites	-0.83	-1.27	-0.78	0.36	9 (7)	9 (7)	47 (37)	42 (33)
Part 1	0.93	1.24	-0.68	0.10	25 (20)	27 (21)	0	3 (2)
Part 2	1.63	1.96	1.44	2.40	67 (53)	77 (60)	6 (5)	1 (1)
Part 3	2.44	2.94	5.26	8.00	76 (60)	83 (65)	0	3 (2)

HTKS-Kids N = 79. HTKS-R (HTKR) N = 78. All values in this table are based on mean scores (with range 0–2) except the first set of descriptives which is for sum scores.

TABLE 2 Sum scores on HTKS-Kids over up to five occasions within n = 12 children.

Child	T1	T2	T3	T4	T5
A	0	4	6		
B	0	1	2		
C	8	6			
D	9	20	16		
E	12	18	20		
F	14	16			
G	14	18	40	34	
H	17	50	60	48	
I	38	40	34		
J	42	58	56		
K	44	36	50	44	42
L	58	44	44		

than correlations among different self-regulation and EF measures, which tend to fall around  $r = 0.3$  or  $0.4$ ; one exception is that among 4-year-olds, the HTKS-R correlates above  $r = 0.54$  with the Dimensional Change Card Sort (DCCS) task (McClelland et al., 2014). One small study found the correlation between the HTKS given to 25 children by researchers and teachers is  $r = 0.99$  (McClelland et al., 2015a; see Table 3).

### RQ2: sources of variability in HTKS-Kids scores

In the first regression model, older children scored significantly higher on HTKS-Kids scores, but the overall model did not explain variance that differed from zero (see Table 4). In each step that

included parent education, children whose parents reported higher education levels scored significantly higher on HTKS-Kids,  $p < 0.01$ . The magnitude of this association was modest: when parents reported 1 SD, or 2.3 years higher than the mean 12.5 years of education, children scored from 0.25 to 0.37 points higher on HTKS-Kids.

In the last model we tested, children whose families identified them as Latino scored significantly lower,  $p < 0.05$ ; however, this model did not explain more variance from the model without Black or Latino indicator variables. Thus, we conclude that in this sample, the only meaningful predictor of children’s HTKS-Kids score was their parent’s level of education.

### RQ3: teacher rated compared to tablet-assessed self-regulation (HTKS-Kids)

HTKS-Kids scores were moderately and positively correlated with teacher ratings of classroom self-regulation, with  $r = 0.40$ . This is similar to correlations between CBRS and HTKS in other studies, with  $r = 0.29$  (Cameron Ponitz et al., 2009) and  $r = 0.35$  (Schmitt et al., 2014; see Table 3).

In simple comparisons without predictors, we found no differences by gender in either HTKS-Kids or teacher ratings. However, teachers scored Black children as having about 0.3-point or half an SD lower classroom self-regulation compared with non-Black children,  $t = -2.00$ ,  $p < 0.05$ ; and teachers rated white children about 0.4-point (also half-SD) higher on classroom self-regulation than non-white children,  $t = 2.22$ ,  $p < 0.05$ . See raw score differences by key groups in Table 5. We also found that Latino children scored 0.4-point lower on HTKS-Kids than non-Latino children,  $z = -2.82$ ,  $p < 0.01$ .

Tables 6A–D show standardized results for the Mplus regressions, which means that the coefficients indicate the percent of SD change in outcome score given a 1 SD increase in the predictor variable. Standardized coefficients can also be compared

TABLE 3 Zero-order correlations by task and task section (above diagonal).

	1	2	3	4	5	6	7	8	9	10	11
1. Kids total	–	0.60	0.63	0.41	0.94	0.60	0.91	0.54	0.80	0.36	0.42
2. HTKS-R total	0.59	–	0.41	0.59	0.63	0.93	0.51	0.91	0.41	0.80	0.41
3. Kids opposites	0.65	0.39	–	0.54	0.46	0.38	0.37	0.25	0.30	0.18	0.34
4. HTKS-R opposites	0.41	0.58	0.51	–	0.33	0.45	0.30	0.31	0.21	0.24	0.33
5. Kids part 1	0.94	0.63	0.50	0.35	–	0.63	0.84	0.60	0.71	0.44	0.38
6. HTKS-R part 1	0.58	0.92	0.36	0.43	0.62	–	0.51	0.80	0.44	0.64	0.39
7. Kids part 2	0.92	0.50	0.39	0.31	0.85	0.50	–	0.50	0.72	0.29	0.35
8. HTKS-R part 2	0.52	0.91	0.23	0.30	0.59	0.80	0.48	–	0.37	0.78	0.36
9. Kids part 3	0.80	0.39	0.32	0.21	0.73	0.41	0.71	0.35	–	0.25	0.31
10. HTKS-R part 3	0.34	0.80	0.17	0.23	0.42	0.63	0.28	0.78	0.24	–	0.31
11. CBRS self-reg mean	0.40	0.38	0.34	0.30	0.38	0.35	0.34	0.34	0.25	0.21	–

Correlations with age partialled out (below diagonal).

Pearson correlations used pairwise deletion. Ns range from 74 to 79 depending on the measures.

All correlations above 0.30 are  $p < 0.01$ . Correlations 0.24–0.29 are  $p < 0.05$ . Correlations 0.20 are  $p < 0.10$ .  $n = 74$ .

TABLE 4 HTKS-Kids sum scores regressed on key child characteristics.

