Evaluating the viability of a structured observational approach to assessing early self-regulation

Steven James Howard, [a,b] Cathrine Neilsen-Hewett, [a] Marc de Rosnay, [b] Elena Vasseleu, & Edward Melhuish[c]

[a]School of Education, University of Wollongong, Wollongong, NSW, Australia, 2522

[b]School of Psychology, University of Wollongong, Wollongong, NSW, Australia, 2522

[c]Department of Education, University of Oxford, UK

Corresponding Author: Dr. Steven Howard, Early Start, University of Wollongong, NSW, Australia, 2522. Telephone: +61 (0)2 4221 5165. E-mail: stevenh@uow.edu.au

Page Count: 40 pages (abstract + text + references + tables)

Declarations of Interest: None.

Funding Source: This work was supported by philanthropic financial support from the Colman Foundation. This funding body had no role in the design, conduct, or reporting of this research.

Evaluating the viability of a structured observational approach to assessing early self-regulation

Abstract

There is now clear evidence that young children's abilities to self-regulate set in motion developmental trajectories that are longitudinally influential, that self-regulation is malleable in childhood with flow-on implications into adulthood, and that efforts to improve self-regulation are enhanced when tailored and adaptive to children's developmental progress. Yet there are significant concerns regarding current approaches to assessing self-regulation, potentially limiting our capacity to respond to children's needs and support their development. In response, this study examined the viability of a novel approach to assessing young children's cognitive, behavioral, and socio-emotional self-regulation that uses observer ratings in semi-structured activities. Eighty 3- to 5-year-old children were observed in their preschool in a small group task (a memory card game with other children) and in two individual tasks (an outdoor obstacle course and a curiosity box guessing game). Exploratory factor analysis of resultant *Preschool Situational Self-Regulation Toolkit* (PRSIST) *assessment* ratings yielded a two-factor solution which corresponded to *cognitive* and *behavioral* self-regulation. Comparison of these observation ratings with task-based (Head-Toes-Knees-Shoulders; HTKS) and adult-report measures of self-regulation (parent and teacher report on the Child Behavior Rating Scale; CBRS) showed strong association with HTKS scores and more modest relations with CBRS ratings. Unlike adult report, however, PRSIST and HTKS were also developmentally sensitive, both showed robust associations with children's school readiness, and neither revealed systematic differences across genders. These findings offer encouraging support for the PRSIST approach, which has high ecological validity and direct relevance to those working with children.

Keywords: self-regulation, early years, preschool, assessment, formative, observation

Introduction

By the end of the preschool years a well self-regulated child can sustain their attention and resist distraction, endure temptations and delay gratification, wait their turn, consider the consequences of their actions, and persist with challenging activities. They are often able to initiate or cease behaviors that conflict with their immediate preferences or impulses, such as listening to the views of other children in group time despite a desire to focus on their own interests. Self-regulation can broadly be considered as the ability to control our thinking, behavior, social interactions, and emotional reactions, despite contrary urges and impulses (occasionally referred to as self-control; Moffitt et al., 2011). While the segmentation of self-regulation (i.e., as some permutation of cognitive, behavioral, and social-emotional dimensions) and its delineation from related abilities remains debated (e.g., Rademacher & Koglin, 2018), one prominent proposal suggests that successful self-regulation in complex real-world contexts involves: *goal setting*, to behave in a self-regulated way; *motivation*, to remain self-regulated when this becomes difficult; and *capacity* to sustain attention toward goals, while resisting contrary impulses and distractions (with this cognitive control enabled by executive functioning; Hoffman, Schmeichel, & Baddeley, 2012).

A myriad of studies underscore the foundational role of self-regulation for influencing developmental trajectories related to: school readiness and academic success (e.g., Ponitz, McClelland, Matthews, & Morrison, 2009); risky lifestyle choices in adolescence (such as substance use, self-harm; Howard & Williams, 2018); and health, wealth, substance abuse, and criminal offending in adulthood (Moffitt et al., 2011). Good self-regulation thus appears to enable important kinds of freedom; freedom from needing constant direction from others, from maladaptive and inappropriate impulses, and from unnecessary distraction. There is also mounting evidence that *early* self-regulation, in particular, is important for establishing initial developmental trajectories that are widely and longitudinally impactful (Howard & Williams,

2018; Moffitt et al., 2011). Rather than these trajectories being fixed early, however, there is compelling evidence for the malleability of self-regulation in childhood, with flow-on effects for diverse outcomes in adulthood (Moffitt et al., 2011). Indeed, it has been suggested that early intervention may produce more pronounced, stable, and lasting positive change across the lifespan (e.g., Wass, Scerif, & Johnson, 2012), and provide better return on investment (Heckman, 2006), than comparable efforts later in life.

Uncertainty remains, however, as to how such changes are best effected and supported. In relation to the cognitive control mechanisms underlying self-regulation (i.e., executive functions; Hofmann et al., 2012), current findings indicate that children benefit more from adaptive training tailored to their current and changing ability, than from less-differentiated and non-adaptive intervention approaches (Chacko et al., 2014; Dunning, Holmes, & Gathercole, 2013; Green et al., 2012). For instance, Dunning et al. (2013) found that increasing the demand on working memory in line with children's expanding capacities led to greater improvements than training with a static working memory demand. Indeed, this idea permeates contemporary educational curricula and theory, and has a strong evidence base across diverse populations and domains (Kremer, Brannen, & Glennerster, 2013). In light of these findings, efforts to foster self-regulation may be improved by a clear understanding of individuals' developmental progress and ongoing change, in ecologically valid situations and across the range of domains to which self-regulation might apply (i.e., cognitive, behavioral, socio-emotional).

Contemporary Approaches to Early Self-Regulation Assessment

There are two predominant approaches for appraising self-regulation in childhood: adult report and task-based. Adult report relies on adults who are familiar with the child – typically a parent or educator – and asks them to rate the frequency, severity, or typicality of a child's everyday self-regulatory behavior (e.g., *persists in difficult tasks, concentrates when working*

on a task; Bronson et al., 1990; Howard & Melhuish, 2017). This approach has the benefit of targeting real-world self-regulatory behaviors that are ecologically valid and predict later-life outcomes (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004). Adult report also has the benefit of generating extensive data at minimal time and cost. Yet the extent to which these ratings accurately reflect children's current self-regulatory abilities remains unclear. For instance, despite the fact that self-regulation develops rapidly across the preschool years – as is clear when comparing the behavior of a child at 3- and 5-years of age – adult-report ratings often portray little developmental change (e.g., Howard & Melhuish, 2017). This discrepancy may be attributable to the subjective nature of adult-reported ratings, such that respondents likely reference a child to their same age-peers (e.g., 'this child persists well in difficult tasks *for a 3-year old*'). As a consequence, the data generated by this approach are more reflective of a child's self-regulatory ranking within their context and age group, which is more stable but is less useful for supporting change or informing interventions than an objective index of development.

Furthermore, the typically modest associations between parent- and educator-reports of children's self-regulatory capacities – despite being on the same child and using the same scale – across a range of child development dimensions further highlights the subjectivity or situational inconsistency inherent in this approach (Goodman, 2001; Murray, Ruble, Willis, & Molloy, 2009). Indeed, these inherent methodological problems may explain findings that adult-report measures show different patterns of association with other child outcomes – and uniquely indicate gender differences, such that girls are usually rated as better self-regulators than boys (Matthews, Ponitz, & Morrison, 2009) – than more direct self-regulation measures (Schmitt, Pratt, & McClelland, 2014).

An alternative tradition in assessing early self-regulation is direct task-based approaches, in which children are asked to undertake an activity that is believed to require self-regulation

to successfully complete. Amongst the most highly researched of these are the Marshmallow and Head-Toes-Knees-Shoulders (HTKS) tasks. In the Marshmallow task (Mischel, Shoda, & Rodriguez, 1989), children are presented with a marshmallow and are informed that they will receive a second one if they can refrain from eating the initial marshmallow for a period of time. In HTKS (Ponitz et al., 2009), a child is taught a correspondence between two or more body parts (e.g., head-knees) and then must touch the body part that is opposite to what they are directed to touch (e.g., touch their head if told to *touch your knees*). In both cases, these tasks require children to overcome natural impulses (e.g., to eat a sweet or to carry out an instruction as given) and instead behave in accordance with the rules of the situation.

