Learning-Related Cognitive Self-Regulation Measures for Prekindergarten Children with Predictive Validity for Academic Achievement

Working Paper

Mark W. Lipsey[a]
Kimberly Turner Nesbitt[a]
Dale C. Farran[a]
Nianbo Dong[b]
Mary Wagner Fuhs[c]
Sandra Jo Wilson[a]

May 2014

[a] Vanderbilt University, [b] University of Missouri-Columbia, [c] University of Dayton

PEABODY research institute

VANDERBILT PEABODY COLLEGE

Peabody Research Institute
Vanderbilt University

Our mission is to conduct research aimed at improving the effectiveness of programs for children, youth, and families. Using field research, program evaluation, and research synthesis (meta-analysis), our faculty and staff help determine which programs are actually making a difference in the lives of the people they serve. PRI research addresses many aspects of child and family programs, such as their implementation, costs, dissemination, and social or political support. But the main focus for all of our work is the effects of programs on children and families.

Author's Note:

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080079 to Vanderbilt University. Kimberly Turner Nesbitt and Mary Wagner Fuhs were supported by an Institute of Education Postdoctoral Training Grant (R305B100016) awarded to Dale C. Farran and Mark W. Lipsey. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Special thanks to Deanna Meador and the Peabody Research Institute research team for their very capable management of the data collection for this project.
Correspondence should be addressed to Mark Lipsey, Peabody Research Institute, Vanderbilt University, Nashville, TN 37203. E-mail: mark.lipsey@vanderbilt.edu.

Funding Source:

Development supported by the Institute of Education Sciences. Grant #R305A080079.

Contact Us:

Phone: 615.322.8540
Fax: 615.322.0293

Mailing Address:

Peabody Research Institute
230 Appleton Place
PMB 181
Nashville, TN 37203-5721

Delivery Address:

Peabody Research Institute
1930 South Drive
Room 410A
Nashville, TN 37212

Abstract

Identifying and understanding the foundational skills children need to participate effectively in formal schooling is an important objective for research in early childhood education. One component of school readiness is cognitive self-regulation (CSR). The question this study addresses is how to assess CSR with prekindergarten-aged children in a way that taps into the learning-related cognitive engagement behaviors teachers observe in the classroom that are predictive of later academic achievement. A number of candidate measures applicable to prekindergarten age children have been generated for research on attention, effortful control, executive function, and related constructs. A diverse set of twelve candidate measures that can be easily administered in school settings was selected from these domains and applied to a sample of prekindergarten children. These measures were then examined for construct validity, developmental change, convergent validity with teacher ratings of CSR, and predictive validity for subsequent academic achievement and achievement gain. Six measures performed well by these criteria: Peg Tapping, Head-Toes-Knees-Shoulders (HTKS), the Kansas Reflection-Impulsivity Scale for Preschoolers (KRISP), Dimensional Change Card Sort (DCCS), Copy Design, and Backwards Digit Span. Cross-validation with a new sample of children confirmed the validity of these measures, estimated their test-retest reliability, identified the best performing individual measures, and demonstrated that a composite score combining results from all these measures performed better than any single measure.

Keywords: school readiness, measurement, cognitive self-regulation, executive function

Learning-Related Cognitive Self-Regulation Measures for Prekindergarten Children with Predictive Validity for Academic Achievement

A critical objective for research in early childhood education is identifying and understanding the foundational abilities and knowledge children need to participate effectively in formal schooling. In the prekindergarten (pre-k) context, this is a question of what skills young children should develop to ensure that they will be ready to take advantage of the learning opportunities in kindergarten and beyond. A key empirical test of the importance of a school readiness skill, therefore, is its predictive relationship to subsequent achievement. Among the recognized aspects of school readiness are emergent literacy and math skills, and these indeed are significant predictors of later achievement (Duncan et al., 2007; LaParo & Pianta, 2000). Another component of school readiness that has garnered attention is self-regulation, broadly defined as the ability to deliberately control the quality, sequence, direction, and persistence of one's behaviors, cognition, and emotions (Schunk & Zimmerman, 1997).

Young children's ability to exert control over their cognition and behaviors within educational contexts has been variously labeled approaches to learning (Kagan, Moore, & Bredekamp, 1995; Meisels, Atkins-Burnett**,** & Nicholson, 1996), learning dispositions (Katz, 1993, 2002), and work-related skills (Cooper & Farran, 1988, 1991). However labeled, research has demonstrated that children's ability to focus on classroom tasks, persist despite difficulty, and engage in learning activities are positively related to academic achievement (Bronson, Tivnan, & Seppanen, 1995; Cooper & Speece, 1988; Duncan et al., 2007; McClelland, Morrison, & Holmes, 2000). More generally, this constellation of skills can be referred to as cognitive self-regulation (Blair, 2002).

Teacher ratings provide one source of measures of cognitive self-regulation (CSR) skills in early childhood educational contexts and they have been found to be predictive of later academic achievement. For example, kindergarten teachers' ratings of children's approaches to learning in the ECLS-K study (persistence, eagerness, attention, learning independence, flexibility, and organization) predicted mathematics achievement gain across the first four years of school (Bodovski, & Farkas, 2007). Similarly, McClelland, Acock, and Morrison (2006) found that teacher ratings of kindergarten CSR predicted reading and math achievement between kindergarten and sixth grade, and growth in literacy and math from kindergarten to second grade.

Teacher ratings have significant advantages as measures of CSR in educational contexts. They are based on relevant classroom behaviors and draw on the many observations teachers are able to make during varied instructional activities. They are also relatively easy to administer and, as noted, demonstrate significant associations with later academic achievement and gains in achievement. Nonetheless, there are circumstances and purposes for which teacher ratings are inappropriate or unattainable. For instance, they require that teachers have sufficient opportunity to observe children's behavior in the classroom and thus cannot measure children's abilities at the onset of school for purposes of screening for poor CSR skills. They are also problematic for research on interventions delivered by teachers for which CSR is an outcome of interest. In this situation, the teachers in the experimental condition cannot be blinded to the intervention and its implications for CSR, and that awareness may bias their ratings relative to those of control teachers who are not as engaged in the intervention or as sensitized to CSR issues.

Our interest is in direct assessments appropriate for pre-k children of those forms of CSR teachers observe in the classroom that are predictive of later academic achievement. We refer to this as *learning-related cognitive self-regulation* (LRCSR). By direct child assessments, we mean measures that can be administered by an objective assessor directly to the children of interest. A number of candidate measures appropriate for pre-k children are available. The central question for this study is which of these show convergent validity with teacher ratings and predictive validity for later achievement and achievement gains.

Direct Assessment of Cognitive Self-Regulation

Research on CSR has been conducted within various conceptual frameworks including attention, effortful control, and executive function. Executive function is generally defined as a set of effortful cognitive abilities that aid in the completion of goal-directed actions (Miyake et al., 2000). These abilities include adapting or shifting actions to changing situational demands (Zelazo, Frye, & Rapus, 1996), active maintenance and manipulation of information in working memory (Baddeley & Hitch, 1974), and inhibition of inappropriate but prepotent responses (Diamond, 1990). Effortful control, in turn, is typically conceptualized as attentional functions such as conscious detection and sustained attention to a target stimulus (Posner & Rothbart, 2000; Rothbart & Ahadi, 1994) and behavioral regulation (Kochanska, Murray, & Harlan, 2000). A number of direct assessments of CSR-related constructs suitable for pre-k children have been developed within these research contexts, and some of those have been shown to be related to concurrent or future academic achievement (Allan & Lonigan, 2011; Blair & Razza, 2007; Gathercole, Brown, & Pickering, 2003; Lan, Legare, Ponitz, Li, & Morrison, 2011) and achievement gains during the pre-k and kindergarten years (Matthews, Ponitz, & Morrison, 2009; McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009).

The objective of the present study was to draw on this pool of existing measures to identify a set of direct assessment LRCSR measures with properties that make them especially suitable as school readiness measures for pre-k children. We considered only measures that could be easily administered in school settings—those that could be completed in a relatively brief period without specialized equipment or an online Internet connection. The properties then used as criteria to determine their suitability as LRCSR measures were, first, that the construct measured was recognizable as CSR or a facet of CSR. Second, the measures should be responsive to developmental change, i.e., show nontrivial increases as CSR skills improve through maturation and facilitation. Third, the measures should have convergent validity with teacher ratings of CSR to ensure their educational relevance and ecological validity for pre-k settings. Fourth, the measures should have predictive validity for subsequent achievement and gains in achievement to affirm their relevance to school readiness. We first selected a range of measures and evaluated them against these criteria with a large sample of pre-k children, then cross-validated the most promising measures with a new sample. The measures we selected for this investigation are described in the methods section that follows.

Methods

Direct Assessments of Cognitive Self-Regulation

To identify candidate measures, we first reviewed the literature on executive function, effortful control, attention, and self-regulation in an attempt to delineate the range of skills likely to be relevant to LRCSR. We then organized the selection of candidate measures to ensure that collectively they encompassed those skills. The skill domains distinguished for this purpose were the following:

(1) Sustained attention—attending to and sustaining focus on a task.