	Step 1 ( $n = 86$ )		Step 2 ( $n = 86$ )		Step 3 ( $n = 83$ )		Step 4 ( $n = 82$ )		Step 5 ( $n = 77$ )	
	Coeff.	$t$	Coeff.	$t$	Coeff.	$t$	Coeff.	$t$	Coeff.	$t$
Int.	–2.10	–1.38	–2.95	–2.17*	–2.90	–1.99*	–2.88	–1.96*	–1.80	–1.19
Age	0.22	2.12*	0.17	1.64	0.13	1.23	0.13	1.22	0.11	1.03
Par. Edu.			0.29	2.63**	0.37	3.95**	0.37	3.94**	0.25	2.71**
Lan. Oth.					0.11	0.88	0.11	0.87	0.21	1.60
Female							0.00	–0.02	0.02	0.18
Black									–0.07	–0.51
Latino									–0.32	–2.53**
R-sq.	0.05	1.06	0.13	1.97*	0.18	2.77**	0.18	2.78**	0.17	2.94**

Par. Edu., parent education; Lan. Oth., family reported a language other than English as child's first language.

\* $p < 0.05$ .

\*\* $p < 0.01$ .

TABLE 5 Raw score means and SDs for distinct subgroups.

	HTKS-Kids				CBRS		
	N	M	SD		N	M	SD
Female	41	22.05	19.77	Female	42	3.58	0.78
Male	37	21.32	17.78	Male	42	3.44	0.69
Black	30	20.83	15.75	Black	29	3.24	0.68
White	16	26.50	22.28	White	17	3.84	0.76
Latino	20	13.40	15.91	Latino	21	3.30	0.69



**TABLE 6A** No gender differences on HTKS-Kids or teacher-reported CBRS.

	HTKS-Kids		CBRS	
	Coeff.	t	Coeff.	t
Int.	-2.88	-1.96 <sup>+</sup>	-1.94	-1.40
Age	0.13	1.22	0.34	3.78**
Par. Edu.	0.37	3.94**	0.26	3.29**
Lan. Oth.	0.11	0.87	0.05	0.52
Female	0.00	-0.02	0.12	1.15
Total r-sq	0.18	2.78**	0.23	2.82**

N = 78.  
 Par. Edu., parent education; Lan. Oth., family reported a language other than English as child's first language.  
 \*\*p < 0.01.  
<sup>+</sup>p = 0.05.

**TABLE 6B** Teachers rate Black children lower on CBRS but HTKS-Kids scores do not differ.

	HTKS-Kids		CBRS	
	Coeff.	t	Coeff.	t
Int.	-2.91	-1.99*	-1.52	-1.08
Age	0.13	1.21	0.33	3.62**
Par. Edu.	0.37	4.00**	0.26	3.57
Lan. Oth.	0.12	0.93	-0.03	-0.29
Female	0.12	0.00	0.11	1.08
Black	0.04	0.37	-0.25	-2.41*
Total r-sq	0.18	2.85**	0.28	3.84**

N = 78.  
 Par. Edu., parent education; Lan. Oth., family reported a language other than English as child's first language.  
 \*p < 0.05.  
 \*\*p < 0.01.

within the same model. Overall, we found the same pattern of results as with simple group comparisons.

### Gender

There were no significant differences by gender in HTKS-Kids or CBRS teacher ratings in the Mplus regressions controlling for age, parent education, and language status (see Table 6A).

### Black vs. non-Black

There were no differences in HTKS-Kids scores for children their families identified as Black vs. children not identified as Black. In contrast, teachers rated Black children lower on CBRS classroom self-regulation: the coefficient of -0.25 if the child was Black was similar in magnitude to the 0.28 coefficient for parent education (see Table 6B).

### Latino vs. non-Latino

In regressions controlling for background variables including first language, there were no score differences in teacher-rated CBRS classroom self-regulation, but Latino children scored

**TABLE 6C** Latino children score lower on HTKS-Kids but teacher ratings do not differ.

	HTKS-Kids		CBRS	
	Coeff.	t	Coeff.	t
Int.	-1.87	-1.23	-1.10	-0.75
Age	0.11	0.99	0.31	3.40**
Par. Edu.	0.26	2.82**	0.22	2.30*
Lan. Oth.	0.22	1.80 <sup>+</sup>	0.12	1.15
Female	0.02	0.13	0.10	0.86
Latino	-0.28	-2.90**	-0.16	-1.26
Total r-sq	0.17	2.99	0.20	2.23*

N = 78.  
 Par. Edu., parent education; Lan. Oth., family reported a language other than English as child's first language.  
 \*p < 0.05.  
 \*\*p < 0.01.  
<sup>+</sup>p = 0.07.

**TABLE 6D** Teachers rate white children higher on CBRS but HTKS-Kids scores do not differ.

	HTKS-Kids		CBRS	
	Coeff.	t	Coeff.	t
Int.	-2.50	-1.63	-1.60	-1.12
Age	0.12	1.07	0.32	3.57**
Par. Edu.	0.30	3.29**	0.23	2.77**
Lan. Oth.	0.13	1.01	0.09	0.93
Female	-0.01	-0.10	0.13	1.26
White	0.13	1.11	0.29	3.14**
Total r-sq	0.15	2.72**	0.27	3.13**

N = 78.  
 Par. Edu., parent education; Lan. Oth., family reported a language other than English as child's first language.  
 \*\*p < 0.01.

significantly lower on HTKS-Kids. The magnitude of this difference if the child was identified as Latino, -0.25, was again similar to the coefficient of 0.27 for parent education (see Table 6C).

### White vs. non-white

There were no differences for white vs. non-white children on HTKS-Kids scores; but teachers rated white children higher on CBRS classroom self-regulation. The coefficient of 0.31 if the child was white was similar in magnitude to the coefficient of 0.24 for parent education (see Table 6D).

## Discussion

The present study examined the initial psychometric properties of a tablet-based version of a popular research measure of self-regulation requiring EF, called HTKS-Kids, including variability, reliability, and validity using comparison with the established HTKS-R measure. We investigated potential sources of difference

in children's scores and also examined whether scores on HTKS-Kids correlated with, and provided different information from teacher ratings of these skills in the classroom. We report three main findings. First, the new HTKS-Kids tablet measure of self-regulation facilitated by preschool teachers showed early evidence of reliability and validity. Second, the best predictor of tablet-based HTKS-Kids self-regulation score was parent education. Third, correlation of HTKS-Kids with teacher ratings was moderate; further, teachers rated Black children lower, but white children higher, on classroom ratings, whereas these differences did not appear in children's HTKS-Kids scores.