Strengths of task-based approaches lie in their structured situation and direct assessment of children's abilities, which contribute to their strong psychometric properties (McClelland et al., 2014), sensitivity to age- and intervention-related changes (Tominey & McClelland, 2011), and association with academic and non-academic outcomes (e.g., Ponitz et al., 2009; Schlam, Wilson, Shoda, Mischel, & Ayduk, 2013). Yet task-based self-regulation measures diverge in the extent to which they approximate the complex and often affective real-world situations in which children typically must self-regulate. While the Marshmallow Task does require children to resist a tangible and affective impulse that is routinely experienced in life (i.e., to wait before eating), HTKS assesses children's ability to follow instructions in a novel task without this affective component. As such, the extent to which performance across these tasks genuinely index children's everyday self-regulatory abilities (e.g., ability to share, take turns, resist emotional outbursts) may vary.

Indeed, rather than a measure of self-regulation, it may be that tasks like HTKS measure *executive function* through their requirements, for example, to maintain rules in mind (i.e., in working memory) and perform opposite actions (i.e., via inhibition). This construal of HTKS (and other similar paradigms) as indexing cognitive control aspects of self-regulation, and not

self-regulation per se (which, at a minimum, additionally involves goal-setting, motivation, and problem solving; Hoffman et al., 2012), is supported by findings that executive function tasks show similar patterns of results to task-based 'self-regulation' measures (Fitzpatrick, McKinnon, Blair, & Willoughby, 2014; Miller, Nevado-Montenegro, & Hinshaw, 2012).

An Alternative Approach to Early Self-Regulation Assessment

An alternative and emerging approach to self-regulation assessment in childhood uses observation in ecologically valid situations to combine some of the strengths of task-based (i.e., greater objectivity, developmental sensitivity) and adult-report approaches (i.e., a focus on authentic self-regulatory behaviors, capturing distinct dimensions and applications of self-regulation), while minimizing their limitations. The importance of ecological validity – that is, whether measures, methods, and materials approximate the real-world conditions of the factor(s) under investigation – is illustrated by research suggesting that task-based executive function measures account for only one-fifth of the variance in everyday executive ability (Chaytor, Schmitter-Edgecombe, & Burr, 2006). While measures such as HTKS similarly strive to approximate and leverage familiar situations (e.g., 'Simon Says'), their individual, highly structured, and less-affective nature means that they miss much of the complexity, interactivity, and emotionality that is prevalent in the daily situations in which children are expected to self-regulate.

By contrast, observational tools seek to emphasize both psychometric and ecological validity in generating profiles of self-regulation across its various areas of application (e.g., cognitive, behavioral, social-emotional). For example, the Response to Challenge (RCS) scale evaluates self-regulation as a child completes routine physical activities in an obstacle course (Lakes, 2012). Observers rate a child's attentiveness, assertiveness, and perseverance. In another paradigm, the Preschool Self-Regulation Rating Assessment (PSRA) engages children in a set of executive-function-like tasks (e.g., do not peek as a present is wrapped)

but rates their attention, impulse control, and regulation of arousal, instead of accuracy or response times typical of executive function task indices (Smith-Donald, Raver, Hayes, & Richardson, 2007). The Observed Child Engagement Scale (OCES; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2005) observes similar dimensions within play or learning situations (e.g., engagement, compliance).

While these observation-based measures are responding to important issues of validity and developmental sensitivity, they are not without issue. For example, to generate a reliable self-regulatory index on the RCS, multiple live observers are required for each child (Lakes, 2012). In the PSRA, use of executive-function-like activities raises similar questions as some task-based approaches regarding children's emotional and motivational investment in these non-typical situations. For the OCES, when unstructured free play is observed, children's self-selection into different types, durations, and complexities of free play makes it difficult to compare ratings across children; this may explain why OCES ratings did not relate to any child outcomes investigated in a comparison of adult-report, task-based, and observational approaches (Schmitt et al., 2014). Notwithstanding these challenges, the potential benefits driving these approaches are worth pursuing. If developmentally sensitive and valid indices of children's early and everyday self-regulation can be captured through direct and structured observation of their real-world activities, our capacity to respond to children's needs and support their development would be enhanced.

Current Study

To explore the viability of a structured observational assessment of children's early self-regulation development, and in response to issues identified above, the Preschool Situational Self-Regulation Toolkit (PRSIST) assessment was created. First, to ensure consistency across the situations in which children would be observed, semi-structured activities were selected on the basis of offering sufficiently high self-regulatory challenge (e.g., children needing to

persist in a challenging task, overcome frustration, stay within the rules of the activity) to evoke stable individual differences in children's self-regulatory development. Further, to preserve ecological validity, the activities also had to be compatible with early childhood contexts and routines, reflect the diversity of experiences children have in these settings (i.e., group, individual), and be similar to common activities that young children already engage in. Based on these considerations, the first activity selected was a group-based memory card game. It was anticipated that this activity would introduce a high degree of challenge across the cognitive (e.g., sustain attention, resist distraction), behavioral (e.g., take turns, follow the rules), and social-emotional domains (e.g., cope with frustration) within which self-regulation applies, as a consequence of its semi-structured and group format, opportunities for impulsive behaviors, and need for sustained attention.

Whilst the group-based activity has obvious advantages – such as its approximation of the dynamic, complex, and emotionally laden contexts in which children must routinely self-regulate – it stands in contrast with the overwhelming precedent in the literature to examine children individually. Further, given the possibility that self-regulatory capacity might differ across group and individual contexts, it was also decided to select an individual activity. An obstacle course activity was included on the basis of its anticipated self-regulatory challenge (e.g., persist if initially unsuccessful, follow instructions and sequences). However, given that difficulties emerged in obtaining reliable self-regulation ratings from children's obstacle course performance using the RCS (which may be influenced by conflation of physical and self-regulation items in the RCS, rather than the activity per se; Lakes, 2012; Lakes & Hoyt, 2004), a briefer curiosity box guessing game was also included. It was intended that one individual activity would be selected from these two options, if necessary, to supplement the group activity.

An associated scale was then created based on existing literature and the observation of children during the activities, to capture dimensions of cognitive (e.g., attention, distraction), behavioral (e.g., self-directedness, behavioral control), and social-emotional self-regulation (e.g., emotional reactions and recovery, helping). This mirrors the approach of adult-report measures, and contrasts with the single-index result of other task-based measures, in that it provides a profile of development across a range of self-regulatory dimensions. Yet to ensure comparable sensitivity to age- and intervention-related change as task-based approaches, an iterative process of developing and refining descriptors and exemplars of each behavior was also conducted so that self-regulatory behaviors could be clearly and consistently situated on a developmental continuum. Specifically, after initial item creation, multiple rounds of rating occurred with seven raters viewing video recordings of children participating in the activities, and then independently rating using a working draft of the scale. Subsequent discussions of ratings, interpretations, and rationales contributed to refining the PRSIST items, descriptions, and developmental continua embedded in the scale. Finally, to ensure maximum objectivity and reliability across raters, a structured process for training and evaluation was created. This potentiates future use of the tool not only by researchers, but also by those who spend the majority of time with young children (e.g., educators). In this respect, the PRSIST assessment contrasts sharply with the vast majority of existing assessments, which are explicitly reserved for trained researchers (e.g., PSRA; Smith-Donald et al., 2007) or yield results that have less-clear application to educational practice (i.e., do not provide explicit guidance on the specific areas requiring support, which are needed to act upon the data; Ponitz et al., 2009).

The current study thus evaluated the viability of a structured observational approach to self-regulation assessment. Specifically, we aimed to: (1) determine whether an observation approach could yield insights about discrete yet correlated cognitive, behavioral, and social-emotional dimensions of self-regulation; (2) evaluate psychometric properties of the PRSIST

assessment; (3) examine the level of agreement between observer ratings of self-regulation with task-based and adult-report assessments; and (4) examine age trends across each of the measurement approaches. It was expected that the PRSIST ratings would yield three related yet distinct self-regulation factors with good psychometric properties. It was also expected that, despite their clear differences, the PRSIST assessment would more strongly relate to an existing task-based measure (i.e., HTKS) than an educator- or parent-report measure. This was expected due to: our additional expectation that the PRSIST assessment would detect age-related changes in self-regulatory development (like task-based measures, but unlike adult-report measures); and that PRSIST was seen as an intermediary between task-based (i.e., increasing the objectivity of adult reports) and observational approaches (i.e., using a questionnaire-like scale of children's everyday behaviors to guide ratings). In addition to these expectations, we also: included an index of children's school readiness as an external validation of the various self-regulatory measures in the study; and explored possible gender differences in children's self-regulation, which have been found more commonly in adult-report measures (Matthews et al., 2009).