(2) Attention shifting—shifting focus within or between tasks as situations demand.

(3) Working memory—active maintenance and manipulation of information in memory.

(4) Inhibitory control—volitional inhibition of a prepotent response in order to complete a task.

(5) Effortful control—suppression of impulsive or premature responses when required by a task.

When selecting candidate measures we prioritized those previously shown to be related to academic achievement or gains in achievement and those most practical for administration in classrooms settings without the need for computer support or specialized equipment. Through this process we identified 10 candidate measures that yield 12 indices of CSR (two measures assess both accuracy and reaction time), which are described below.[1]

Sustained attention. Sustained attention, the capacity to attend to and maintain focus on a task, was assessed with the *Copy Design* task (Davie, Butler, & Goldstein, 1972; Osborn, Butler, & Morris, 1984) and the *Kansas Reflection-Impulsivity Scale for Preschoolers* (KRISP; Wright, 1971). For *Copy Design*, children copied eight geometric designs of increasing difficulty from a printed model. Children had two attempts to replicate each design and the quality of the best attempt was scored 0 or 1 according to defined criteria (e.g., should be approximately symmetrical; cannot be rotated). Total scores could range from 0 to 8 with higher scores indicating more accurate copies.

The KRISP assesses the ability attend to detail before making a response. Children were presented with a series of drawings, each with a target picture and four to six other pictures, all but one of which differed from the target in minor ways. They were asked to identify the picture that was a duplicate of the target and were allowed up to three errors before continuing to the next trial. Three practice trials with feedback were followed by 12 progressively more difficult test trials. Each trial was scored for number of errors (up to 3) and reaction time (RT) for the first drawing selected by the child. The final accuracy score was the number of errors on the 12 test items subtracted from the total errors possible (36). The final RT score was the difference between the mean RTs for the 5 hardest and the 7 easiest trials divided by the mean RTs for the

[1] Further information about these measures and how they are administered can be found at https://my.vanderbilt.edu/cogselfregulation/

hardest trials, thus indexing how much the child slowed down to reflect on the harder items.

Attention shifting. Attention shifting, the ability to shift focus appropriately from one task to another, was assessed using the *Dimensional Change Card Sort* (DCCS; Zelazo, 2006). Children were first asked to sort a set of cards according to one dimension (red vs. blue color), and then according to a different dimension (star vs. truck shape). If they were largely successful in making that switch, they were given a set of similar cards with a black border around some of them and asked to sort by color if the card had a border and by shape if it did not. The assessor first demonstrated color sorting with two test cards, asked the children where they would sort the cards, and provided feedback for incorrect responses. The children were then given six color sorting trials with the rule for the sort stated before each. If at least five of these were correct, the child was then asked to sort the same cards by shape in six additional trials; if not, the task was terminated. If at least five of the six shape trials were correct, the assessor explained and demonstrated the border sort task, asked the child what they were to do if the card had a border and if it did not, and provided feedback for incorrect responses. These children were then given 12 trials of the border sort with the card described (e.g., “here is one with a border”) and the rules restated before each. Children received a score of 0 if they did not pass the initial color sort, a 1 if they passed the color sort but not the shape sort, a 2 if they passed the shape sort, and a 3 if they also passed the border sort with correct responses on at least 9 of the 12 trials.

Working memory. Working memory, the ability to temporarily store and manage the information to carry out tasks, was assessed using the *Operation Span* (Blair & Willoughby, 2006) and *Backwards Digit Span* tasks (Davis & Pratt, 1996). For *Operation Span*, children were told that they would look at houses that have animals and colors in them, and that they should remember the animals. A practice item with feedback for an incorrect response was then followed by six test trials, two each with two, three, or four items to remember. The child was asked to label the colors and animals on the first display and then recall the animal in each house on a second display of empty houses. Each item was scored 0 for incorrect or no responses and 1 for a correct response, with the sum across all items as the final score (range 0 to 18).

Backwards Digit Span (Davis & Pratt, 1996) requires children to remember and manipulate (reverse) a series of numbers. An example was provided ("If I say 1, 3, you say 3, 1) with a two digit practice trial with feedback, followed by six test trials with an increasing number of digits (2, 3, 4). The task was terminated with the first incorrect response. One point was scored for each number recalled correctly in sequence (e.g., 587 recalled backwards as 875 would be scored as 1). The final score was the sum of digits correctly recalled across the practice and test trials (range 0 to 23). For children who could not pass the practice item, a final score of 1 was given for an incorrect numeric response and a 0 for a non-numeric response.

Inhibitory control. Inhibitory control, the ability to suppress a prepotent response when necessary to complete a task, was assessed with *Head-Toes-Knees-Shoulders* (HTKS; Ponitz et al., 2009), *Peg Tapping* (Diamond & Taylor, 1996), and *Spatial Conflict* (Blair & Willoughby, 2006). HTKS asks a child to respond to two oral prompts, "touch your head" and "touch your toes," by doing the opposite-- touching their heads when the assessor says "touch your toes" and vice versa. Six practice trials with feedback were followed by 10 test trials. If responses on five or more of these trials were correct, two new prompts were added—"touch your shoulders" and "touch your knees"—and the instructions were again reversed. Four practice trials with feedback were followed by 10 test trials scored 0 for an incorrect response, 1 for an incorrect motion that was corrected, and 2 for a correct response. Ponitz et al., (2009) summed over the 20 test items (range 0 to 40). We also included the six practice items in the total score (range 0 to 52); we found this provided more differentiation in the lower range and somewhat stronger relationships with the other measures of CSR. The two scoring methods were highly related (r = .98 at the beginning of pre-k and .99 at the end).

The *Peg Tapping* task asked children to tap once when the examiner tapped twice and twice when the examiner tapped once (Diamond & Taylor, 1996). After two practice trials with feedback, children had eight more practice trials. If successful, they then had 16 test trials with no feedback; if not, the task was terminated. Test trials were scored 0 for incorrect responses, 1 for correct responses, and -1 if the task was aborted. Final scores ranged from -1 to 16.

The *Spatial Conflict* task (Blair & Willoughby, 2006) was a paper adaptation of the computer-based version (Gerardi-Caulton, 2000). Children were given a card with two black buttons (circles), one on the right-hand side and one on the left, and shown a series of arrows that pointed either left or right. They were asked to touch the button on the side the arrow pointed to, touching the button on the right with the right hand and the one on the left with the left hand. The assessor demonstrated the task twice and the children then completed four practice trials with feedback. This was followed by 14 congruent trials (arrow on the same side of the page it pointed to), then 16 trials with 4 congruent trials mixed in with 14 incongruent trials (arrow position on opposite side of the page it pointed to). Each trial was scored 0 for the incorrect button, 1 for the correct button with the wrong hand, and 2 for the correct button with the correct hand. Previous studies (e.g., Carlson, 2005) scored only the last 16 mixed trials (range 0 to 32). However, we scored all the trials (range 0 to 72) because we found this provided more variability and a final score with a somewhat stronger relationship to the other measures of CSR. These scoring methods were highly related (r = .88 at the beginning of pre-k, .90 at the end).

Effortful control. Effortful control, the ability to suppress impulsive or premature responses, was assessed with the *Whisper* and *Turtle-Rabbit* tasks (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). In the *Whisper* task children were shown pictures of 12 cartoon characters and asked to whisper their names, or whisper that they did not know, and the assessor whispered throughout the task to model that behavior. The cartoon characters varied in familiarity for most pre-k children, providing the opportunity for the child to act impulsively (shout) when a very recognizable one came up. Each trial was scored 0 for a shout, 1 for a normal or mixed voice, 2 for no response, and 3 for a whisper (range 0 to 36).

The *Turtle-Rabbit* task (Kochanska et al., 1996) presented children with a mat on which a curving path with five bends was drawn, and they were then asked to move a toy figure from the beginning to the end without straying off the path. Children completed two baseline trials with a boy and a girl token (neutral condition), then two trials with a rabbit they were told was "fast" to stimulate an impulse to get to the finish line quickly, and two with a turtle they were told was "slow." The trials were scored for accuracy and time. Each curve was scored 0 if the child bypassed it, 1 if the figure was above the mat but followed the general curvature, and 2 if the figure stayed on the mat and within the path. Also, the time to complete each trial was recorded to the nearest second. The final accuracy score was the total for all curves and all trials (range 0 to 60). For the reaction time score, Kochanska et al., (1996) used the difference between the means of the turtle and rabbit trials. However, we used the ratio of that difference to the mean turtle time, thus representing the faster time of the rabbit in proportion to the turtle time rather than as a simple difference. This version provided a final time score with more variability and a somewhat stronger relationship to the other measures of CSR. The correlations between the two scoring methods were moderately large (r = .68 at the beginning of pre-k; .53 at the end).

Teacher Ratings of Cognitive Self-Regulation

Teacher rating scales for children's behaviors in the classroom were selected to mirror as much as possible the aspects of CSR identified in our initial literature review and assessed in the candidate direct child measures. The following subscales were combined in a single rating form.