## HTKS-Kids showed strong internal consistency and correlated with HTKS-R

HTKS-Kids captured individual differences in self-regulation among 4-year-old children. Additionally, a subgroup of children who took the HTKS-Kids more than once within a short time period achieved highly similar scores. Based on correlations with the existing, established measure of self-regulation, HTKS-Kids measured self-regulation: Children who scored higher on HTKS-Kids also scored higher on the original HTKS-R research task, and were rated with better classroom self-regulation by their teachers. These findings all suggest that HTKS-Kids is a promising new tablet-based, direct measure of self-regulation that can be used inside preschool classrooms and facilitated by early childhood teachers. As assessment demands on teachers increase, practical measures that provide direct information on children's skills across a range of school readiness domains become more important (Maves, 2022).

We found very few differences when comparing analogous individual items on the two measures, except children scored lower on the first few HTKS-Kids items where they interacted directly with the tablet, as compared with corresponding items on the HTKS-R. Their lower HTKS-Kids item scores make sense because those particular HTKS-Kids items represent a transition from listening to the teacher and speaking responses, to holding and pressing the tablet screen. This transition may pose higher self-regulatory demands than corresponding items on the HTKS-R where there is a transition from listening/speaking, to listening/responding with gross motor movements without a tablet involved.

The correlation of  $r = 0.60$  between the total scores on HTKS-Kids and HTKS-R was not as high as we expected in a task that includes the same items delivered in a different format and setting. As we previously noted, HTKS-Kids included fewer overall items as well as greater variation in how teachers administered the task (inside or outside the classroom, though we do not have this information at the child level). In contrast, the HTKS-R was administered by a researcher unknown to the child, in a relatively quiet space outside the classroom. These differences likely accumulated, resulting in a lower-than-expected correlation. The correlation alone, however, does not mean that HTKS-Kids is not providing valuable information about children's self-regulation; as noted previously,

the HTKS-R and DCCS are correlated around 0.60 and both are considered robust measures of self-regulation that require children to apply EF processes to their behavior (McClelland et al., 2014). Future research should continue to test HTKS-Kids and HTKS-R associations with larger and more diverse samples of children.

## Parent education explained the most variance in children's HTKS-Kids scores

The best predictor of children's HTKS-Kids scores was parent education. Parent education is a proxy for early learning experiences and resources, which are consistently linked to EF development and overall self-regulation (Davis-Kean, 2005; Waters et al., 2021). Child age was not a significant predictor of HTKS-Kids scores, but this may be due to the combination of the relatively narrow variation in age and the sample size under 80. Importantly, we did not find that whether children were identified by their parents as Black or white explained any variance in HTKS-Kids scores. Given that race is a social, not biological construct, it is encouraging that children performed similarly on the new HTKS-Kids assessment regardless of this identity variable.

Although sub-sample sizes were small, we found that non-Black, Latino-identified children had lower HTKS-Kids scores, even after controlling for parent-reported child first language status and parent education. Another study found that among low-income children, Latino children showed less developed self-regulation and improved more slowly over time as compared to white children (Wanless et al., 2011b). We note that our analyses are based on small subgroup numbers: only eight of 17 Latino children were identified by their families as speaking Spanish as their first language, with the other nine Latino children identified with a first language of English. All our study children were given only the English version of HTKS-Kids, so it is not possible to address whether Spanish-language administration may have improved scores. Therefore, the extent to which language and culture played a role in HTKS-Kids performance was not possible to fully explore in this study but needs to be better understood.

## Is HTKS-Kids providing different information for Black children than teacher ratings?

As in other studies including the HTKS (Cameron Ponitz et al., 2009; Schmitt et al., 2014), HTKS-Kids was also moderately positively correlated with teacher ratings of self-regulation. Modest or moderate positive correlations are common among measures of regulatory processes that vary in design features (Rimm-Kaufman et al., 2009; Vitiello et al., 2011; McClelland et al., 2014), such as different formats (e.g., paper or tablet), settings (e.g., individualized or naturalistic), response modalities (e.g., gross motor actions, points, or key presses), and administrators (researchers or teachers). Teacher ratings are based on their

observations and interactions with children in their classroom over several weeks or months, and the CBRS items in particular asked teachers for their aggregate impressions on how well children manage attention, materials, and behavior across various learning situations. On the other hand, HTKS-Kids performance reflects a score derived from a single individualized assessment where the teacher was present in one-on-one interaction. Distractions from peers did vary, because some teachers reported that they took children outside the classroom to play HTKS-Kids, though we did not collect information on each child's specific assessment context. Teacher discretion on administration setting is an important part of naturalistic assessment, but this variability along with the other differences between CBRS ratings and HTKS-Kids are likely sources of other findings based on child demographic characteristics. Specifically, teachers rated children similarly regardless of Latino ethnicity, but we found other differences in teacher ratings of children's classroom self-regulation if the child was Black. That is, teachers rated white children more favorably than non-white children, and Black children worse than non-Black children.