Material and Methods

Participants

Participants were 80 3- to 5-year-old children (M_{age} = 4.46, *SD* = 0.73; range = 3.10 - 5.88) who were enrolled at one of seven preschool services in or proximal to a metropolitan city in Australia. More than 80% of 3- to 5-year old children in Australia are enrolled in pre-school education (which is in line with the OECD average; OECD, 2018). Preschools ranged in their statutory government quality ratings against the National Quality Standard (ranging from Meeting to Exceeding this standard). Children who attended these centres resided in middle- to high-SES areas (i.e., deciles 6-8), based on the Australian Bureau of Statistics' Socioeconomic Indexes for Areas (SEIFA; ABS, 2008). The sample consisted of a similar

number of 3-year-olds (*n* = 28, 60% girls), 4-year-olds (*n* = 31, 42% girls), and 5-year-olds (*n* = 21, 43% girls). All children spoke English as a first language and had no formal diagnoses of developmental delay.

Measures

Preschool Situational Self-Regulation Toolkit (PRSIST) assessment. The PRSIST assessment engages children in a series of activities that were selected on the basis of their expected self-regulatory challenge and consistency with activities children routinely engage in within preschool contexts. Specifically, children participated in one group (memory card game) and two individual activities (curiosity boxes guessing game, obstacle course) in their preschool centre. Activities were facilitated by a trained observer, who on completion of each activity rated the participating child(ren) on the PRSIST scale. The PRSIST scale, training in its use, and activity protocols are outlined below.

PRSIST scale. Observations and ratings of children's self-regulation within each activity were structured by the 11-item PRSIST scale. The scale requires observers to rate a child's behavior and responses to each activity along a 7-point Likert scale, representing frequency and/or severity of that behaviour. Items pertain to dimensions of cognitive (e.g., Did the child sustain attention, and resist distraction, throughout the instructions and activity?), behavioral (e.g., Did the child control their behaviors and stay within the rules of the activity?), and social-emotional self-regulation (e.g., Did the child resist emotional impulses, and not show overly positive or negative emotional reactions?). For each item a description of its focus is provided, as well as examples of how relevant behaviors can manifest. Further, clusters of behaviors that would yield a rating of 1 and 7 are described as developmental anchors. In all cases, items focused not on the number of pairs in the memory game, correct guesses in the curiosity boxes, or physical ability within the obstacle course, but instead on self-regulatory behaviors around: sustaining attention; engagement; self-direction; impulsiveness; remaining

within the rules of the activity; fidgeting; following social conventions such as taking turns; willingness to try even under conditions of uncertainty; as well as emotional responses and recovery.

While the PRSIST items were rated in the same manner for each activity, there were two items that were predominantly relevant to, and thus only rated for, a particular activity: *taking opportunities to be helpful*, in the group memory game; and *willingness to risk being wrong under conditions of uncertainty*, in the guessing game. There was also one item on recovery from emotional responses that could be scored N/A if a frustration response did not occur. A full version of the PRSIST scale can be found at Appendix A.

PRSIST training. To ensure reliability of these ratings, prior to data collection observers must complete an in-depth online training module that provides background, administration protocols, practice ratings with feedback, and culminates in an inter-rater reliability check against a full observation and rating. This reliability check requires a minimum threshold of consistency against a benchmark rating as follows: mean difference in average rating ≤ 0.75 points; a correlation between item ratings of at least $r = .70$; and at least 80% of item ratings within 1 point. The PRSIST online training module is freely available at www.prsist.com.au. We additionally video-recorded PRSIST administration for further in-field reliability checks. For these checks we randomly selected ~10% of participants ($n = 9$) to be independently rated by one of the scale's creators, who was blind to the in-field ratings. ICCs considering single measures and the absolute agreement between ratings (to estimate the consistency of a single rater) were acceptable to excellent for all activities and subscales (ICCs ranging from .72 to .92; Koo & Li, 2016).

PRSIST Memory Game. The first PRSIST assessment activity was a group memory card game. In this activity, a group of four children who were familiar with each other were seated around a table with the activity facilitator (observer) and were introduced to the memory card

game. Instructions focused on relaying and demonstrating important elements of the activity: what a pair is; that two cards must be flipped over to try to find a pair; if a pair is found you get to keep it and take another turn; or if not, it is the next person's turn. The number of pairs increased with the children's age (6 pairs for 3-year-olds, 8 pairs for 4-year-olds, 14 pairs for 5-year-olds) to ensure gameplay took around 10 minutes. During the activity, the facilitator intervened with support only as needed, ensuring sufficient pauses to first see what children would do. Giving children an opportunity to engage in the activity independently, intervening only when needed (e.g., after a delay in which a child hadn't taken their turn, or if a child took their turn prematurely, confirming whose turn it is), permitted evaluation of a child's self-direction and whether they acceded to impulses that ran contrary to rules of the activity. The facilitator made brief and occasional notes throughout the activity as a memory prompt, and then completed the rating scale for each child at the conclusion of the activity. While we did not specifically assess numeracy abilities, and supported children's counting as needed, numeracy cards were used in this study (digits with matching quantities of depicted animals, such as two dolphins) to illustrate how the PRSIST assessment could be embedded within learning experiences, rather than require 'assessment-specific' time or resources in preschool settings. As such, during the 'assessment time' children experienced interactive game play, received practice in numeracy skills, and did so in their usual classroom setting. As such, this was simultaneously play, a learning experience with curricular links, and an assessment that could provide insight about children's self-regulation.

With respect to group assignment in the memory game, a priori we needed groups of four children who were of the same age, familiar with each other, and were in a familiar environment. The only other constraint that we imposed on the group context was that children should not be best friends nor should they be uncomfortable with one another. Within those constraints, assignment was opportunistic based on child age and availability.

This activity took ~10 minutes to complete (for an effective duration of 2.5 minutes per child). Full activity protocols are provided at http://www.prsist.com.au/assessment.html.

PRSIST Obstacle Course. The second activity was an obstacle course undertaken with individual children. In this activity a child attempted a sequence of increasingly challenging physical activities, one at a time. The obstacle course progressed as follows: 'balance' along a line of tape on the ground; two-foot hop through the squares of a 2m ground ladder; throw a bean bag between two pylons at a distance of 1m; two-foot hop in and out of the squares of a ground ladder; and throw a bean bag into squares of a ground ladder at increasing distance. Children continued until completion or, more commonly, after 10 minutes. Each physical activity was described and demonstrated to the child, after which the child attempted that activity themselves. Where performance was initially unsuccessful, a child was told that they needed to start again and were prompted with 'Let's keep trying until we get it right, or you can let me know when you've had enough and want to stop.' If children indicated that they wanted to stop, the next activity in the sequence was introduced and demonstrated for the child to attempt. As with the first activity, the facilitator completed the ratings for a child after the child's completion of the full obstacle course (or 10 minutes had elapsed).

PRSIST Curiosity Boxes. The third activity was a curiosity boxes guessing activity done individually with children. In this activity, children were presented with a sequence of three boxes of increasing size, one at a time, and were asked to guess their contents. The facilitator initially explained the sequence of guessing to the child, as follows: first, guess based only on the size of the box (no touching); second, guess after gently lifting the box to feel its weight (no shaking); third, guess after shaking the box (no opening); and lastly, guess after closing your eyes and feeling the object inside (no peeking). Children were additionally told that a gift for them was contained in the final box, as a thank you for them playing the games. As with the other two activities, after explaining the structure and sequence of the activity the

facilitator gave sufficient pause to permit the child to engage with the activity independently, intervening only as necessary (e.g., if a child skipped a step, reminding them of the rules; if a child did not guess, asking what they think might be in the box). In line with disappointing gift protocols that are well established (Davis, 1995), after receiving the disappointing gift in the final box (i.e., a broken CD/DVD) a series of probes were used to evaluate the child's reaction: 'Do you like it?'; 'Why do you like it?'; 'What do you think you will do with it?' After responses to these questions, or if a child appeared upset, the facilitator indicated they had made a mistake and put the wrong gift in the box. They then corrected this by giving the child a picture book as their actual gift. This activity took about 5 minutes to complete. The facilitator rated the child's behaviors on the PRSIST scale at the conclusion of the activity.