Persistence. The Persistence subscale of the *Temperament Assessment Battery for Children* (TABC; Martin, 1988) was used to obtain teachers' assessment of each child's ability to sustain attention. The eight items on this subscale are rated on a 1 to 7 Likert scale (1 = hardly ever, 7 = almost always) and include such behaviors as "child can continue at the same activity for an hour" and "if child's activity is interrupted, he/she tries to go back to the activity."

Distractibility. The Distractibility subscale of the TABC was used to obtain teachers' assessment of each child's ability to ignore distractions. The eight items on this subscale are also rated from 1 (hardly ever) to 7 (almost always) and cover such behaviors as "child is easily drawn away from his/her work by noises, something outside the window, another child's whispering, etc." and "if other children are talking or making noise while teacher is explaining a lesson, this child remains attentive to the teacher."

Impulsivity. This dimension was assessed with the Impulsivity subscale of the *Children's Behavioral Questionnaire* (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001). CBQ items are rated on a 1 to 7 Likert scale (l = extremely untrue of student, 7 = extremely true). The 13 items on this subscale cover such behavior as "sometimes interrupts others when they are speaking" and "usually stops and thinks things over before deciding to do something."

Attention shifting. The CBQ Attention Shifting subscale was used for this dimension. The twelve items on this subscale were also rated from 1 (extremely untrue of student) to 7 (extremely true) and covered such behaviors as "needs to complete one activity before being asked to start on another one" and "can easily shift from one activity to another."

Work-Related Skills. An additional scale that spanned a variety of children's CSR skills as manifested in the classroom was included in the teacher rating form—the Work-Related Skills subscale of the *Cooper-Farran Behavior Rating Scale* (CFBR; Cooper & Farran, 1988). The 16 items on this scale ask about children's independent work, compliance with instructions, memory for instructions, and completion of games and activities. These items are rated from 1 to 7 using behavioral anchors distinctive to each item.

Academic Achievement Measures

Academic achievement was measured with five subscales from the *Woodcock Johnson III* achievement battery (Woodcock, McGrew, Mather, 2001). These included two math subtests: Applied Problems, which assesses children's ability to solve numerical and spatial problems presented verbally with accompanying pictures, and Quantitative Concepts, which assesses children's knowledge of numbers, sequencing, shapes, symbols, and the like. Language and literacy skills were assessed with Letter-Word Identification, Picture Vocabulary, and Oral Comprehension subtests. Letter-Word Identification asks children to identify and pronounce alphabet letters and read words. For Picture Vocabulary, children name objects presented in pictures and point to the picture that goes with a word. Oral Comprehension assesses children's ability to complete an orally presented passage by providing the appropriate missing word using semantic and syntactic cues. Data analysis was conducted with the IRT-scaled W-scores, but standard scores (mean of 100, standard deviation of 15) are more descriptive and showed fall pre-k baseline mean values for the analytic sample of 98 on Applied Problems, 90 on Quantitative Concepts, 104 on Letter-Word Identification, 100 on Picture Vocabulary, and 97 on Oral Comprehension.

Participants and Assessment Procedure

Parental consent was obtained for 608 children recruited from 58 pre-k classrooms in 38 schools/centers across 4 school systems and several community childcare centers for low-income families. The consent rate was approximately 60% across all classrooms, ranging from 13% to 100%. Consented children identified by their teachers as English Language Learners were screened for English proficiency using the *Pre-LAS* (Duncan & DeAvila, 1985). Thirty-six children did not pass the *Pre-LAS*, five children did not provide assent, and 32 children moved during the course of the study, leaving 535 in the final analytic sample.

The participating schools/centers were in urban, suburban, and rural settings and provided an ethnically and economically diverse sample of children. Racial diversity in these schools/centers ranged from 0% to 87% African American (*M* = 16%), 2% to 34% Hispanic (*M* = 11%), and 13% to 95% non-Hispanic white (*M* = 71%). Economic diversity ranged from 16% to 100% of the children qualifying for free or reduced lunch programs (*M* = 55%). The mean age of the 535 children in the analytic sample at the first assessment session was 4.6 years, ranging from 3.8 to 5.4, and 52% were male.

Procedure**.** Children were assessed twice during the pre-k year– near the beginning of the school year (early September through October) and near the end (mid-March to early May), further referred to as Time 1 and Time 2, respectively. They were then assessed again at the end of kindergarten (mid-March to early May), referred to as Time 3. The Time 1 and Time 2 assessments were administered in three testing sessions of 20-30 minutes each with nearly all sessions occurring within a period of fewer than 10 weeks. The Time 3 assessments were administered in two sessions spanning fewer than five days on average. Each child was assessed individually in a quiet area away from the classroom with a varying order for the sessions but a fixed order for the measures within a session. In pre-k, the sessions included (a) Operation Span, Whisper, Peg Tapping, and WJ-III Applied Problems and Quantitative Concepts; (b) DCCS, HTKS, Digit Span, Copy Design, and WJ-III Picture Vocabulary; and (c) Spatial Conflict, Turtle-Rabbit, KRISP, and WJ-III Letter-Word Identification and Oral Comprehension. Based on the findings from pre-k, a reduced set of measures was administered in the two sessions at the end of kindergarten: (a) Peg Tapping, HTKS, Copy Design, and WJ-III Applied Problems, Quantitative Concepts, and Picture Vocabulary; and (b) DCCS, KRISP, Digit Span, and WJ-III Letter-Word Identification and Oral Comprehension.

Teacher ratings were made at approximately the same times as the child assessments. These thus occurred near the beginning, and again near the end of the pre-k year. Kindergarten teachers then completed the same rating scales near the end of the kindergarten year.

Missing data. Of the 535 children who comprised the initial pre-k analytic sample, 47 could not be located for the Time 3 end of kindergarten assessments, leaving 488 children in the follow-up sample. The children missing Time 3 data were compared with those providing data on the available demographic variables and the T1 and T2 CSR and achievement measures. T-tests with Benjamini-Hochberg corrections for the large number of multiple comparisons showed no significant differences between children assessed and not assessed in kindergarten. Given no indications that the missing cases made the follow-up sample unrepresentative of the initial sample, analyses with pre-k data were conducted on the analytic sample of 535 children while those with kindergarten data were conducted on the sample of 488.

Cross-validation sample and assessment procedure. The cross-validation sample was drawn from a later cohort of children enrolled in pre-k in the four school systems that provided most of the original sample. These children were assessed three times during the pre-k year—

near the beginning (Time 1), approximately two weeks later (Retest) to assess the test-retest reliability of the measures, and near the end of the school year (Time 2). Parental consent was obtained for 593 children from 43 classrooms in 23 schools (overall consent rate of 69%). Of these, 416 were randomly selected, but 21 did not pass the *Pre-LAS* screen for English language proficiency, 4 could not complete the study tasks, 18 moved prior to the reliability retest, and 4 were withdrawn due to assessor error. This left 369 children in the sample for the test-retest reliability data collected in the fall of the pre-k year. After that, 13 children moved before the end of pre-k, leaving 356 in the sample with data from both the beginning and end of the pre-k year.

The mean age of the children in both the test-retest and final samples was 4.4 years and 53% were male. As in the initial sample, the schools from which these children were drawn were economically and racially diverse: the proportion of students at each school qualifying for free or reduced price lunch ranged from 26% to 95% (M = 52%); the proportion who were African American ranged from 0% to 49% (M = 12%), the proportion Hispanic ranged from 1% to 38% (M = 9%), and the proportion non-Hispanic white ranged from 33% to 97% (M = 75%).

At Times 1 and 2, there were two assessment sessions, one for CSR and one for achievement. The order of these sessions was varied, but the measures were administered in a fixed order at each session (described later). Only CSR measures were administered at Retest. In addition, at Time 1, Retest, and Time 2, teachers completed ratings on selected CSR measures (described later). Most of these teachers had also participated in the initial phase of this study, but some were new.

Analytic Approach

The overall objective of the analysis was to determine which of the 12 candidate direct child assessment measures performed best as indicators of LRCSR for pre-k children. To accomplish this, a series of analyses was conducted to investigate the extent to which each measure met the criteria we deemed important for this purpose. In particular, we sought evidence that each measure:

(1) Represents a facet of CSR. Exploratory factor analysis examined the extent to which each measure loaded on the common factor representing the latent CSR construct the measures were expected to reflect.

(2) Captures developmental change. Gains in task performance on each measure from the beginning to end of pre-k were assessed for magnitude and statistical significance.

(3) Demonstrates convergent validity with teachers' ratings of CSR in the classroom. Correlations between each measure and teacher ratings were examined.

(4) Demonstrates predictive validity for subsequent academic achievement. Regression analyses examined the ability of each measure, and gain on each measure over the pre-k year, to predict achievement scores at the end of pre-k and kindergarten and achievement gains from the beginning of pre-k to each of those end points.

In the cross-validation phase we then examined key features of the measures that best met the above criteria with data from the new sample. Construct validity was investigated with a confirmatory factor analysis, and analyses of convergent and predictive validity were repeated to assess the stability of the initial results. The test-retest reliability coefficients for the selected measures were also estimated with data from the cross-validation sample.