In explaining these findings, the literature on implicit bias must be considered along with the aforementioned discussion of assessment differences. First is the possibility that there is some "true" difference in children's self-regulatory behaviors in the classroom; for example, perhaps Black children were able to be as successful as non-Black children in the more structured HTKS-Kids context, but exhibited greater levels of distraction in typical classroom learning settings, leading to lower teacher ratings. We note that "true" self-regulation differences could arise from a classroom system that is less supportive for Black children than for white children. That is, teachers may interact with Black and white children differently in regular classroom interactions, which could lead to Black children responding and self-regulating differently. For example, teachers might unconsciously use a less warm tone with Black children; this inconsistent emotional support could activate the child's nervous system and lead to problems self-regulating as children struggle to focus on the task at hand, possibly worrying about their teachers' attitude toward them (Curby et al., 2013). Similarly, Black children may exhibit some behaviors—such as physical and vocal expressiveness—that Boykin explains are a rich legacy of their African American heritage (Boykin and Allen, 1988), but which may be incorrectly interpreted as indicators of poor self-regulation by some teachers. As a reminder, over 80% of this study's teachers identified as white. Finally, it is possible that teachers see the same behaviors differently depending on children's ethnicity. For example, literature supports the idea that teachers discipline Black children more harshly for offenses compared with white children (Ispa-Landa, 2018).

One of the key goals of this study and the broader program of research is to use technology to develop more equitable assessment for young children. The possibility of HTKS-Kids providing information about children that could enrich their teachers' preexisting views of their potential, based on conclusions from observing the child in traditional classroom situations alone, is an important one in this broader context.

## Limitations and future directions

Overall study results were preliminary but promising, and the small sample size necessitates further work to establish psychometric properties and relatedly (and perhaps most importantly), to identify age-based norms and/or screening cutoffs so that HTKS-Kids can be more useful to early childhood programs. While HTKS-Kids is based on the well-established research instruments HTKS and HTKS-R (Cameron Ponitz et al., 2009; McClelland et al., 2014; Gonzales et al., 2021), HTKS-Kids is also different in meaningful ways, given that it involves children interacting with a tablet, and is teacher-facilitated for use in typical classroom settings. Thus, similar research using HTKS-Kids with a larger sample and to continue to examine validity using more than just the HTKS-R, including other measures of self-regulation and early academic skills, is needed. Statistically, a larger sample can enable analysis of items to see how well HTKS-Kids captures child and item differences, norming to identify average scores for children of a given age, and/or cutoff scores that indicate further assessment for possible intervention is needed. More data with both younger and older children are needed. Because the EF processes that underlie a child's ability to self-regulate are implicated among students with ADHD (Barkley, 2004), a modified, age-appropriate version for older students could provide school psychologists for an additional tool for their work with this population.

Given the cultural-embeddedness of self-regulation and its assessment, broadening the socio-demographic characteristics of children given HTKS-Kids is perhaps the most important. It is promising that the HTKS has been translated into 28 languages and used worldwide, and a meta-analysis indicated no differences in how well HTKS predicted young children's academic achievement by country or cultural context (Kenny et al., 2023). Collecting data with a larger sample of racially diverse children can help establish whether HTKS-Kids could mitigate implicit bias against Black children, which may contribute to the differences we found in teacher ratings of children's classroom self-regulation. Future research should also examine the HTKS-Kids in larger samples of multi-lingual children, perhaps using a screening tool to determine a child's need for Spanish language assessment, and subsequently, a Spanish version of HTKS-Kids. Other potential directions include offering children a choice about their language of assessment, and/or incorporating symbols, which may help broaden the task beyond English- and Spanish-speakers.

Future applications for HTKS-Kids are broad and could have major impact. Within early childhood systems, screenings help provide children access to early intervention services (AAP Council on Early Childhood and AAP Council on School Health, 2016; Bertram and Pascal, 2016). One future longitudinal study could examine the potential for HTKS-Kids to predict referral to special education or intervention services. HTKS-Kids was also designed to be incorporated into a holistic assessment system, specifically Cognitive Toybox ([cognitivetoybox.com](http://cognitivetoybox.com)), which provides observation and individualized game-based assessment measures across whole child development. This design allows HTKS-Kids to be used in conjunction with other academic and

non-academic measures to achieve a full understanding of a child's school readiness (Tripathy et al., 2020).

A strong and scalable instrument that measures self-regulation with reliability and validity and that is both child- and teacher-friendly could change the face of kindergarten entry assessment. Common observational tools like Teaching Strategies Gold (TS Gold) are resource-intensive, and misuse of TS Gold and other early childhood assessment tools is common (Ackerman and Lambert, 2020; Olson and Lepage, 2022; Cameron et al., 2023). And as screeners grow more widespread—used in more than half of U. S. states in 2022—currently available tools like the Brigance remain focused on academic and related skills, which are highly dependent on family resources and are not always culture-fair (Olson and Lepage, 2022). Yet EF processes and the overall self-regulation it supports form the foundation of whether children can learn from academic opportunities (Blair and Ku, 2022). Self-regulation is also part of social-emotional learning (SEL) which is increasingly recognized as critical to support. Tellingly, in our collaborations, programs have had to supplement Creative Curricula, which is used by the large majority of Head Start programs, with other programs such as Second Step that more intentionally support SEL. TS Gold also poses heavy burdens on teachers which can take away from their time to interact effectively with children (Kim, 2016; Cameron et al., 2023). In other words, both kindergarten and preschool programs stand to benefit from a child-friendly, teacher-friendly, scalable tablet-based assessment of self-regulation requiring EF.

## Conclusion

Equitable direct assessment of self-regulation is increasingly sought by early childhood systems. Tablet-based assessments can directly measure children's skills in several learning domains and reduce teacher burden. This study suggests that the HTKS-Kids tablet-based self-regulation assessment requiring EF captures individual differences among children with item scores, variability, and floor effects similar to the original HTKS-R task; measures skills that are similar to this established research task; positively relates with teacher ratings of children's classroom skills; and provides a different picture from teacher ratings of children's classroom self-regulation, which was especially evident for Black children. Assessing the potential of children from historically oppressed groups is crucial equity work, and practical tools can support often-under resourced early childhood professionals and programs. Future efforts should further test and refine HTKS-Kids. These early results point to the potential for equitable direct assessment of self-regulation, a capacity that forms the foundation for children's success in and beyond school.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material,

further inquiries can be directed to the corresponding author.