Head-Toes-Knees-Shoulders (HTKS). HTKS is often considered as a task-based measure of child self-regulation (Ponitz et al., 2009) and has been shown to have strong psychometric properties (inter-rater reliability, $\kappa = 0.90$; internal consistency, $\alpha > .90$) and good concurrent validity (McClelland et al., 2014). This task asks children to remember a correspondence between body parts (e.g., head and knees), perform an opposite action to what is indicated by the administrator (e.g., touch their knees when the facilitator says *touch your head*) and, at higher levels of challenge, flexibly switch between correspondences. The task consists of six practice trials and 10 test trials at each of three levels of difficulty: (1) correspondence between head and toes; (2) correspondence between knees and shoulders; and (3) switching between correspondences of head-knees and shoulders-toes. The task continues until either completion (~8 minutes) or failing to achieve four points at any level (2 points are awarded for a correct response and 1 point for a self-corrected correct response). Assessors completed the online training module prior to in-field data collection to ensure accuracy of scoring and inter-rater reliability. Self-regulation was indexed by a sum score of the correct and self-corrected items across all practice and test trials.

Child Behavior Rating Scale (CBRS). CBRS is an adult-report measure of children's task and social behavior (Bronson et al., 1990), from which a reduced 10-item scale of task self-regulation has been derived (Matthews et al., 2009). Items ask adults to rate the frequency of target behaviors (e.g., attempts new challenging tasks, concentrates when working on a task) on a 5-point scale ranging from 1 (Never) to 5 (Always). In the current study, both parent- and educator-report data were collected, to be considered separately. Self-regulation was indexed by a score that averages all 10 items of the reduced CBRS.

Bracken School Readiness (BSRA). BSRA (3rd edition; Bracken, 2007) is a standardized assessment of key areas deemed important for school readiness. This includes subscales for colours (10 items), letters (15 items), numbers/counting (18 items), sizes/comparisons (22 items), and shapes (20 items). For each subscale, the test continues until completion or three consecutive incorrect responses. All subscales are administered regardless of performance on any one subscale. This task takes 10-15 minutes to complete and has shown good validity and reliability in previous research (Panter & Bracken, 2009). School readiness was indexed by a raw accuracy score across all assessment items.

Procedure

All tasks were administered to children in a quiet area of their preschool centre across four sessions in the same day, to maximize attention and minimize fatigue. Measures were administered in the same order to all children, as follows: (1) HTKS and BSRA; (2) PRSIST memory; (3) PRSIST obstacle course; and (4) PRSIST curiosity boxes. PRSIST assessments and non-PRSIST assessments (i.e., HTKS, BSRA) were conducted by different fieldworkers to remove opportunities for ratings/scores to be biased by a child's performance on the other measures. Each session took between 5-20 minutes to complete, with at least a 1-hour break between sessions. Parents completed the CBRS at the time of consent. Educators completed

the CBRS during the fieldworkers' visit to their centre (yet without knowledge of children's performance on the direct assessments).

Results

Initial Exploratory Data Analysis

Missing data were explored in terms of the extent and patterns of missingness. Four of the children did not have complete PRSIST data due to early departure from their preschool on the day of assessment. Inspection of these cases against available HTKS data showed that while these children were older than the mean age of the sample, their scores were within the overall distribution, each of the values fell within the 95% CI of the sample mean, and their scores were not clustered at an extreme (e.g., very high or very low). Given that this study has as its core focus examination of the PRSIST tool, which is a novel instrument, rather than estimate values for these children they were excluded listwise from relevant analyses. Other missingness was due to parents (n = 4) or educators (n = 1) not returning their adult-report measure. The same inspections of these cases showed that, on the other adult-report measure, they were similarly inside the overall distribution, within the 95% CI of the sample mean, and scores were not clustered at an extreme. Given this result, and that the scales were missing in their entirety, these cases were deleted pairwise so their data were only absent for correlations of this measure with comparison measures. The remaining sample (>91%) had complete data.

As two individual activities were initially selected, further exploratory analyses sought to evaluate the psychometric properties of the two individual activities. Results indicated similar properties of the PRSIST ratings for the group memory game and individual curiosity boxes activities in: factor structure, reliability estimates, and concurrent validity with child outcome measures (reported more fully below). The individual obstacle course activity, in contrast, displayed a discrepant and unanticipated (one-factor) structure that appeared to be influenced by a concentration of high ratings amongst 4- and 5-year-old children; that is, most children

tended to persist with the obstacles, regardless of their success, especially by 4 years of age. In light of this issue, and given the aim of adopting activities that are ideally suited to early years contexts (i.e., briefer, requiring fewer materials, in alignment with current early years practices), the curiosity boxes activity was selected as the individual activity to be subjected to further analysis.

Finally, we examined whether and to what extent the group context of the memory game may have exerted an influence on individual children's responses. While this is a common issue for research that involves multiple ratings within a particular context (e.g., classroom) or within social interactions (e.g., dyads), it presents issues for statistical assumptions of the independence of observations. While planned analyses with comparison measures provide evidence around individual indices derived from the group context correlating with indices from individual assessments, we also ran ICCs to investigate the correlation amongst ratings within groups (Alferes & Kenny, 2009). These ICCs were low (cognitive: .16; behavioral: .07) and non-significant for both subscales, suggesting there was no significant association within group ratings.

Exploratory Factor Analysis

Although the group memory game was expected to impose the greatest self-regulatory challenge, both the group and individual activities were expected to be sufficient to capture dimensions of cognitive, behavioural, and social-emotional self-regulation. While different items were created to reflect specific dimensions of self-regulation, data were first subjected to exploratory factor analyses (EFA) to establish the inherent groupings between scale items, which were compared to a priori theoretical groupings. EFAs of each activity used maximum likelihood estimation and an oblique (direct oblimin) factor rotation, due to expectations that the extracted self-regulation factors would be distinct yet related. For all EFAs, the Kaiser-Meyer-Olkin (KMO) statistic indicated sufficient sampling, KMO > .86, and Bartlett's test of

sphericity was significant (*p*s < .001) indicating inter-item correlations were sufficiently large to justify EFA. Activities were initially analysed separately to evaluate whether the factor structures and item clusters were comparable across situations.

Initial EFAs were conducted on all items that contained sufficient data. For this reason, Item 11 (recovery from an emotional reaction) was excluded, as scoring of this item required an emotional reaction to occur and the incidence of this was low (30.0% for curiosity boxes and 23.8% for memory). Items pertaining exclusively to the memory game (Item 8: being helpful) or curiosity boxes (Item 9: willing to take measured risks) were included in EFAs for those activities in which they could be rated. The number of factors extracted was determined by eigenvalues > 1 (Kaiser, 1960) in conjunction with inspection of scree plots, while item assignment was decided by factor loadings (>.30) and theoretical justification (in cases of cross-loadings). A final EFA was then run on averaged item scores across the two activities, as a final evaluation of the emergent factor solution.

For the memory game and curiosity boxes activities, EFA results yielded two dimensions that could reliably be measured (EFA 1; Table 1). These dimensions could be considered as: *cognitive self-regulation*, consisting of items regarding attention/distraction, engagement, thought/planning, self-direction, and helpfulness; and *behavioral self-regulation*, consisting of items for behavioral control, fidgeting, following social conventions, and resisting strong emotional impulses. However, unexpectedly high ratings and low incidence for one of the items suggested that it may be inappropriate for inclusion. As such, a subsequent EFA (EFA 2; Table 1) was conducted to exclude Item 10 (emotional reaction), given that these activities did not elicit emotional responses or frustration as frequently as initially anticipated (mean ratings ranged from 5.82 to 6.30 across the two activities). Factor loadings and reliability estimates were largely maintained in this subsequent EFA, except for an increased reliability of the behavioral self-regulation factor and stronger correlation between factors across both

activities. A final EFA (Aggregate; Table 1) was conducted on averaged item scores from the two activities (except Items 8 and 9, which reflected ratings from the one activity for which they were rated). While eigenvalues suggested a one-factor solution, the scree plot indicated a clear two-factor solution. A forced extraction of two factors showed clear item loadings on one of the two factors (Table 1), in parallel to the results of individual activity EFAs. This two-factor solution was adopted for all subsequent analyses, on the basis of empirical (two-factor structure, item loadings) and theoretical considerations (items remaining with a priori factors where realignment could not be theoretically justified).

Distributions by Age

Age-related change for each activity and subscale were examined with linear regressions, with self-regulation scores within each activity regressed on age. Results from these analyses showed a significant effect of age for both subscales in the memory game: cognitive, $F(1, 76) = 30.36$, $p < .001$, $R^2 = .29$, $B_{std} = .54$; behavioral, $F(1, 76) = 14.44$, $p < .001$, $R^2 = .16$, $B_{std} = .40$. As depicted in Figure 1 and presented in Table 2, there were also good distributions by age that did not approach floor or ceiling. There were also significant effects of age in the curiosity boxes activity: cognitive, $F(1, 75) = 71.06$, $p < .001$, $R^2 = .49$, $B_{std} = .70$; behavioral, $F(1, 75) = 29.07$, $p < .001$, $R^2 = .28$, $B_{std} = .53$. There were again good distributions by age that did not approach floor or ceiling. As expected, age effects were also apparent for HTKS, $F(1, 79) = 38.53$, $p < .001$, $R^2 = .33$, $B_{std} = .58$, but not for parent-reported, $F(1, 75) = 0.01$, $p = .972$, $R^2 = .00$, $B_{std} = .00$, or educator-reported self-regulation ratings, $F(1, 78) = 2.70$, $p = .105$, $R^2 = .03$, $B_{std} = .18$.