Results

Relationships to a Common Factor

Principal factor analyses were conducted separately for the CSR assessments obtained from the children at the beginning (Time 1) and end of pre-k (Time 2). The one-factor solutions accounted for 31.3% of the total variance at Time 1 and 30.1% at Time 2. The factor loadings (see Table 1) varied, but none was smaller than 0.31. Three measures had loadings greater than .60 at both time points– HTKS, KRISP Accuracy, and Peg Tapping. Two others– Copy Design and DCCS– had loadings greater than .50 at both time points. These measures were thus the ones that most strongly exemplified the underlying CSR construct our original selection of measures was intended to represent.

A possible artifact in this analysis relates to how well children understood the directions for the various tasks; poor performance could represent failure to understand the task rather than inability to accomplish it. As an alternative perspective on the relationships of these measures to the presumed common factor, we repeated the factor analyses with scores that were adjusted for each child's performance on the WJ-III Oral Comprehension scale. As shown in the last columns of Table 1, the correlations between Oral Comprehension and the CSR measures ranged from .12 to .52 and were highest for some of the measures with the largest factor loadings. The adjusted scores were generated by predicting each CSR score from the Oral Comprehension W-scores with ordinary least squares regressions. The residuals from those regressions, that is, the portions of the children's scores that could not be predicted from their Oral Comprehension scores, were then used as the adjusted scores for the alternate factor analyses. Note that this is a conservative procedure. Some of the shared variance between scores on the CSR tasks and Oral Comprehension is almost certainly due to the influence of general cognitive development on both measures. With all the variance on the CSR tasks associated with oral comprehension stripped out, the adjusted scores measure very narrowly restricted skills on the CSR tasks.

The two middle columns in Table 1 report the results of the factor analyses with the adjusted scores. Though the loadings are smaller, as expected because of the reduced reliability of the residualized scores, the pattern is similar to that found with the original observed values.

The largest loadings at both times appear for Copy Design, HTKS, KRISP Accuracy, and Peg Tapping. The greatest decrease in the factor loadings with adjusted scores is for DCCS. Though hardly definitive, this suggests that children's performance on the DCCS may be more dependent on the ability to understand the directions and general cognitive ability, and less on specific CSR ability, than the other measures.

Developmental Change

Children's scores on each of the candidate CSR measures at the beginning of pre-k were compared to their scores at the end of the year to assess change over that period. These analyses were conducted with multilevel models in which a dummy code for time predicted each CSR score with Time 1 and Time 2 scores nested within children and children nested within classrooms and schools. As Table 2 shows, performance was significantly better at Time 2 for all the measures except Turtle-Rabbit Accuracy. Pre-post effect sizes for the gains on the other measures were positive and ranged from .31 to .69, with the greatest gains for Copy Design, DCCS, HTKS, KRISP Accuracy, and Peg Tapping (effect sizes greater than 0.50).

Table 2 also shows the correlations between children's scores at the beginning and end of pre-k. These were all statistically significant and ranged from .12 to .66. The largest of them showed reasonable consistency in children's relative ranking over the school year. Nevertheless, they are not so large as to indicate that only stable individual differences are measured with no room for influence from differential experiences in and outside the classroom during this period.

Convergent Validity with Teacher Ratings

To investigate the convergent validity with teacher's ratings of the CSR measures, we examined the correlations between each measure and each of the five teacher rating scales (CFBR Work Related Skills, TABC Distractibility, TABC Persistence, CBQ Attention Shifting, and CBQ Impulsivity) and a composite scale created by summing the z-scores computed for each scale. The correlations of each CSR measure with this composite and with each of the individual teacher rating scales are reported in Table 3 for the beginning and end of pre-k.

As Table 3 shows, all these correlations were statistically significant except for a few involving CBQ Impulsivity. The largest correlations with the Teacher Rating Composite appeared for Peg Tapping, HTKS, and KRISP Accuracy (.34 to .42). Close behind were Copy Design, DCCS, and Turtle-Rabbit Accuracy with correlations of at least .25. Moreover, these correlations were substantially similar for the ratings at the beginning and end of pre-k. The correlations with the individual teacher rating scales showed similar patterns, though they were generally much lower for the CBQ scales.

Predictive Validity for Academic Achievement

The most important consideration for our purposes in assessing the CSR measures was their predictive validity for academic achievement, measured here with the WJ-III Quantitative Concepts, Applied Problems, Oral Comprehension, Picture Vocabulary, and Letter-Word Identification subtests. The intercorrelations among these five subtests at Times 1, 2, and 3 were positive and relatively high, and principal components analyses showed strong one-factor solutions with loadings from .61 to .84. To represent overall academic achievement, therefore, we created a composite score for each time of measurement by combining the W-scores across the five subscales for each child with each subtest given equal weight.

CSR predicting achievement. One aspect of predictive validity is the ability of the CSR measures to predict later academic achievement. Table 4 reports the predictive correlations between each CSR measure and achievement measured later. In each instance, the correlation is computed as the standardized regression coefficient for the CSR measure predicting the respective achievement measure in a multilevel model that takes into account the nesting of children within classrooms and classrooms within schools when estimating the standard errors.

Table 4 shows that all these correlations are statistically significant. For those representing predictions from Time 1 or Time 2 to later achievement, the largest correlations were found for Backwards Digit Span, Copy Design, DCCS, HTKS, KRISP Accuracy, and Peg Tapping. From the beginning of pre-k to the end of pre-k (Time 2) and then to the end of kindergarten (Time 3), these correlations ranged from .37 to .56. From the end of pre-k (Time 2) to the end of kindergarten, they ranged from .38 to .57. Thus whether measured at the beginning or end of pre-k, some CSR measures showed relatively strong predictive validity for later achievement.

CSR predicting achievement gain. A more demanding test of the predictive validity of the CSR measures is their ability to predict the *gain* children make in academic achievement over a subsequent period. This was assessed with standardized regression coefficients from multilevel models predicting later achievement from each CSR measure and initial achievement. The CSR measures, therefore, were predicting residualized gain in achievement—the parts of the later achievement scores that could not be predicted from the initial levels on the respective achievement measures. This gain measure is thus an indication of what children learned during the measurement interval. Note that by controlling for the initial score on the achievement composite (which includes language measures), this analysis controls for children's language ability, thus helping disentangle language ability from the influence of the CSR measures.

The first four columns of Table 5 show standardized regression coefficients from these analyses. It is not surprising that they are relatively small given the strong relationship between initial and later achievement and the inherent unreliability of gain scores. Nonetheless, many of the CSR measures had statistically significant predictive relationships from the beginning to the end of pre-k and the end of kindergarten, as well as from the end of pre-k to the end of kindergarten. The five measures with significant predictive relationships for all three intervals, at least at $p < .10$, were Backwards Digit Span, Copy Design, HTKS, KRISP Accuracy, and Peg Tapping. Table 5 also shows the predictive relationships between CSR at the beginning of pre-k and achievement gain between the end of pre-k and end of kindergarten. Except for Digit Span, those relationships were also significant for these five CSR measures.

CSR gain predicting achievement gain. To further probe the predictive validity of the CSR measures, we asked whether CSR gains during the pre-k year, and between the end of pre-k and end of kindergarten, were correlated with the achievement gains made over those same periods. That is, if a child showed gain on one of the CSR measures, was there an associated gain in achievement? For this gain-with-gain analysis, we first estimated gain for each CSR measure over the respective periods by predicting later CSR scores from the initial values on the same measure and saving the residuals. Those residualized gain scores were then used as independent variables in multilevel regression analyses to predict later achievement with initial achievement controlled.

The last three columns of Table 5 report the standardized regression coefficients from these analyses. Here also the coefficients are relatively small because the much larger relationships between pre-post CSR and pre-post achievement have been controlled and because of the generally low reliability of gain scores. The relationships of CSR gain during pre-k with achievement gain during that year, and with achievement gain between the beginning of pre-k and the end of kindergarten, are nonetheless statistically significant for many of the CSR measures. The better performing CSR measures across these various intervals were Backwards

Digit Span, Copy Design, DCCS, HTKS, KRISP Accuracy, and Peg Tapping.

Summary of Findings on the Selected Criteria

Table 6 summarizes the findings reported above by identifying the CSR measures that were the top performers in each analysis. The measures are listed with the better performing ones first rather than in alphabetical order as in the previous tables. Four CSR measures were among the top performers in every analysis: Copy Design, HTKS, KRISP Accuracy, and Peg Tapping. DCCS was very close behind, appearing in the top performing group in all but one analysis. Consideration must also be given to Backwards Digit Span, which showed good performance in the predictive validity analyses, though it was not among the top performers in the other analyses. The most notable feature of this summary is the great consistency of the CSR measures that performed well in these analyses. For the most part, those that were strong in one analysis were strong in all or nearly all of them, and those weak in any one analysis were weak in all or nearly all.

Validity of the Top CSR Measures in Combination

The final series of analyses examined the convergent and predictive validity of the six top performing CSR measures in combination. For that purpose, another series of multilevel regression analyses was conducted with all six of these measures used together as predictors. To examine their collective performance, multiple correlations were estimated for their relationship to the different dependent variables of interest. This was done by fitting unconditional models with the six measures omitted, followed by conditional models in which they were included. The proportion of the total variance identified in the unconditional models accounted for in the conditional models was determined (R-squared) and the square root of that estimate was taken as the multiple correlation of interest. In addition, the standardized regression coefficient for each

CSR measure in the set of predictors indicated the independent contribution of that measure to predicting the respective dependent variable.