## Ethics statement

The study involving humans was approved by University at Buffalo Institutional Review Board. The study was conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

CC created the HTKS-Kids items for use on a tablet, conceptualized the manuscript, oversaw data collection, performed all analyses, drafted Results and Discussion sections, and readied the paper for publication. MM helped to create the HTKS-Kids items for use on the tablet, helped design the study and craft research questions, advised on analyses, and refined all manuscript sections. TK contributed to Cognitive Toybox programmed HTKS-Kids for use on the tablet and helped to refine all manuscript sections. KS performed most of the data collection and data entry and drafted the Method section. TL drafted the literature review and helped refine the Discussion section. All authors contributed to the article and approved the submitted version.

## Funding

This publication is based on research funded by the Bill & Melinda Gates Foundation (INV-034807). The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

## Conflict of interest

TK was employed by Cognitive ToyBox, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



## References

- AAP Council on Early Childhood and AAP Council on School Health (2016). The pediatrician's role in optimizing school readiness. *Pediatrics* 138, 2293. doi: 10.1542/peds.2016-2293
- Ackerman, D. J., and Lambert, R. (2020). Introduction to the special issue on kindergarten entry assessments: policies, practices, potential pitfalls, and psychometrics. *Early Educ. Dev.* 31, 629–631. doi: 10.1080/10409289.2020.1769302
- Anderson, V. (1998). Assessing executive functions in children: biological, psychological, and developmental considerations. *Neuropsychol. Rehabil.* 8, 319–349. doi: 10.1080/713755568
- Arizmendi, T., Paulsen, K., and Domino, G. (1981). The Matching Familiar Figures Test: a primary, secondary, and tertiary evaluation. *J. Clin. Psychol.* 37, 812–818.
- Bailey, R., and Jones, S. M. (2019). An integrated model of regulation for applied settings. *Clin. Child Fam. Psychol. Rev.* 22, 2–23. doi: 10.1007/s10567-019-00288-y
- Barkley, R. A. (2004). "Attention-deficit/hyperactivity disorder and self-regulation: taking an evolutionary perspective on executive functioning," in *Handbook of Self-regulation: Research, Theory, and Applications*, eds. R. F. Baumeister and K. D. Vohs (Guilford), 301–323. Available online at: <http://search.epnet.com/login.aspx?direct=true&db=psyh&an=2004-00163-014>
- Barron, A. E., Kemker, K., Harmes, C., and Kalaydjian, K. (2003). Large-scale research study on technology in K–12 schools: technology integration as it relates to the National Technology Standards. *J. Res. Technol. Educ.* 35, 489–507. doi: 10.1080/15391523.2003.10782398
- Bertram, T., and Pascal, C. (2016). *Early Childhood Policies and Systems in Eight Countries: Findings From IEA's Early Childhood Education Study*. New York, NY: Springer International.
- Blair, C. (2002). School readiness: integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *Am. Psychologist* 57, 111–127. doi: 10.1037/0003-066X.57.2.111
- Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behav. Brain Sci.* 29, 109–160. doi: 10.1017/S0140525X06009034
- Blair, C., and Ku, S. (2022). A hierarchical integrated model of self-regulation [Conceptual Analysis]. *Front. Psychol.* 13:725828. doi: 10.3389/fpsyg.2022.725828
- Blair, C., and Raver, C. C. (2015). School readiness and self-regulation: a developmental psychobiological approach. *Ann. Rev. Psychol.* 3, 711–731. doi: 10.1146/annurev-psych-010814-015221
- Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x
- Boykin, A. W. (1986). "The triple quandary and the schooling of African-American children," in *The School Achievement of Minority Children*, ed. U. Neisser (Mahwah, NJ: Erlbaum), 57–92.
- Boykin, A. W., and Allen, B. A. (1988). Rhythmic-movement facilitation of learning in working-class Afro-American children. *J. Genet. Psychol.* 149, 335–347. doi: 10.1080/00221325.1988.10532162
- Bronson, M. B. (1994). The usefulness of an observational measure of young children's social and mastery behaviors in early childhood classrooms. *Early Childh. Res. Quart.* 9, 19–43. doi: 10.1016/0885-200690027-2
- Bull, J., and McKenna, C. (2003). *A Blueprint for Computer-Assisted Assessment*. London: Routledge.
- Cameron Ponitz, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., and Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childh. Res. Quart.* 23, 141–158. doi: 10.1016/j.ecresq.2007.01.004
- Cameron Ponitz, C. E., McClelland, M. M., Matthews, J. S., and Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Dev. Psychol.* 45, 605–619. doi: 10.1037/a0015365
- Cameron, C. E., Kenny, S., and Chen, Q. H. (2023). How Head Start professionals use and perceive Teaching Strategies Gold: associations with individual characteristics including assessment conceptions. *Teach. Teacher Educ.* 121, 103931. doi: 10.1016/j.tate.2022.103931