Concurrent Associations between Self-Regulation and Related Measures

Concurrent associations amongst self-regulation measures and with related abilities were additionally examined (see Table 3). PRSIST subscales were highly correlated across both of the activities, and especially so when averaging across the two activities (ps < .001). While

these factors were highly correlated, mean cognitive self-regulation ratings were significantly lower than ratings for behavioral self-regulation (Table 2): memory game, $t(76) = 4.88$, $p < .001$; curiosity boxes, $t(75) = 6.71$, $p < .001$; aggregate, $t(76) = 7.61$, $p < .001$. As such, the two factors appeared to be indexing different developmental levels, although children higher in behavioral self-regulation were also more likely to be higher in cognitive self-regulation.

That these factors were indeed distinct, yet related, was also supported by their differing pattern of associations with the other self-regulation measures. Aggregate PRSIST cognitive self-regulation ratings showed stronger association with HTKS ($r = .63$, $p < .001$) than did ratings of behavioral self-regulation ($r = .50$, $p < .001$) – a statistically significant difference in correlations, $z = 2.71$, $p = .003$. In contrast, educators' ratings of children's self-regulation were associated similarly with aggregate PRSIST ratings of cognitive self-regulation ($r = .49$, $p < .001$) and behavioral self-regulation ($r = .52$, $p < .001$). Correlations of aggregate PRSIST ratings with parent ratings of self-regulation were not as strong (cognitive: $r = .29$, $p = .012$; behavioral: $r = .41$, $p < .001$). This was also the case for correlations of HTKS with educator-report ($r = .36$, $p = .001$) and parent-report ratings ($r = .08$, $p = .483$).

Whereas age was significantly associated with the PRSIST ratings and HTKS scores (ps $< .001$), there was no significant association with age for educator ($p = .105$) or parent ratings of self-regulation ($p = .972$). Further, neither PRSIST nor HTKS was significantly associated with child sex (ps $> .05$), whereas adult reports of self-regulation displayed modest to strong associations with child sex: educator report, $r_{pb} = .48$, $p < .001$; parent report, $r_{pb} = .30$, $p = .010$).

Aggregated PRSIST ratings of cognitive self-regulation were strongly related to school readiness scores ($r = .75$, $p < .001$), as were ratings of behavioral self-regulation ($r = .66$, $p < .001$), albeit to a slightly lesser extent. As expected, HTKS also related to children's school readiness ($r = .77$, $p < .001$), to a similar extent as PRSIST cognitive ratings. Among the

adult-report measures, educator ratings of child self-regulation were modestly associated with school readiness (r = .43, p < .001) and parent ratings showed no significant association (r = .20, p = .084).

Finally, to examine differential associations with PRSIST subscales, validation measures (i.e., HTKS, CBRS, school readiness) were regressed on PRSIST cognitive and behavioral self-regulation indices (step 2), controlling for age and sex (step 1). The model for HTKS was significant, $F(4, 76) = 14.81$, $p < .001$, $R^2 = .45$, and showed that cognitive ($\beta = .53$, $p = .014$) but not behavioral self-regulation ($\beta = -.07$, $p = .699$) significantly predicted HTKS ($\Delta R^2 = .11$). The model for parental CBRS ratings was also significant, $F(4, 72) = 5.27$, $p = .001$, $R^2 = .24$, although the reverse pattern of association was found; behavioural ($\beta = .55$, $p = .014$) but not cognitive self-regulation ($\beta = -.16$, $p = .530$) emerged as a significant predictor ($\Delta R^2 = .13$). For educator CBRS ratings the overall model was again significant, $F(4, 75) = 13.69$, $p < .001$, $R^2 = .44$, and the addition of PRSIST at step 2 significantly improved the model ($\Delta R^2 = .17$, $p < .001$). However, neither behavioral ($\beta = .21$, $p = .239$) nor cognitive self-regulation indices ($\beta = .37$, $p = .085$) emerged as significant independent predictors of educators' CBRS ratings. This same pattern emerged for school readiness, such that the overall model was significant, $F(4, 76) = 30.76$, $p < .001$, $R^2 = .63$, and the addition of PRSIST at the second step significantly improved the overall model ($\Delta R^2 = .12$, $p < .001$). Once again, however, neither behavioral ($\beta = .19$, $p = .208$) nor cognitive self-regulation indices ($\beta = .31$, $p = .074$) independently predicted school readiness scores.

Given the close association between PRSIST and HTKS, we further examined whether the association between PRSIST and school readiness was still observed after additionally controlling for HTKS at step 1. Once again, the overall model was significant, $F(4, 76) = 41.24$, $p < .001$, $R^2 = .73$. Even after controlling for HTKS, we found that the addition of PRSIST at the second step yielded significant improvement in the model ($\Delta R^2 = .04$, p =

.004). Again, neither PRSIST cognitive (β = .07, p = .639) nor behavioral subscales (β = .22, p = .081) independently predicted school readiness. These results thus provided additional evidence for PRSIST's association with the validation measures (that are not accounted for by age or sex), and two distinct PRSIST factors that show different patterns of association across the comparison measures.

Discussion

The current study provides strong initial evidence of the viability – indeed, strength – of a structured observational approach to self-regulation assessment. This is in contrast to the challenges and limitations of previous attempts to adopt this approach (e.g., Lakes, 2012). In particular, the current study demonstrated the PRSIST assessment could clearly identify two distinct yet related cognitive and behavioral self-regulation dimensions using an observation approach. In addition, the PRSIST assessment showed: strong psychometric properties that paralleled those of direct task-based approaches; a high degree of association with a direct task-based measure of self-regulation; and good developmental sensitivity. Rather than just introducing a new self-regulation assessment that replicates the strengths of an existing task-based measure, the PRSIST assessment provides distinct advantages and possibilities over existing tools. These include: (1) the ability to distinguish between discrete dimensions of self-regulation, in contrast to the singular self-regulation index typically derived from task-based measures; (2) the ability to generate actionable, formative-like information that is linked to real-world self-regulatory behaviors and developmental sequences; and (3) the integration of assessment within existing routines and practices in early years settings.

The ability to derive valid and reliable indices of distinct dimensions of self-regulation was a feature uniquely demonstrated by PRSIST amongst the measures considered. Within task-based approaches, despite strong validity and reliability evidence, assessments typically yield only a single index of self-regulation (Ponitz et al., 2009). Whether this score indexes a

particular form of self-regulation, or represents some combination of multiple self-regulation dimensions, is unclear. This is problematic insofar as latent cognitive and behavioral self-regulation factors appeared distinct in their association in the current study. That is, while children who were higher in one form of self-regulation tended to also be higher in the other form of self-regulation, there were numerous cases in which these indices were discrepant. While studies using adult-report measures document adults' perceived discrepancies between children's cognitive, behavioral, and social-emotional self-regulation (Howard & Melhuish, 2017), their lack of developmental sensitivity and subjectivity in ratings create issues for tracking developmental progress. Still, prevention, education, and intervention efforts may be enhanced with an improved understanding of the particular area(s) and condition(s) in which a child is experiencing self-regulatory difficulties. Indeed, there is clear evidence for superior effects of targeted and adaptive over general intervention strategies in a range of educational areas (Adams, Sallis, Norman, Hovell, Hekler, & Perata, 2013; Green et al., 2012; Holmes, Gathercole, & Dunning, 2009). In relation to self-regulation, there are programs that similarly provide different strategies, practices, and child activities to target its distinct domains (e.g., focused on sustaining attention, behavioral control, social interaction, or emotions; Howard, Vasseleu, Neilsen-Hewett, & Cliff, 2018). While it is noted that correlations among PRSIST subscale scores were high, the fact that disparate patterns of association occurred for the two factors indicated that these factors were distinct. As such, of the measures evaluated, the PRSIST assessment was uniquely able to capture multiple valid and reliable dimensions of self-regulation.