The results of these analyses are summarized in Table 7. The first panel shows the multiple correlations that index the convergent validity of the set of six CSR measures with the composite teacher ratings at the beginning (T1) and end (T2) of pre-k. Those multiple correlations (.49 and .50, respectively) can be compared with the analogous correlations for each individual CSR measure reported in the first two columns of Table 3, all of which are smaller. The standardized regression coefficients in Table 7, in turn, indicate that Peg Tapping, KRISP Accuracy, and HTKS made the strongest independent contributions to those relationships. These are also the measures with the largest individual correlations in Table 3.

The second panel of Table 7 reports the collective relationship of the six CSR measures to composite achievement measured later, and thus addresses predictive validity. The multiple correlations, ranging from .68 to .72, can be compared with the analogous correlations for the individual measures shown in Table 4, all of which are smaller. The standardized regression coefficients indicate that the strongest independent contributions were made by HTKS, KRISP Accuracy, Peg Tapping, and Backwards Digit Span.

The third panel of Table 7 provides the results for the six CSR measures collectively predicting achievement gain over various periods. The multiple correlations, ranging from .23 to .28, can be compared with the standardized regression coefficients in the first four columns of Table 5, all of which are notably smaller. The regression coefficients in Table 7 indicate that KRISP Accuracy and HTKS have the strongest independent relationships to achievement gain, followed by Copy Design and Backwards Digit Span.

The fourth and final panel in Table 7 reports the results for the most important predictive validity relationships—those between gain on the CSR measures and achievement gain. The multiple correlations, which can be compared with the smaller standardized regression coefficients for the individual measures in the last three columns of Table 5, ranged from .32 to .39. For this aspect of predictive validity, the individual measures making the strongest independent contributions were Backwards Digit Span and Peg Tapping.

In general, the results in Table 7 show that a combination of the six top performing individual CSR measures has greater convergent and predictive validity than any single measure. Moreover, in most instances the improvement in the magnitude of the respective relationships is great enough to show that a composite of these measures holds more promise as a general measure of LRCSR for pre-k children than any one of them used alone. With a primary emphasis on convergent validity with teacher ratings and predictive validity for achievement gain, the overall strongest independent contributions were made by Peg Tapping, KRISP Accuracy, and HTKS. In addition, Backwards Digit Span had an especially strong influence in the relationship between CSR gain and achievement gain.

Cross-Validation

As described above, six of the candidate CSR child assessments performed well on our criteria for measures of LRCSR. The large number of analyses conducted to identify those six, however, allow ample opportunity for chance factors in the particular sample of children and the data they provided to influence the results. In a follow-up cross-validation study, therefore, we administered those six measures[2] to a new sample of children to check the stability of the key features that favored them in the initial analyses. We also used this new sample to examine the

---

[2] Scores for the Backward Digit Span measure in the cross-validation reflect the longest span correctly recalled (range = 1 to 8) based on administration procedures from the Wechsler Intelligence Score for Children – Forth Edition (Wechsler, 2003). For the KRISP, we added more advanced items from version B to provide a better ceiling (maximum score of 48). Scoring was altered for Copy Design; every attempt was scored of the two allowed for each item, making the scores range from 0 to 16. The other cognitive self-regulation measures were the same as before.

test-retest reliability of the selected measures.

The selected CSR measures and the same WJ-III measures used in the initial phase were administered in two sessions at the beginning (Time 1) and end (Time 2) of the pre-k year. The order of these sessions was varied, but the measures were administered in a fixed order at each session (for CSR this was: Peg Tapping, KRISP, HTKS, DCCS, Backwards Digit Span, and Copy Design). The CSR measures were administered a second time approximately two and a half weeks after the assessment sessions at the beginning of the year to allow estimation of test-retest reliability. In addition, at Time 1, Retest, and Time 2, teachers completed ratings on the 20 items that had the largest correlations with the child measures in the initial phase: 10 items from the CFBR Work-Related Skills scale, 3 items from the CBQ, and 7 items from the TABC.

Construct validity. Confirmatory factor analyses in Mplus 7 were conducted with Time 1 and Time 2 data to test the assumption that, as in the initial analysis, the CSR measures have strong relationships with a single common underlying construct. Robust maximum likelihood estimation was used to account for any non-normality in the data distributions and adjust the standard errors for the nesting of children within classrooms and schools. The fit statistics for both Time 1 and Time 2 data confirmed a one factor-solution [$\chi^2(9)$ tests of model fit: 32.9, $p < .001$ and 28.0, $p < .001$; RMSEA of .09 and .08, respectively]. Moreover, the factor loadings were statistically significant at $p < .001$ for all six measures at both times. The largest loadings were found for Peg Tapping (.73, .79) and HTKS (.75, .72); the smallest for KRISP (.50, .46) and Copy Design (.55, .46); and those for DCCS (.61, .52) and Backwards Digit Span (.55, .57) fell in between.

Test-retest reliability. The mean interval between the CSR assessments for the 369 children in the test-retest sample was 16.7 days (SD=5.0). Test-retest reliability coefficients were estimated using multilevel regression to account for the effect of the nesting of children within classrooms and schools on the standard errors. For each measure, the initial score was used to predict the retest score with the standardized regression coefficients then representing test-retest correlations. In descending order, those reliability coefficients and their standard errors were as follows: HTKS .80 (.03); Peg Tapping .80 (.03); Backwards Digit Span .73 (.04); Copy Design .72 (.04); KRISP Accuracy .64 (.04); and DCCS .47 (.05). The KRISP reliability coefficient is modest and that for DCCS is marginal, but the others are in a generally acceptable range. Test-retest reliability was also estimated for a composite of all six measures using the factor scores from the factor analyses described above, yielding a reliability coefficient of .89 (.02).

***Convergent validity*.** Factor analysis and the convergent validity analysis of the 57 teacher rating items used in the initial phase of the study identified 20 of those items as a sufficient set for representing teachers' assessments while maintaining alignment with the direct assessment measures. These 20 teacher rating items were then used for the check on convergent validity in the cross-validation phase of the study. These items were all rated on 7-point scales and they showed a high level of internal consistency (Cronbach alpha values of .98 at both Time 1 and 2). A total score was computed as the mean of the 20 items and used as the dependent variable in multilevel regression models with each CSR direct assessment measure in turn as the sole independent variable. The standardized regression coefficients that represent the correlations between each CSR measure and the teacher rating total score are reported in Table 8. They range from .27 to .47 and all are statistically significant. The largest correlations were found for Copy Design, Peg Tapping, HTKS, and KRISP Accuracy; those for Backwards Digit Span and DCCS were notably smaller. Compared with the analogous values from the initial sample shown in the first two columns of Table 3, all but two of these correlations are larger and those two are close to the prior values. For a broader view, Table 8 also reports the correlation between the total teacher rating scores at Times 1 and 2 and the factor score from the confirmatory factor analysis of the six child measures described earlier. Those correlations were .63 and .60 respectively, and demonstrate the greater convergent validity of a composite of the six child CSR measures.

Predictive validity. To assess predictive validity in the cross-validation sample, we first examined the correlations between the CSR measures at Time 1 and the composite achievement score at Time 2 (column 1, Table 9). As with the initial sample, these were estimated with standardized regression coefficients in multilevel models. These coefficients were statistically significant and very similar to those found in the initial sample (column 1, Table 4).

As with the initial sample, the ability of each CSR measure to predict the gain children made in achievement over the pre-k year was assessed with standardized regression coefficients from multilevel models in which Time 1 achievement was controlled. These coefficients were statistically significant for DCCS, HTKS, KRISP Accuracy, and Peg Tapping (column 2, Table 9), but showed some modest inconsistencies with the initial sample results (column 1, Table 5) for Copy Design (.05 vs .12), DCCS (.11 vs. .07), and HTKS (.07 vs. .11). The predictive validity coefficients for the more revealing relationships between gains on the CSR measures and gains in achievement over the pre-k year (Time 1 to Time 2) were statistically significant for all the measures (column 3 in Table 9). The strongest relationships were for Peg Tapping, DCCS, Copy Design, and KRISP Accuracy but here also there were some modest inconsistencies with the estimates from the original sample (column 5, Table 5) for some measures, specifically Backwards Digit Span (.06 vs. .12), Copy Design (.10 vs. .07), and Peg Tapping (.15 vs. .11). The predictive validity coefficients for the factor score from the confirmatory factor analysis that combined all six CSR measures are also shown in Table 9 and again demonstrate that the combined set of items performs better than any one item.

The final series of predictive validity analyses investigated the independent contribution of each of the six CSR measures relative to the others when they were all used simultaneously as independent variables in multilevel regressions predicting the various achievement outcomes. The results are reported in Table 10 and can be compared with the analogous values from the initial sample in Table 7. Across all the outcomes, Peg Tapping, DCCS, and KRISP Accuracy showed the largest independent relationships to later achievement or achievement gain and these were the only three CSR measures for which the coefficients were statistically significant with every outcome. However, Copy Design showed a significant independent gain-with-gain relationship and HTKS and Backwards Digit Span showed significant independent contributions to predicting Time 2 Achievement. Comparing these results with the analogous ones for the initial sample (Table 7), the coefficients that were most similar in terms of statistical significance and magnitude across all the achievement outcomes were for KRISP Accuracy; Peg Tapping, DCCS, and HTKS also showed relatively good replication.