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Dev. Neuropsychol.* 28, 595–616. doi: 10.1207/s15326942dn2802\_3
- Carlson, S. M., and Zelazo, P. D. (2014). *Minnesota Executive Function Scale: Test Manual*. Saint Paul, MN: Reflection Sciences.
- Carson, K., Boustead, T., and Gillon, G. (2014). Predicting reading outcomes in the classroom using a computer-based phonological awareness screening and monitoring assessment (Com-PASMA). *Int. J. Speech-Lang. Pathol.* 16, 552–561. doi: 10.3109/17549507.2013.855261
- Christakis, D. A. (2014). Interactive media use at younger than the age of 2 years: time to rethink the American Academy of Pediatrics guideline? *J. Am. Med. Assoc. Pediatr.* 168, 399–400. doi: 10.1001/jamapediatrics.2013.5081
- Collier, S., Weinburgh, M. H., and Rivera, M. (2004). Infusing technology skills into a teacher education program: change in students' knowledge about and use of technology. *J. Technol. Teacher Educ.* 12, 447–468. Available online at: <https://link.gale.com/apps/doc/A123128011/AONE?u=anon~93c448&sid=googleScholar&xid=ab953524>
- Curby, T. W., Brock, L. L., and Hamre, B. K. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Educ. Dev.* 24, 292–309. doi: 10.1080/10409289.2012.665760
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of students' knowledge and the home environment. *J. Fam. Psychol.* 19, 294–304. doi: 10.1037/0893-3200.19.2.294
- De Jong, M. T., and Bus, A. G. (2004). The efficacy of electronic books in fostering kindergarten children's emergent story understanding. *Read. Res. Quart.* 39, 378–393. doi: 10.1598/RRQ.39.4.2
- Delis, D. C., Kaplan, E., and Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. Washington, DC: PsycTests.
- Delpit, L. (2006). *Other People's Children: Cultural Conflict in the Classroom*. New York, NY: The New Press.
- Denham, S. A., Ferrier, D. E., and Bassett, H. H. (2020). Preschool teachers' socialization of emotion knowledge: considering socioeconomic risk. *J. Appl. Dev. Psychol.* 69, 101160. doi: 10.1016/j.appdev.2020.101160
- Denham, S. A., Ji, P., and Hamre, B. (2010). Compendium of preschool through elementary school socialemotional learning and associated assessment measures. Deliverable for Assessments for Social, Emotional, and Academic Learning with Preschool/Elementary-School Children. Collaborative for Academic, and Emotional Learning and Social and Emotional Learning Research Group. University of Illinois at Chicago, Chicago, IL. 1–134.
- Diamond, A. (2016). "Why improving and assessing executive functions early in life is critical," in *Executive Function in Preschool-Age Children: Integrating Measurement, Neurodevelopment, and Translational Research*, eds. J. A. Griffin, P. McCurdle, and L. S. Freund (Washington, DC: American Psychological Association), 11–43.
- Dixon, A., and Rousseau, C. (2005). Toward a CRT of education. *CRT Educ.* 2005, 11–30. doi: 10.4324/9781315709796
- Downer, J. T., Goble, P., Myers, S. S., and Pianta, R. C. (2016). Teacher-child racial/ethnic match within pre-kindergarten classrooms and children's early school adjustment. *Early Childh. Res. Quart.* 37, 26–38. doi: 10.1016/j.ecresq.2016.02.007
- Entwisle, D. R., Alexander, K. L., and Olson, L. S. (2007). Early schooling: the handicap of being poor and male. *Sociol. Educ.* 80, 114–138. doi: 10.1177/003804070708000202
- Fletcher, E. N., Whitaker, R. C., Marino, A. J., and Anderson, S. E. (2014). Screen time at home and school among low-income children attending Head Start. *Child Indic. Res.* 7, 421–436. doi: 10.1007/s12187-013-9212-8
- Franzen, M. D., and Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. *Ecol. Valid. Neuropsychol. Test.* 1996, 91–112.
- Garon, N., Bryson, S. E., and Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychol. Bull.* 134, 31–60. doi: 10.1037/0033-2909.134.1.31
- Glock, S., and Kovacs, C. (2013). Educational psychology: using insights from implicit attitude measures. *Educ. Psychol. Rev.* 25, 503–522. doi: 10.1007/s10648-013-9241-3
- Gnys, J. A., and Willis, W. G. (1991). Validation of executive function tasks with young children. *Dev. Neuropsychol.* 7, 487–501. doi: 10.1080/87565649109540507
- Gokierto, R. J., Noble, L., and Littlejohns, L. B. (2013). *Directions for Professional Development: Increasing Knowledge of Early Childhood Measurement*. Charlotte, NC: Head Start Dialog.
- Gonzales, C. R., Bowles, R., Geldhof, G. J., Cameron, C. E., Tracy, A., and McClelland, M. M. (2021). The Head-Toes-Knees-Shoulders Revised (HTKS-R): development and psychometric properties of a revision to reduce floor effects. *Early Childh. Res. Quart.* 56, 320–332. doi: 10.1016/j.ecresq.2021.03.008
- Gregory, A., Cornell, D., and Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *Am. Educ. Res. J.* 48, 904–934. doi: 10.3102/0002831211398531
- Gresham, F. M., Reschly, D. J., and Carey, M. P. (1987). Teachers as "tests": classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychol. Rev.* 16, 543–553.



- Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., and Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics* 145, e20191579. doi: 10.1542/peds.2019-1579
- Hartmann, D. P., and Pelzel, K. E. (2015). "Design, measurement, and analysis in developmental research," in *Developmental Science: An Advanced Textbook*, Vol. 7th, ed. M. H. B. M. E. Lamb (New York, NY: Psychology Press).
- Hernández-Ramos, P. (2005). If not here, where? Understanding teachers' use of technology in Silicon Valley schools. *J. Res. Technol. Educ.* 38, 39–64.
- Hirsh-Pasek, K., Kochanoff, A., Newcombe, N. S., and de Villiers, J. (2005). Using scientific knowledge to inform preschool assessment: making the case for "empirical validity". *Soc. Pol. Rep.* XIX(1), 3–19. doi: 10.1002/j.2379-3988.2005.tb00042.x
- IBM Corp (2021). *IBM SPSS Statistics for Windows, Version 27.0*. Armonk, NY: IBM Corp.
- Isaacs, J., Sandstrom, H., Rohacek, M., Lowenstein, C., Healy, O., and Gearing, M. (2015). *How Head Start Grantees Set and Use School Readiness Goals*. In *OPRE Report #2015-12a*. Washington, DC: The Urban Institute.
- Ispa-Landa, S. (2018). Persistently harsh punishments amid efforts to reform: using tools from social psychology to counteract racial bias in school disciplinary decisions. *Educ. Research.* 47, 384–390. doi: 10.3102/0013189x18779578
- Jones, M., and Liu, M. (1997). Introducing interactive multimedia to young children: a case study of how two-year-olds interact with the technology. *J. Comput. Childh. Educ.* 8, 313–343.
- Kenny, S. A., Cameron, C. E., Karing, J. T., Ahmadi, A., Braithwaite, P. N., and McClelland, M. M. (2023). A meta-analysis of the validity of the Head-Toes-Knees-Shoulders task in predicting young children's academic performance [Systematic Review]. *Front. Psychol.* 14:1124235. doi: 10.3389/fpsyg.2023.1124235
- Kim, K. (2016). Teaching to the data collection? (Un)intended consequences of online child assessment system, "Teaching Strategies GOLD". *Glob. Stud. Childh.* 6, 98–112. doi: 10.1177/2043610615627925
- Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., and Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: a comparative evaluation of the educational relevance of selected measures. *J. Educ. Psychol.* 2017, 203. doi: 10.1037/edu0000203
- Martin, R. (2008). *New Possibilities and Challenges for Assessment Through the Use of Technology. Towards a Research Agenda on Computer-Based Assessment. Challenges and needs for European Educational Measurement* (Luxembourg), 6–9.
- Mashburn, A. J., Hamre, B. K., Downer, J. T., and Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *J. Psychoeducat. Assess.* 24, 367–380. doi: 10.1177/0734282906290594
- Matthews, J. S., Cameron Ponitz, C., and Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *J. Educ. Psychol.* 101, 689–704. doi: 10.1037/a0014240
- Maves, S. (2022). *Using Assessments to Create More Equitable Early Learning Systems: What Do Pre-K Leaders Think?* Available online at: <https://www.mdrc.org/publication/using-assessments-create-more-equitable-early-learning-systems-what-do-pre-k-leaders>
- McCabe, L. A., Rebello-Britto, P., Hernandez, M., and Brooks-Gunn, J. (2004). "Games children play: observing young children's self-regulation across laboratory, home, and school settings," in *Handbook of Infant, Toddler, and Preschool Mental Health Assessment*, eds. R. DelCarmen-Wiggins and A. Carter (Oxford: Oxford University Press), 491–521.
- McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., and Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childh. Res. Quart.* 28, 314–324. doi: 10.1016/j.ecresq.2012.07.008
- McClelland, M. M., Ahmadi, A., and Wanless, S. B. (2022). *Self-regulation. In Reference Module in Neuroscience and Biobehavioral Psychology*. Amsterdam: Elsevier.
- McClelland, M. M., and Cameron, C. E. (2012). Self-regulation in early childhood: improving conceptual clarity and developing ecologically valid measures. *Child Dev. Perspect.* 6, 136–142. doi: 10.1111/j.1750-8606.2011.00191.x
- McClelland, M. M., and Cameron, C. E. (2018). Developing together: the role of executive function and motor skills in children's early academic lives. *Early Childh. Res. Quart.* 3, 14. doi: 10.1016/j.ecresq.2018.03.014
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: the Head-Toes-Knees-Shoulders task. *Front. Psychol.* 5:599. doi: 10.3389/fpsyg.2014.00599
- McClelland, M. M., Díaz, G., Lewis, K., Cameron, C. E., and Bowles, R. P. (2015a). "The Head-Toes-Knees-Shoulders task as a measure of school readiness, in *Poster Presented at US Department of Education, Institute for Education Sciences (IES) December Grantee Meeting*. Washington, DC.
- McClelland, M. M., Geldhof, G. J., Cameron, C. E., and Wanless, S. B. (2015b). "Development and self-regulation," in *Handbook of Child Psychology and Developmental Science, 7th Edn, Vol. 1: Theory and Method*, eds. W. F. Overton and P. C. M. Molenaar (Hoboken, NJ: Wiley), 523–566.
- McClelland, M. M., Gonzales, C. R., Cameron, C. E., Geldhof, G. J., Bowles, R. P., Nancarrow, A. F., et al. (2021). The Head-Toes-Knees-Shoulders revised: links to academic outcomes and measures of EF in young children. *Front. Psychol.* 12:721846. doi: 10.3389/fpsyg.2021.721846
- Miller-Cotto, D., Smith, L. V., Wang, A. H., and Ribner, A. D. (2022). Changing the conversation: a culturally responsive perspective on executive functions, minoritized children and their families. *Infant Child Dev.* 31, e2286. doi: 10.1002/icd.2286
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., and Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 1010076108. doi: 10.1073/pnas.1010076108
- Muthén, B. O., and Muthén, L. K. (1998). *Mplus Version 8.0*. Los Angeles, CA: Muthén & Muthén.
- National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education (2003). *Early Childhood Curriculum, Assessment and Program Evaluation*. Washington, DC: NAEYC.
- National Research Council (2008). *Early Childhood Assessment: Why, What, and How*. Washington, DC: The National Academies Press.
- Neuman, M. J., and Devercelli, A. E. (2013). "What matters most for early childhood development: a framework paper," in *Systems Approach for Better Education Results (SABER) Working Paper Series No 5*. Washington, DC: World Bank. Available online at: <https://openknowledge.worldbank.org/handle/10986/20174>
- Olson, L., and Lepage, B. (2022). *Tough Test: The Nation's Troubled Early Learning Assessment Landscape*. G. University. Available online at: [https://www.future-ed.org/wp-content/uploads/2022/02/REPORT\\_Tough\\_Test.pdf](https://www.future-ed.org/wp-content/uploads/2022/02/REPORT_Tough_Test.pdf)
- Outwaite, L. A., Gulliford, A., and Pitchford, N. J. (2017). Closing the gap: efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Comput. Educ.* 108, 43–58. doi: 10.1016/j.compedu.2017.01.011
- Papageorge, N., Gershenson, S., and Kang, K. (2016). *Teacher expectations matter (IZA Discussion Paper No. 10165)*. Bonn: Institute for the Study of Labor (IZA).
- Reardon, S. F., Kalogrides, D., and Shores, K. (2017). *The Geography of Racial/Ethnic Test Score Gaps. CEPA Working Paper No. 16-10*. Stanford, CA: Stanford Center for Education Policy Analysis.
- Rideout, V. J., and Katz, V. S. (2016). *Opportunity for All? Technology and Learning in Lower-Income Families: A Report of the Families and Media Project*. Available online at: <https://files.eric.ed.gov/fulltext/ED574416.pdf>
- Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., and Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Dev. Psychol.* 45, 958–972. doi: 10.1037/a0015861
- Rothbart, M. K., Posner, M. I., and Kieras, J. (2006). "Temperament, attention, and the development of self-regulation," in *Blackwell Handbook of Early Childhood Development*, eds. K. McCartney and D. Phillips (Blackwell Publishing), 338–357. Available online at: <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2006-04286-017&site=ehost-live>
- Salthouse, T. A., Atkinson, T. M., and Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J. Exp. Psychol.* 132, 566. doi: 10.1037/0096-3445.132.4.566
- Schilder, D., and Carolan, M. (2014). *State of the States Policy Snapshot: State Early Childhood Assessment Policies*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.
- Schmitt, S. A., Pratt, M. E., and McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Educ. Dev.* 25, 641–660. doi: 10.1080/10409289.2014.850397
- Smith, T. E. C., Patton, J. R., Polloway, E. A., Dowdy, C. A., and McIntyre, L. J. (2015). *Teaching Students With Special Needs in Inclusive Settings*. London: Pearson.
- Spawls, N., and Wilson, D. (2017). *Why Are Young Children So Good With Technology? Young Children in a Digital Age: Supporting Learning and Development With Technology in the Early Years* ed L. Kaye (London; New York, NY: Taylor & Francis), 18–29.
- Teddle, C., and Tashakkori, A. (2009). *Foundations of Mixed Methods Research*. Brooklyn, NY: Sage.
- Tripathy, R., Burke, R., and Kwan, T. (2020). *Case Study: Implementing Cognitive ToyBox Alongside a Traditional Observation-Only Approach to Assessment in the Early Childhood Classroom*. San Francisco, CA: WestEd.
- Tymms, P. (2001). The development of a computer-adaptive assessment in the early years. *Educ. Child Psychol.* 18, 20–30.