Further problems with adult-report data were highlighted by their lack of developmental sensitivity and apparent gender biases. It is perhaps for this reason that previous research has found that adult-report indices of self-regulation predict different outcomes than task-based indices (Schmitt et al., 2014). Whereas in Schmitt et al.'s (2014) study no associations were

found for their observation measure with HTKS or academic outcomes, in this study PRSIST was strongly related with HTKS. This is particularly impressive given their highly discrepant approaches (i.e., in HTKS touching a body part that is the opposite to what was instructed; in PRSIST playing a group memory card game and a guessing game with boxes). In the current study, both the observation and task-based measures also showed strong predictive validity with a standardized measure of school readiness. This is consistent with prior findings for task-based approaches (McClelland et al., 2014; Ponitz et al., 2009; Schlam et al., 2013), and further clarifies suggestions that "the ability to observe children's regulatory behaviors may be more difficult during free choice time, which in turn, could limit the predictive utility of observer reports to achievement" (Schmitt et al., 2014, p. 24). Indeed, observations in that study occurred during unstructured play, introducing the possibility that self-regulation was rated in the context of children self-selecting in or out of more complex forms of free play on the basis of their self-regulatory abilities. In the current study and approach, children engaged in a common set of activities (albeit semi-structured and therefore variable in the behaviors, responses, and interactions encountered) selected to ensure self-regulatory challenge. While the two activities varied in their degree of structure, neither was structured in the manner of a typical behavioral assessment (i.e., fixed instructions, corrections, and trials, and controlled environments) or was completely unstructured (e.g., self-selected free play). Indeed, the level of structure in the PRSIST activities was consistent with what children routinely experience (e.g., in individual or group games that ask children to follow a sequence, take turns, and/or remain within the rules of the game).

Nevertheless, this inconsistency across PRSIST assessment situations and instances (due to variable dynamics, behaviors, and responses of the children) is uncommon and is typically avoided in traditional assessment design. Rather than a limitation, this variability can instead be viewed as one of the strengths of the PRSIST approach. First, it privileges ecological

validity in that it more closely reflects children's real-world self-regulatory behaviors in genuine situations. Second, that no two iterations of an activity are exactly the same serves to minimize potential learning effects of repeated administrations. This is a challenge for task-based measures once a child internalizes its rules, strategies, or sequences (Basso, Bornstein, & Lang, 1999; Muller, Kerns, & Konkin, 2012), and thereby limits their utility in educational applications that require repeated formative assessments. This is not to say that the PRSIST assessment appeared inappropriate for research use; rather, its strong and similar properties to our task-based measure suggests it is indeed suitable for such use. Yet if the goal is a single accurate estimate of self-regulation from young children, as expediently as possible, HTKS – given its comparable administration time but shorter training time – may be better suited to these purposes. That is, while the association of PRSIST with HTKS suggests its concurrent validity with this task-based measure, HTKS' correlation with PRSIST additionally suggests that HTKS scores associate well with ratings of real-world self-regulatory behaviors derived from structured observation.

Yet using a task-based measure is not without its limitations. As previously mentioned, a single index of self-regulation precludes understanding potential differences in ability across different dimensions of self-regulation. It also fails to provide clear information that informs educational practice, planning, or interventions; what does a score of 14, for example, mean for adults' practices and planning for that child? In contrast, PRSIST provides data that can: draw attention to distinct dimensions and constituent behaviors of self-regulation, as well as their developmental continua (improving educators' knowledge of what self-regulation and its development looks like). From these data, the identification of areas in which a child may be challenged can inform the implementation of targeted practices and strategies, such as those specified for different self-regulation dimensions in some programs (Howard et al., 2018). By linking the items and observations to authentic situations, using PRSIST can also

structure the lens through which educators can then view children's behaviors, responses, and interactions outside of PRSIST activities. For these possibilities to be fully realized, further research is required. For instance, the possibility of educator use of the PRSIST assessment in a way that meets the rigor and reliability of researcher use warrants investigation. This could follow the structure adopted by Howard, Neilsen-Hewett, de Rosnay, Buckley-Walker, and Melhuish (2018) to yield test-retest reliability and educator-researcher comparability data. Moreover, investigations with larger samples would permit latent variable analyses to more purely model development of self-regulation, its fluctuations over time and contexts (e.g., group versus individual, across groupings), and its associations with related abilities (e.g., executive function, school readiness).

The extent to which observation approaches to self-regulation assessment can promote practice change, in and of itself, is also an interesting area of investigation. For instance, can providing structure to educators' observations of children lead to practice change and better child outcomes through more comprehensive, accurate, and actionable insights about these children, even in lieu of programs that prescribe specific actions to remediate or extend their current self-regulatory abilities?

The inability to capture a social-emotional dimension of self-regulation in the current study suggests either that this dimension is indistinguishable from behavioral and cognitive aspects of self-regulation, or that the current activities were not well suited to elicit emotional responses and recovery. While the current data are not able to conclusively decide amongst these options, the prevalence of maximum ratings on the emotional reaction items and N/A ratings on the emotional recovery items (where no emotional reaction was evident) suggests the latter. It may be that other situations or paradigms are better suited to elicit these aspects (Fink, Heathers, & de Rosnay, 2015), or that assessment of socio-emotional self-regulation is best achieved through positive evidence that children manage relationships well and exhibit

social-emotional maturity (Fink, de Rosnay, Peterson, & Slaughter, 2013), thereby avoiding strong emotion elicitation. Low emotional reactivity across the PRSIST activities suggests that eliminating the disappointing gift element from the curiosity boxes activity, paralleling removal of emotion-related items of the scale, may eliminate a point of concern for educators without influencing function of the scale. That two social self-regulation items unexpectedly clustered on the cognitive and behavioral self-regulation factors additionally warrants further investigation in larger and more diverse samples.

Lastly, while final analyses did control for age and sex, additional factors that could have influenced between-task associations are also plausible (e.g., maternal education, quality of home learning environment) and subsequent studies are needed to investigate the potential contributions of these factors. This includes research with larger and more diverse samples, given the modest sample size and middle- to high-SES catchment area for the current sample. Extending this research longitudinally would permit investigation of the growth trajectories that different self-regulation measures elicit.

Conclusion

While these areas of further investigation are indeed important, the direction and strength of the current results are clear. Specifically, the current study evidenced the viability of a structured observational approach to self-regulation assessment, which displayed very strong concurrent and predictive validity, and good to excellent psychometric reliability. Viability of this approach also proffers wide-ranging possibilities unique to this approach, such as: the ability to put the tool in the hands of preschool educators for use as a research or formative assessment tool; providing data from which to tailor education and intervention efforts to profiles of development across distinct self-regulation dimensions; and use as an in-service professional development tool, illustrating the domain-specificity of self-regulation and structuring educators' everyday observations toward these distinct domains. At the same

time, this study serves as a framework from which similar assessment tools in other domains of development could be created, evaluated, and extended. Given the need for valid tools that capture young children's real-world abilities in self-regulation and beyond, and the growing prominence of formative assessment (i.e., assessment *for* learning) in early years and primary curricula and frameworks, the current study provides an important first step in the creation of rigorous yet flexible tools that have both summative and formative utility.

References

ABS (Australian Bureau of Statistics). (2008). *Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA)* (cat. no. 2039.0). Retrieved August 16 2016 from http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0

Adams, M. A., Sallis, J. F., Norman, G. J., Hovell, M. F., Hekler, E. B., & Perata, E. (2013). An adaptive physical activity intervention for overweight adults: A randomized controlled trial. *PLOS One, 8*(12), e82901. doi: 10.1371/journal.pone.0082901

Alferes, V. R., & Kenny, D. A. (2009). SPSS programs for the measurement of nonindependence in standard dyadic designs. *Behavior Research Methods, 41*(1), 47-54. doi: 10.3758/BRM.41.1.47

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. The Clinical Neuropsychologist, 13(3), 283-292. doi: 10.1076/clin.13.3.283.1743

Bracken, B. A. (2002). *Bracken School Readiness Assessment* [Measurement instrument]. San Antonio, TX: Psychological Corporation.

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). *Child behavior rating scale*. Cambridge, MA: Abt Associates.