Conclusions

The objective of this study was to identify direct assessment measures of CSR for pre-k aged children that are indicative of their ability to engage in the learning opportunities available in the classroom and, as such, are predictive of their later academic achievement, especially their achievement gains. We have referred to the construct indexed by these measures as *learning-related cognitive self-regulation* (LRCSR). Such measures are useful for screening pre-k children to assess their readiness for kindergarten and for further research on the factors and interventions that influence LRCSR.

In pursuing this objective, we did not attempt to develop new measures but, rather, focused on assessing existing measures in a comparative fashion to determine which were better for the intended purposes. Many potentially relevant measures have been created in the context of research on attention, executive function, effortful control, and the like, not all of which could be examined in this study. We selected the measures to consider by reviewing the pertinent research literature, delineating the potentially relevant components of CSR, and choosing candidate measures that encompassed those components. In that selection, we favored measures that were relatively well known, could be used in the classroom without requiring computer support or specialized equipment, could be administered in a relatively brief period, and, when possible, had already demonstrated a relationship to academic achievement in prior research. Of the 12 measures selected, analyses with an initial sample of 535 pre-k children identified six that performed well against our criteria. Cross-validation with a new sample of 356 children explored the stability of those findings.

The two most important criteria for identifying good LRCSR measures were, first, convergent validity with teacher ratings of children's CSR in the classroom and, second, predictive validity for measures of children's academic achievement. Convergent validity in this context reflects the ecological validity of the measures—their relevance to what teachers observe in classroom settings. Of equal importance, predictive validity for academic achievement, especially later achievement gains, provides assurance that the LRCSR measures were, in fact, learning-related. By these and other secondary criteria, the best performing measures were Copy Design, HTKS, KRISP Accuracy, Peg Tapping, DCCS, and Backwards Digit Span. The single best performing measure across all our analyses was Peg Tapping, with the functionally similar HTKS close behind and KRISP Accuracy in third place.

However, each of the six measures had a strong showing in some analyses and no single measure did nearly as well as a composite of all six measures. Our procedure for administering those six measures with the cross-validation sample demonstrated that it was feasible to include them all in a single assessment session of 35-45 minutes. It might be tempting to shorten the battery by omitting Copy Design and Backwards Digit Span, but analyses not reported here across both samples show that this produced a significant decrement in the performance of the composite for predicting later achievement and achievement gains. The six measures are scored on quite different scales, complicating the integration of them into a single composite measure. For research purposes, computing z-scores for each, then summing them provides one solution; using factor scores from a factor analysis of the six measures provides another. For general use, each measure can be rescaled into a 0 to 5 point format with all six then summed to create a simple additive total score that works well. The appendix to this paper describes the rationale, procedure, and results of this rescaling process.

Since our own work began in this area, the National Institutes of Health has issued its Tool Box, which includes some measures of self-regulation similar to the ones found to perform well in this study (Weintraub et al., 2013; Zelazo et al., 2013). Currently the Tool Box requires connection to the Internet in order to access those IRT scaled instruments. Other researchers have focused on a single self-regulation measure such as the work by McClelland and colleagues with HTKS (Schmitt, Pratt, & McClelland, 2014). In contrast, the work reported in this paper is focused on assessment procedures that can be readily used without computer support and reveals the advantages of using a battery of measures, the combination of which performs better than any single measure.

The fact that the LRCSR measures identified here are predictive of later achievement and achievement gains, of course, does not mean that they represent causal factors for those outcomes. However, with validated measures in hand, a key question for future research can be further investigated—whether certain practical interventions or particular teacher practices in preschool are capable of increasing children's LRCSR skills. There is some evidence using one or another of the measures identified here that this might be possible (e.g., Biermn, Nix, Greeberg, Blair, & Domitrovich, 2008; Raver et al., 2011), but also some less encouraging findings (e.g., Barnett et al., 2008; Clements, Sarama, Unlu, & Layzer, 2012). Assuming that LRCSR can be boosted, an even more important question is whether doing so for pre-k children, in fact, will lead to greater learning and increased academic achievement.

References

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology, 47*, 905-915. doi: 10.1037/a0023748

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York, NY: Academic Press.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., & Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, *23*, 299-313. doi: 10.1016/j.ecresq.2008.03.001

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, *3*, 821-843. doi: 10.1017/S0954579408000394

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111-127. doi: 10.1037//0003-066X.57.2.111

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647-663. doi: 10.1111/j.1467-8624.2007.01019.x

Blair, C., & Willoughby, M. T. (2006). *Measuring Executive Function in Young Children.* Chapel Hill, NC: The Pennsylvania State University and The University of North Carolina at Chapel Hill.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal*, *108*, 115-130. doi: 10.1086/525550

Bronson, M. B., Tivnan, T., & Seppanen, P. S. (1995). Relations between teacher and classroom activity variables and the classroom behaviors of prekindergarten children in Chapter 1 funded programs. *Journal of Applied Developmental Psychology, 16*, 253-282. doi:1016/0193-3973(95)90035-7

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616. doi: 10.1207/s15326942dn2802_3

Clements, D. H., Sarama, J., Unlu, F., & Layzer, C. (2012, March). *The efficacy of an intervention synthesizing scaffolding designed to promote self-regulation with an early mathematics curriculum: Effects of executive function.* Paper presented at the meeting of Society for Research on Educational Effectiveness, Washington, D.C.

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk in kindergarten. *Early Childhood Research Quarterly, 3*, 1-20. doi:10.1016/0885-2006(88)90026-9

Cooper, D. H., & Farran, D. C. (1991). *The Cooper-Farran Behavioral Rating Scales*. Brandon, VT: Clinical Psychology Publishing Co., Inc.

Cooper, D. H., & Speece, D.L. (1988). A novel methodology for the study of children at risk for school failure. *The Journal of Special Education, 22*, 186–198. doi:10.1177/002246698802200205

Davie, R., Butler, N. R. & Goldstein, H. (1972). *From birth to seven: The second report of the National Child Development Study (1958 Cohort).* London: Longman and the National Children's Bureau.

Davis H. L., Pratt C. (1996). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47,* 25–31. doi:10.1080/00049539508258765

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608,* 637-676. doi: 10.1111/j.1749-6632.1990.tb48913.x

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the ability to remember what I said and to "do as I say, not as I do." *Developmental Psychobiology, 29*, 315-334. doi: 0012-1630/96/040315-20

Duncan, S. E. and DeAvila, E. A. (1985). *How to Administer Pre-LAS*. Monterey,CA: CTB/McGraw-Hill.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446. doi: 10.1037/0012-1649.43.6.1428

Gathercole, S., Brown, L., & Pickering, S. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology*, *20*, 109-122.

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24-36 months of age. *Developmental Science, 3,* 397-404. doi:10.1111/1467-7687.00134

Kagan, S. L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Washington, DC: National Education Goals Panel.

Katz, L. G. (1993). Dispositions: Definitions and implications for early childhood practices.

Perspectives from ERIC/EECE: A Monograph Series, No. 4. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.

Katz, L. G. (2002). 'Not all dispositions are desirable': Implications for assessment. *Assessment in Education: Principles, Policy & Practice*, *9,* 53-54. doi: 10.1080/09695940220119175

Kochanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490-507. doi:10.1111/j.1467-8624.1996.tb01747.x

Kochanska, G., Murray, K. T., & Harlan, E.T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220-232. doi: 10.1037/0012-1649.36.2.220

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*, 443-484. doi: 10.3102/00346543070004443

Lan, X., Legare, C. H., Ponitz, C. C., Li, S., & Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American preschoolers. *Journal of Experimental Child Psychology, 108*, 677-692. doi: 10.1016/j.jecp.2010.11.001

Martin, R. P. (1988). *The Temperament Assessment Battery for Children*. Brandon, VT: Clinical Psychology Publishing.

Matthews, J., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*, 689-704. doi: 10.1037/a0014240

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471-490. doi: 10.1016/j.ecresq.2006.09.003

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959. doi: 10.1037/0012-1649.43.4.947

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Research Quarterly, 15*, 307-329. doi: 10.1016/s0885-2006(00)00069-7

Meisels, S. J., Atkins-Burnett, S., & Nicholson, J. (1996). *Assessment of social competence, adaptive behaviors, and approaches to learning with young children* (Working Paper No. 96-18). Retrieved from the National Center for Education Statistics website: http://nces.ed.gov/pubs96/9618.pdf

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi:10.1006/cogp.1999.0734

Osborn, A. F., Butler, N. R., & Morris, A. C. (1984). *The social life of Britain's five year olds: A report of the child health and education study*. London: Routledge and Kegan Paul.