- Ursache, A., Noble, K. G., the Pediatric Imaging, N., and Study, G. (2016). Socioeconomic status, white matter, and executive function in children. *Brain Behav.* 6, e00531. doi: 10.1002/brb3.531
- Vandewater, E. A., Rideout, V. J., Wartella, E. A., Huang, X., Lee, J. H., and Shim, M. S. (2007). Digital childhood: electronic media and technology use among infants, toddlers, and preschoolers. *Pediatrics* 119, e1006–e1015. doi: 10.1542/peds.2006-1804
- Vitiello, V. E., Greenfield, D. B., Munis, P., and George, J. (2011). Cognitive flexibility, approaches to learning, and academic school readiness in Head Start preschool children. *Early Educ. Dev.* 22, 388–410. doi: 10.1080/10409289.2011.538366
- Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S. H., Lan, X., et al. (2011a). Measuring behavioral regulation in four societies. *Psychol. Assess.* 23, 364. doi: 10.1037/a0021768
- Wanless, S. B., McClelland, M. M., Lan, X., Son, S. H., Cameron, C. E., Morrison, F. J., et al. (2013). Gender differences in behavioral regulation in four societies: the United States, Taiwan, South Korea, and China. *Early Childh. Res. Quart.* 28, 621–633. doi: 10.1016/j.ecresq.2013.04.002
- Wanless, S. B., McClelland, M. M., Tominey, S. L., and Acock, A. C. (2011b). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Educ. Dev.* 22, 461–488. doi: 10.1080/10409289.2011.536132
- Waterman, C., McDermott, P. A., Fantuzzo, J. W., and Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—or whose score is it anyway? *Early Childh. Res. Quart.* 27, 46–54. doi: 10.1016/j.ecresq.2011.06.003
- Waters, N. E., Ahmed, S. F., Tang, S., Morrison, F. J., and Davis-Kean, P. E. (2021). Pathways from socioeconomic status to early academic achievement: the role of specific executive functions. *Early Childh. Res. Quart.* 54, 321–331. doi: 10.1016/j.ecresq.2020.09.008
- Wiley, A. L., Brigham, F. J., Kauffman, J. M., and Bogan, J. E. (2013). Disproportionate poverty, conservatism, and the disproportionate identification of minority students with emotional and behavioral disorders. *Educ. Treat. Child.* 36, 29–50. doi: 10.1353/etc.2013.0033
- Willoughby, M. T., Blair, C. B., Wirth, R. J., and Greenberg, M. (2012). The measurement of executive function at age 5: psychometric properties and relationship to academic achievement. *Psychol. Assess.* 24, 226–239. doi: 10.1037/a0025361
- Xie, H., Peng, J., Qin, M., Huang, X., Tian, F., and Zhou, Z. (2018). Can touchscreen devices be used to facilitate young children's learning? A meta-analysis of touchscreen learning effect. *Front. Psychol.* 9, 2580. doi: 10.3389/fpsyg.2018.02580
- Zelazo, P. D. (2020). Executive function and psychopathology: a neurodevelopmental perspective. *Ann. Rev. Clin. Psychol.* 16, 431–454. doi: 10.1146/annurev-clinpsy-072319-024242
- Zelazo, P. D., Blair, C., and Willoughby, M. T. (2016). *Executive Function: Implications for Education (NCER 2017-2000)*. Washington, DC: National Center for Education Research: Institute of Education Sciences, U.S. Department of Education. Available online at: <http://ies.ed.gov/>