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., … Ramon, M. (2014). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: a replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry, 55*(3), 247-255. doi: 10.1111/jcpp.12146

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*(3), 217-227. doi: 10.1016/j.acn.2005.12.002

Dunning, D. J., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science, 16*(6), 915-925. doi: 10.1111/desc.12068

Fink, E., de Rosnay, M., Peterson, C., & Slaughter, V. (2013). Validation of the Peer Social Maturity Scale for assessing children's social skills. *Infant and Child Development, 22*(5), 539-552. doi: 10.1002/icd.1809

Fink, E., Heathers, J.A., & de Rosnay, M. (2015). Young children's affective responses to another's distress: Dynamic and physiological features. *PLoS One, 10*(4), e0121735. doi: 10.1371/journal.pone.0121735

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction, 30*, 25-31. doi: 10.1016/j.learninstruc.2013.11.003

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1337-1345. doi: 10.1097/00004583-200111000-00015

Green, C. T., Long, D. L., Green, D., Iosif, A.-M., Dixon, J. F., … Schweitzer, J. B. (2012). Will working memory training generalize to improve off-task behavior in children with attention-deficit/hyperactivity disorder? *Neurotherapeutics, 9*(3), 639-648. doi: 10.1007/s13311-012-0124-y

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900-1902. doi: 10.1126/science.1128898

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2014). Executive functions and self-regulation. *Trends in Cognitive Sciences, 16*, 174-180. doi: 10.1016/j.tics.2012.01.006

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*(4), 9-15. doi: 10.1111/j.1467-7687.2009.00848.x

Howard, S. J., & Melhuish, E. C. (2017). An Early Years Toolbox (EYT) for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment, 35*(3), 255-275. doi: 10.1177/0734282916633009

Howard, S. J., Neilsen-Hewett, C., de Rosnay, M., Grimmond, J., Buckley-Walker, K., & Melhuish, E. C. (2019). Early Years Toolbox (EYT) Numbers: Validity, reliability, and viability of pre-school educators' use of a game-like assessment of early number understanding. Manuscript submitted for publication.

Howard, S. J., Vasseleu, E., Neilsen-Hewett, C., & Cliff, K. (2018). Evaluation of the Preschool Situational Self-Regulation Toolkit (PRSIST) program for supporting children's early self-regulation development: Study protocol for a cluster randomized controlled trial. *Trials, 19*, 64. doi: 10.1186/s13063-018-2455-4

Howard, S. J., & Williams, K. E. (2018). Early self-regulation, early self-regulatory change, and their longitudinal relations to adolescents' academic, health, and mental wellbeing outcomes [online first]. *Journal of Developmental & Behavioral Pediatrics*. doi: 10.1097/DBP.0000000000000578

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi: 10.1016/j.jcm.2016.02.012

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science, 340*(6130), 297-300. doi: 10.1126/science.1235350

Lakes, K. (2012). The Response to Challenge Scale (RCS): The development and construct validity of an observer-rated measure of children's self-regulation. International Journal of Educational and Psychological Assessment, 10(1), 83-96.

Lakes, K. D., & Hoyt, W. T. (2004). Promoting self-regulation through school-based martial arts training. *Journal of Applied Developmental Psychology, 25*, 283-302. doi: 10.1016/j.appdev.2004.04.002

Matthews, J.S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., … Pratt, M. E. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Frontiers in Psychology, 5*, 599. doi: 10.3389/fpsyg.2014.00599

Miller, M., Nevado-Montenegro, A. J., & Hinshaw, S. P. (2012). Childhood executive function continues to predict outcomes in young adult females with and without childhood-diagnosed ADHD. *Journal of Abnormal Child Psychology, 40*(5), 657-668. doi: 10.1007/s10802-011-9599-y

Mischel, W., Shoda, Y, & Rodriguez, M. I. (1989). Delay of gratification in children. *Science, 244*(4907), 933-938. doi: 10.1126/science.2658056

Moffitt, T., Arsenault, L., Belsky, D., Dickson, N., Hancox, R. J., … 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. doi: 10.1073/pnas.1010076108

Muller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist, 26*(2), 271-287. doi: 10.1080/13854046.2011.645558

Murray, D. S., Ruble, L. A., Willis, H., & Molloy, C. A. (2009). Parent and teacher report of social skills in children with Autism Spectrum Disorders. *Language, Speech, and Hearing Services in Schools, 40,* 109-115. doi: 10.1044/0161-1461(2008/07-0089)

OECD (Organisation for Economic Co-operation and Development). (2018). *Education at a glance 2018: OECD indicators*. Paris: OECD Publishing. doi: 10.1787/eag-2018-en

Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken School Readiness Assessment for predicting first grade readiness. *Psychology in the Schools, 46*(5), 397-409. doi: 10.1002/pits.20385

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 605-619. doi: 10.1037/a0015365

Rademacher, A., & Koglin, U. (2018). The concept of self-regulation and preschoolers' social-emotional development: A systematic review. *Early Child Development and Care.* Advance online publication. doi: 10.1080/03004430.2018.1450251

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*(4), 958-972. doi: 10.1037/a0015861

Schlam, T. R., Wilson, N. L., Shoda, Y., Mischel, W., & Ayduk, O. (2013). Preschoolers' delay of gratification predicts their body mass 30 years later. *Journal of Pediatrics, 162*(1), 90-93. doi: 10.1016/j.jpeds.2012.06.049

Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development, 25*(5), 641-660. doi: 10.1080/10409289.2014.850397

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*(2), 173-187. doi: 10.1016/j.ecresq.2007.01.002

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). The Effective Provision of Pre-School Education (EPPE) project: Findings from pre-school to end of Key Stage1. Nottingham, UK: DfES.

Tominey, S. L., & McClelland, M. M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education and Development, 22*(3), 489-519. doi: 10.1080/10409289.2011.574258

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – is younger better? *Developmental Review, 32*(4), 360-387. doi: 10.1016/j.dr.2012.07.001

Table 1

EFA Factor Loadings, Reliability Estimates, and Latent Factor Correlations, by Activity and EFA subscale

	Memory				**Curiosity Boxes**				**Aggregate**	
	EFA 1		**EFA 2**		**EFA 1**		**EFA 2**			
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Cognitive Self-Reg.	**α = .88**		**α = .88**		**α = .95**		**α = .95**		**α = .95**	
1. Attention/Distract	**.32**	.76	**.15**	.84	**.74**	.28	**.29**	.72	**.42**	.62
2. Engagement	**.89**	.07	**.87**	.08	**.94**	.06	**.73**	.33	**.87**	.13
3. Thought/Plan	**.92**	.07	**.90**	.08	**.95**	-.02	**.82**	.21	**.90**	.11
4. Self-Directed	**.85**	.12	**.82**	.13	**.93**	-.04	**.73**	.27	**.93**	.03
8. Helpful / 9. Risk	**.63**	-.07	**.66**	-.09	**.90**	-.27	**1.05**	-.24	**.82**	-.12
Behavioral Self-Reg.		**α = .81**		**α = .87**		**α = .69**		**α = .74**		**α = .86**
5. Behavior Control	.30	**.69**	.11	**.80**	.87	**.14**	.58	**.46**	.51	**.49**
6. Fidget/Restless	.05	**.80**	-.14	**.89**	.22	**.49**	-.07	**.56**	-.10	**.86**
7. Social Convention	.25	**.73**	.08	**.81**	.44	**.58**	.22	**.51**	.11	**.82**
10. Emotional Reaction	-.19	**.62**	-	-	-.17	**.68**	-	-	-	-
SR factor correlations	**r_{f12} = .41**		**r_{f12} = .58**		**r_{f12} = .25**		**r_{f12} = .56**		**r_{f12} = .62**	

Note. EFA 1 = exploratory factor analysis factor loadings for EFA with all eligible items. EFA 2 = exploratory factor analysis excluding Items 8 and 9. Aggregate = EFA after averaging item scores derived from each activity. Cronbach alphas (α**)** and correlation between factors (r_{f12}) are also reported for each subscale. Items are presented and bolded for the factors they were assigned to for Cronbach alpha computations.