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

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12*, 427-441. doi:10.1017/S0954579400003096

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' pre-academic skills: Self-regulation as a mediating mechanism. *Child Development, 82,* 362-378. doi: 10.1111/j.1467-8624.2010.01561.x

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal psychology*, *103*, 55-66. doi: 10.1037/0021-843X.103.1.55

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*, 1394-1408. doi:10.1111/1467-8624.00355

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*, 641-660. doi: 10.1080/10409289.2014.850397

Schunk, D. H. & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32,* 195-208. doi: 10.1207/s15326985ep3204_1

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV). San Antonio, TX: The Psychological Corporation.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., . . . Gershon, R. C. (2013). I. NIH Toolbox Cognition Battery (CB): Introduction and pediatric data. *Monographs of the Society for Research in Child Development*, *78*, 1-15. doi: 10.1111/mono.12031

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.

Wright, J. C. (1971). *Kansas Reflection-Impulsivity Scale for Preschoolers (KRISP)*. St. Louis: CEMREL, Inc.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297-301. doi: 10.1038/nprot.2006.46

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78,* 16–33. doi:10.1111/mono.12032

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37-63. doi: 10.1016/S0885-2014(96)90027-1

Table 1

Factor Loadings for One-Factor Principal Factor Analysis at the Beginning (Time 1) and End of Pre-k (Time 2): Results for the Observed Values and the Values Adjusted for WJ-III Oral Comprehension with the Correlation of the Observed Values and Oral Comprehension

	Factor Loadings for Observed Values		Factor Loadings for Adjusted Values[a]		Correlation with Oral Comprehension	
CSR Measure	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Backwards Digit Span	.40	.45	.28	.33	.32	.35
Copy Design	.51	.60	.51	.58	.21	.22
DCCS	.53	.54	.36	.40	.45	.42
HTKS	.68	.69	.56	.54	.46	.52
KRISP Accuracy	.66	.66	.62	.61	.35	.31
KRISP Reaction Time	.40	.35	.40	.32	.15	.18
Operation Span	.39	.34	.35	.36	.21	.18
Peg Tapping	.70	.69	.59	.60	.44	.43
Spatial Conflict	.50	.38	.46	.34	.24	.19
Turtle-Rabbit Accuracy	.35	.33	.30	.32	.20	.12
Turtle-Rabbit Reaction Time	.41	.45	.31	.36	.28	.29
Whisper Task	.43	.31	.29	.19	.33	.28

Notes. N = 535. CSR = cognitive self-regulation; DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers.

[a] OLS residuals after predicting the CSR values from WJ-III Oral Comprehension.

Table 2

Change in Scores on the CSR Measures from the Beginning (Time 1) to End of Pre-k (Time 2)

CSR Measure	Time 1 Mean (SD)	Time 2 Mean (SD)	T1 – T2 Effect Size	T1 – T2 Correlation
Backwards Digit Span	1.31 (1.20)	2.05 (2.13)	0.43	.46
Copy Design	1.40 (1.43)	2.27 (1.70)	0.55	.59
DCCS	1.47 (0.57)	1.75 (0.52)	0.51	.38
HTKS	14.43 (15.55)	23.65 (17.57)	0.56	.66
KRISP Accuracy	28.94 (4.09)	31.44 (3.13)	0.69	.56
KRISP Reaction Time	0.15 (0.34)	0.30 (0.26)	0.50	.12
Operation Span	8.57 (3.87)	9.67 (3.18)	0.31	.38
Peg Tapping	6.99 (6.01)	10.21 (5.48)	0.56	.62
Spatial Conflict	56.06 (9.77)	59.38 (8.34)	0.37	.32
Turtle-Rabbit Accuracy	54.18 (9.96)	54.22 (6.62)	0.00	.20
Turtle-Rabbit Reaction Time	0.38 (0.40)	0.55 (0.35)	0.45	.43
Whisper Task	30.04 (8.13)	32.82 (6.04)	0.39	.35

Notes. N = 535. The pre-post difference is statistically significant at $p < .001$ for all measures except Turtle-Rabbit Accuracy. Effect size is Cohen's *d* for the difference between the means at Time 1 and Time 2. CSR = cognitive self-regulation; DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers.

Table 3

Concurrent Correlations Between the Child CSR Measures and the Teacher Rating Scales at the Beginning (Time 1) and End of Pre-k (Time 2)

	Teacher Rating Composite		CFBR - Work Related Skills		TABC - Distractibility		TABC - Persistence		CBQ - Attention Shifting		CBQ – Impulsivity	
CSR Measure	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Backwards Digit Span	.21	.17	.24	.20	.22	.15	.22	.17	.14	.14	.04	.05
Copy Design	.31	.29	.34	.32	.33	.28	.31	.32	.18	.20	.11	.08
DCCS	.28	.25	.27	.25	.29	.24	.27	.25	.21	.17	.13	.12
HTKS	.34	.40	.37	.39	.35	.36	.28	.40	.29	.29	.10	.19
KRISP Accuracy	.36	.40	.38	.39	.35	.35	.32	.38	.28	.28	.12	.24
KRISP RT	.19	.16	.22	.20	.19	.13	.21	.16	.16	.13	.01	.03
Operation Span	.20	.17	.23	.19	.23	.14	.15	.16	.14	.13	.06	.08
Peg Tapping	.42	.38	.43	.39	.42	.36	.36	.36	.34	.28	.16	.19
Spatial Conflict	.28	.24	.24	.21	.29	.20	.24	.24	.20	.17	.19	.17
Turtle-Rabbit Accuracy	.26	.28	.24	.23	.27	.26	.20	.25	.23	.23	.13	.18
Turtle-Rabbit RT	.18	.23	.23	.27	.21	.22	.13	.22	.13	.19	.04	.05
Whisper Task	.24	.20	.27	.17	.27	.19	.19	.20	.22	.10	.03	.14

Notes. N = 535. Correlations greater than .09 are statistically significant at $p < .05$ in multilevel analysis. CSR = cognitive self-regulation; DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers; RT = reaction time.

Table 4

Standardized Regression Coefficients Between Each of the CSR Measures and Later Academic Achievement

CSR Measure	Time 1 CSR & Time 2 Achievement	Time 1 CSR & Time 3 Achievement	Time 2 CSR & Time 3 Achievement
Backwards Digit Span	.42	.37	.47
Copy Design	.41	.40	.38
DCCS	.45	.44	.42
HTKS	.54	.49	.57
KRISP Accuracy	.48	.50	.43
KRISP Reaction Time	.25	.23	.21
Operation Span	.26	.27	.21
Peg Tapping	.56	.51	.52
Spatial Conflict	.35	.34	.22
Turtle-Rabbit Accuracy	.22	.23	.18
Turtle-Rabbit Reaction Time	.32	.27	.40
Whisper Task	.37	.36	.25

Notes. N = 535 at Time 1 and 2 and 488 at Time 3. All correlations are statistically significant at $p < .01$ in multilevel analysis. CSR = cognitive self-regulation; DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers. Academic achievement is the composite measure combining five Woodcock-Johnson subscales. Time 1= beginning of pre-k; Time 2 = end of pre-k; Time 3 = end of kindergarten.

Table 5

Standardized Regression Coefficients for the Relationship Between Each of the CSR Measures and Gains on the Academic Achievement Composite

CSR Measure	Time 1 CSR & T1-T2 Ach Gain	Time 1 CSR & T1-T3 Ach Gain	Time 1 CSR & T2-T3 Ach Gain	Time 2 CSR & T2-T3 Ach Gain	T1-T2 CSR Gain & T1-T2 Ach Gain	T1-T2 CSR Gain & T1-T3 Ach Gain	T1-T2 CSR Gain & T2-T3 Ach Gain
Backwards Digit Span	$.06^{*}$	$.05^{\dagger}$	.04	$.08^{*}$	$.12^{*}$	$.14^{*}$	$.06^{*}$
Copy Design	$.12^{*}$	$.12^{*}$	$.06^{*}$	$.05^{*}$	$.07^{*}$	$.06^{*}$	.02
DCCS	$.07^{*}$	$.10^{*}$	$.09^{*}$	.04	$.10^{*}$	$.06^{*}$	.00
HTKS	$.11^{*}$	$.09^{*}$	$.06^{\dagger}$	$.14^{*}$	$.09^{*}$	$.14^{*}$	$.10^{*}$
KRISP Accuracy	$.09^{*}$	$.17^{*}$	$.14^{*}$	$.10^{*}$	$.09^{*}$	$.08^{*}$	.03
KRISP RT	$.09^{*}$	$.09^{*}$	.03	.02	$.05^{*}$	$.05^{\dagger}$	.02
Operation Span	$.07^{*}$	$.09^{*}$	$.06^{*}$	.01	$.05^{*}$	.02	-.02
Peg Tapping	$.09^{*}$	$.09^{*}$	$.10^{*}$	$.05^{\dagger}$	$.11^{*}$	$.07^{*}$	-.01
Spatial Conflict	$.08^{*}$	$.08^{*}$	$.05^{\dagger}$	.02	$.06^{*}$	$.05^{\dagger}$	.01
Turtle-Rabbit Accuracy	.03	$.05^{\dagger}$	$.04^{\dagger}$	-.02	$.08^{*}$	.03	-.03
Turtle-Rabbit RT	$.04^{\dagger}$	.01	.01	$.08^{*}$	$.07^{*}$	$.11^{*}$	$.07^{*}$
Whisper Task	$.06^{*}$	$.07^{*}$	$.05^{\dagger}$	$-.05^{*}$	$.09^{*}$	-.01	$-.07^{*}$

Notes. N = 535 at T2 and 488 at T3. Ach= Achievement; CSR = cognitive self-regulation; DCCS=Dimensional Change Card Sort; HTKS=Head Toes Knees Shoulders; KRISP=Kansas Reflection-Impulsivity Scale for Preschoolers; RT=Reaction Time. Academic achievement is the composite measure combining five Woodcock-Johnson subscales. Time 1(T1) = beginning of pre-k; Time 2 (T2) =end of pre-k; Time 3 (T3) =end of kindergarten.