Table 2

Descriptive statistics for self-regulation and school readiness measures overall and by age

	Overall	**Age 3**		**Age 4**		**Age 5**	
	M (SD)	M (SD)	Range	M (SD)	Range	M (SD)	Range
PRSIST Aggregate							
Cognitive SR	3.55 (1.00)	2.82 (0.80)	1.10-4.20	3.69 (0.83)	1.10-4.80	4.41 (0.72)	2.70-5.70
Behavioral SR	4.00 (0.98)	3.44 (0.85)	1.13-5.00	4.12 (0.82)	1.50-5.38	4.64 (0.99)	3.00-6.00
PRSIST Memory							
Cognitive SR	3.65 (1.06)	2.98 (0.97)	1.00-5.00	3.85 (0.91)	1.00-5.20	4.34 (0.84)	2.40-5.40
Behavioral SR	4.06 (1.11)	3.54 (0.89)	1.00-5.50	4.20 (1.09)	1.25-5.75	4.62 (1.16)	2.25-6.25
PRSIST Curiosity							
Cognitive SR	3.46 (1.16)	2.66 (0.91)	1.20-4.40	3.56 (0.99)	1.20-5.40	4.47 (0.85)	2.80-6.00
Behavioral SR	3.94 (1.05)	3.34 (1.06)	1.25-4.75	4.04 (0.75)	1.75-5.50	4.66 (0.96)	2.75-6.00
HTKS	30.54 (28.88)	10.96 (18.66)	0-63	35.77 (26.52)	0-90	48.90 (28.91)	2-85
CBRS Parent	3.49 (0.54)	3.53 (0.34)	2.90-4.30	3.47 (0.49)	2.50-4.30	3.46 (0.82)	1.80-4.70
CBRS Educator	3.54 (0.74)	3.46 (0.67)	1.00-6.50	3.49 (0.72)	2.00-4.90	3.70 (0.84)	2.30-5.00
School Readiness	48.54 (18.63)	33.04 (14.48)	1-65	52.77 (14.17)	6-78	62.95 (14.28)	25-82

Note. SR = self-regulation. HTKS and school readiness are reported as raw scores. PRSIST and CBRS are presented as mean (sub)scale scores.

Table 3

Concurrent associations between measures

	1	2	3	4	5	6	7	8	9	10	11	12
1. CSR (Aggreg.)	-	.86*	.89*	.75*	.91*	.83*	.69*	.09	.63*	.75*	.29*	.49*
2. BSR (Aggreg.)		-	.78*	.92*	.79*	.90*	.51*	.16	.50*	.66*	.41*	.52*
3. CSR (Memory)			-	.77*	.63*	.64*	.54*	.08	.49*	.62*	.34*	.45*
4. BSR (Memory)				-	.59*	.65*	.40*	.15	.36*	.54*	.43*	.47*
5. CSR (Curiosity)					-	.84*	.70*	.07	.65*	.73*	.21	.44*
6. BSR (Curiosity)						-	.53*	.13	.57*	.68*	.32*	.48*
7. Age							-	-.12	.58*	.67*	.00	.18
8. Sex								-	-.03	.12	.30*	.48*
9. HTKS									-	.77*	.08	.36*
10. School Read.										-	.20	.43*
11. CBRS Parent											-	.40*
12. CBRS Educ.												-

Note. CSR = cognitive self-regulation. BSR = behavioral self-regulation. Mem = PRSIST memory game activity. Cur = PRSIST curiosity boxes activity. Aggreg. = subscale scores after averaging the items across the two activities. HTKS = Head-Toes-Knees-Shoulders task. *$p < .05$.

Figure Captions

Figure 1. *PRSIST self-regulation scores by subscale, age and activity*. Error bars represent +/- 1 standard error.

Appendix A

Preschool Situational Self-Regulation Toolkit (PRSIST) Assessment

Child Name/ID: _____ Child Sex: M / F **Child Age: _____ Rater: _____ Date: _____ Activity Rated:** M CB

Observer notes: For each item, please rate the degree to which the child engages in the described behaviour in the activity (circle one rating for each item):

1.	**Did the child sustain *attention,* and resist *distraction,* throughout the instructions and activity?**	1	2	3	4	5	6	7
	This item focuses on the child paying attention throughout the activity: during the facilitator's instructions, during their turn and during others' turns. Internal or external occurrences, sounds or objects rarely distract their attention. To rate this item, you have to pay attention to where children are looking throughout the game or activity. *At a score of 1*, a child pays virtually no attention to any aspects of the instructions or game. *At a score of 7*, a child has virtually no lapses in attention and pays careful attention at all times.							
2.	**Was the child *engaged* in the activity throughout its duration?**	1	2	3	4	5	6	7
	This item focuses on engagement, which is related to their involvement and investment in the activity. A child can be paying attention (as in Item 1), but showing little active investment in the activity. Behaviours such as reacting (e.g., to someone else getting a pair), responding (e.g., to words or actions of others), asking and answering questions (e.g., responding to questions of 'What do you think is in the box?'), and/or following requests (e.g., not shaking the box before permitted) would all be indicators of engagement. *At a score of 1*, a child is barely engaged in the activity. They may often look at the facilitator and aspects of the activity, but their involvement is entirely passive and reactionary (or non-existent). *At a score of 7*, a child is reacting to the things that happen in the activity, such as responding to and asking questions and following requests. They are constantly active participants for the duration of the activity.							
3.	**Was the child *thoughtful* and *planful* before acting?**	1	2	3	4	5	6	7
	This item focuses on the child's mental effort put toward the activity. Evidence of being thoughtful includes pauses to think and consider, revision of initial responses (e.g., revises guess of what's in the box based on new information) and not perseverating on the same mistakes (e.g., keeps flipping the same two cards). *At a score of 1*, a child is responding quickly (almost reflexively) and repeating the same mistakes in their responses or actions. *At a score of 7*, a child is taking time and effort to think, remember and avoid repeating mistakes. They may also revise initial responses (e.g., start to reach for a card but, adopting a better strategy, stops and selects another).							
4.	**Was the child *self-directed,* engaging in the activity with little prompting?**	1	2	3	4	5	6	7
	This item focuses on whether the child is able to internalise the sequence and requirements of the activity, and independently enact this. This includes not only knowing when it is their turn and/or what to do next, but enacting this with little to no prompting. *At a score of 1*, a child is highly other-regulated. Even if they appear to know what the next step/requirement is, they consistently require prompting and affirmation from the facilitator to do this (e.g., 'Yes, it's your turn', 'Now you can shake the box'). *At a score of 7*, after some initial time to familiarise with the activity, the child does not require reminders, prompts or affirmations to continue the activity (e.g., if a child finds a pair, they independently commence their second turn).							
5.	**Did the child *control their behaviours* and stay within the rules of the activity?**	1	2	3	4	5	6	7
	This item focuses on the extent to which children resist behavioural impulses. This could include skipping turns or steps (e.g., skipping the requirement to gently lift), breaking rules of the game or activity or, in more extreme cases, being physically or verbally aggressive. *At a score of 1*, a child is showing complete disregard for the rules of the game, has engaged in serious physical or verbal aggression towards others and/or is highly disruptive of the game (e.g., wiping cards off the table, immediately lifting the lid of the box). *At a score of 7*, a child remains within the rules of the activity, is calm and controlled, and has not given in to negative impulses.							

6.	**Did the child *remain in their seat* and *rarely fidget*?**	1	2	3	4	5	6	7
	This item focuses on whether the child is able to remain reasonably still. Fidgeting with their hands, legs, feet, body, or leaving their chair entirely (other than briefly, to reach something required for the activity) apply to this item. *At a score of 1*, a child is almost always fidgeting and/or leaving their chair. *At a score of 7*, a child is rarely fidgeting and does not leave their chair for reasons unrelated to the activity's requirements.							
7.	**Did the child follow *social conventions* of the situation?**	1	2	3	4	5	6	7
	This item focuses on whether a child adheres to the general social requirements of the situation. This includes things like waiting their turn, not beginning the activity before instructions are complete, being considerate to others (e.g., not talking over others), acknowledging others' successes and responding to questions. *At a score of 1*, a child is consistently lacking consideration for others or is being actively inconsiderate (e.g., not waiting their turn, actively ignoring). If there are some positive and negative social behaviours, consider the frequency and gravity of these to decide a rating. *At a score of 7*, a child is consistently considerate of others' turns and speech, is responsive to and considerate of others, and celebrates others' successes.							
8.	*Memory game only:* **Did the child take opportunities to be *helpful* and supportive to the adult or other children?**	1	2	3	4	5	6	7
	This item focuses on whether, in a social situation like a group memory card game, a child takes opportunities to be helpful to others. Opportunities to help are prevalent – answering questions that are posed, helping a child that is unclear about next steps or is struggling with an action, providing honest guidance, and giving support. *At a score of 1*, a child engages in no instances of helping behaviour. *At a score of 7*, although some potential opportunities for help may be missed, a child generally and frequently offers help. For a score of 7, you should feel that the child has a general disposition toward helping, and routinely acts upon this.							
9.	*Curiosity boxes only:* **Was the child willing to *risk* being wrong when uncertain?**	1	2	3	4	5	6	7
	This item focuses on whether, under conditions of uncertainty, a child is willing to engage with a problem, challenge or activity despite the risk of being wrong. For instance, despite having little information from which to guess an item that is inside a box, will they engage in the challenge of guessing? *At a score of 1*, a child does not make any guesses while the boxes are closed, and quickly peeks once feeling the object in their hand. *At a score of 7*, a child will almost invariably guess what is in the boxes, regardless of the level of uncertainty.							