* $p < .05$; † $p < .10$.

Table 6

Summary of the Performance of the CSR Measures on the Selection Criteria

				Predictive Validity		
CSR Measure	Relationship to Common Factor[a]	Developmental Change[b]	Convergent Validity with Teacher Ratings[c]	T1 & T2 CSR & Later Achievement[d]	T1 & T2 CSR & Achievement Gains[e]	PreK CSR Gains & Achievement Gains[f]
Copy Design	X	X	X	X	X	X
HTKS	X	X	X	X	X	X
KRISP Accuracy	X	X	X	X	X	X
Peg Tapping	X	X	X	X	X	X
DCCS	X	X	X	X		X
Backwards Digit Span				X	X	X
Turtle-Rabbit Accuracy			X			
Turtle-Rabbit Reaction Time						X
KRISP Reaction Time						
Operation Span						
Spatial Conflict						
Whisper Task						

Notes. The better performing Cognitive Self-Regulation (CSR) measures on each criterion are indicated by X. DCCS=Dimensional Change Card Sort; HTKS=Head Toes Knees Shoulders; KRISP=Kansas Reflection-Impulsivity Scale for Preschoolers. Time 1(T1) = beginning of pre-k; Time 2 (T2) =end of pre-k; Time 3 (T3) =end of kindergarten.

[a]. Observed factor loadings at Time 1 (T1) and Time 2 (T2) are $\geq .50$ and adjusted factor loadings at T1 and T2 are $\geq .35$.

[b]. Effect size for change from T1 and T2 is $\geq .50$.

[c]. T1 and T2 correlations with the Teacher Rating Composite are $\geq .25$ and significant at $p \leq .05$.

[d]. Correlations for T1 predicting to T2 and T3 achievement, and T2 predicting to T3 achievement, are $\geq .35$ and significant at $p \leq .05$.

[e]. Correlations for T1 predicting T1-T2 and T1-T3 achievement gain, and T2 predicting T2-T3 achievement gain, are significant at $p \leq .10$ or better.

[f]. Correlations for T1-T2 gain predicting T1-T2 and T1-T3 achievement gain are significant at $p \leq .05$.

Table 7

Convergent and Predictive Validity Coefficients for the Six Best CSR Measures Analyzed Together

		Standardized Regression Coefficients for CSR Measures					
IVs: Independent Variables DV: Dependent Variable	Multiple Correlation	Copy Design	HTKS	KRISP Accuracy	Peg Tapping	DCCS	Backwards Digit Span
IVs: T1 CSR Measures DV: T1 Teacher Ratings	.49*	.16*	.10*	.15*	.24*	.08†	.02
IVs: T2 CSR Measures DV: T2 Teacher Ratings	.50*	.11*	.25*	.22*	.18*	.01	-.03
IVs: T1 CSR Measures DV: T2 Achievement	.72*	.10*	.21*	.19*	.22*	.16*	.18*
IVs: T1 CSR Measures DV: T3 Achievement	.68*	.09*	.15*	.25*	.18*	.18*	.16*
IVs: T2 CSR Measures DV: T3 Achievement	.70*	.07*	.28*	.16*	.17*	.10*	.23*
IVs: T1 CSR Measures DV: T1-T2 Achievement Gain	.28*	.08*	.07*	.05*	.03	.04	.04†
IVs: T1 CSR Measures DV: T1-T3 Achievement Gain	.28*	.06*	.04	.14*	.02	.07*	.04
IVs: T2 CSR Measures DV: T2-T3 Achievement Gain	.23	.02	.12*	.08*	.00	.00	.06*
IVs: T1-T2 CSR Gain DV: T1-T2 Achievement Gain	.39*	.04*	.05*	.06*	.07*	.07*	.10*
IVs: T1-T2 CSR Gain DV: T1-T3 Achievement Gain	.32*	.04	.12*	.05†	.03	.03	.12*

Notes. CSR = cognitive self-regulation; DCCS=Dimensional Change Card Sort; HTKS=Head Toes Knees Shoulders; KRISP=Kansas Reflection-Impulsivity Scale for Preschoolers. Academic achievement is the composite measure combining five Woodcock-Johnson subscales. Time 1(T1) = beginning of pre-k; Time 2 (T2) =end of pre-k; Time 3 (T3) =end of kindergarten.

* $p < .05$; † $p < .10$.

Table 8

Concurrent Correlations Between CSR Measures and the Teacher Rating Total Score at the Beginning (Time 1)and End of Pre-k (Time 2) in the Cross-Validation Sample

CSR Measure	Time 1	Time 2
Backwards Digit Span	.30	.27
Copy Design	.42	.47
DCCS	.36	.30
HTKS	.42	.41
KRISP Accuracy	.41	.38
Peg Tapping	.45	.41
CSR Factor Score	.63	.60

Notes: N = 356. CSR = cognitive self-regulation; DCCS=Dimensional Change Card Sort; HTKS=Head Toes Knees Shoulders; KRISP=Kansas Reflection-Impulsivity Scale for Preschoolers. The CSR Factor Score is based on the 6 individual CSR measures shown. All correlations are significant at $p < .001$.

Table 9

Standardized Regression Coefficients for the Relationship Between Each of the CSR Measures and Later Academic Achievement Outcomes for the Cross-Validation Sample

CSR Measure	Time 1 CSR & Time 2 Achievement	Time 1 CSR & T1-T2 Achievement Gain	T1-T2 CSR Gain & T1-T2 Achievement Gain
Backwards Digit Span	$.46^{**}$	.05	$.06^{*}$
Copy Design	$.40^{**}$	.05	$.10^{**}$
DCCS	$.50^{**}$	$.11^{**}$	$.10^{**}$
HTKS	$.54^{**}$	$.07^{*}$	$.08^{**}$
KRISP Accuracy	$.46^{**}$	$.09^{**}$	$.09^{**}$
Peg Tapping	$.58^{**}$	$.10^{**}$	$.15^{**}$
CSR Factor Score	$.78^{**}$	$.18^{**}$	$.16^{**}$

Notes: N = 356; CSR = cognitive self-regulation; HTKS = Head Toes Knees Shoulders; DCCS = Dimensional Change Card Sort; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers. Time 1(T1) = beginning of pre-k; Time 2 (T2) =end of pre-k. The CSR Factor Score is based on the 6 individual CSR measures shown.

** $p \leq .01$. * $p \leq .05$.

Table 10
Predictive Validity Coefficients for the CSR Measures Analyzed Together in the Cross Validation Sample

IVs: Independent Variables DV: Dependent Variable	Multiple Correlation	Standardized Regression Coefficients for CSR Measures Backwards Digit Span	Copy Design	DCCS	HTKS	KRISP Accuracy	Peg Tapping
IVs: T1 CSR Measures DV: T2 Achievement	.73**	.19**	.02	.22**	.15**	.21**	.24**
IVs: T1 CSR Measures DV: T1-T2 Achievement Gain	.27**	.04	-.01	.10**	.00	.08**	.06†
IVs: T1-T2 CSR Gain DV: T1-T2 Achievement Gain	.37**	.04	.08**	.07**	.05†	.06*	.11**

Notes: CSR = cognitive self-regulation; DCCS=Dimensional Change Card Sort; HTKS=Head Toes Knees Shoulders; KRISP=Kansas Reflection-Impulsivity Scale for Preschoolers. Academic achievement is the composite measure combining five Woodcock-Johnson subscales. Time 1(T1) = beginning of pre-k; Time 2 (T2) =end of pre-k.
** $p \leq .01$. * $p \leq .05$. † $p \leq .10$.

Appendix: Scoring Scheme for the Child Measures of LRCSR

The six LRCSR measures identified in this paper were scored on different scales (e.g., 0-3 for DCCS, 0-52 for HTKS), complicating the construction of a total score for all six measures together. One solution is to rescale the scores on each measure to a common scale, then sum them for a total score. We found that a 0-5 point scale format worked well for this purpose. To determine which original scores should be rescored into each value on this common scale, we took advantage of the linear relationship between children's age and their scores on each measure. Using data from the initial sample, we regressed the scores for each measure on age and used the results to estimate the scores in the original metric expected at ages 4.0, 4.5, 5.0, 5.5, and 6.0, spanning the pre-k age range. These estimates were then used as break points for rescaling each original score into the 0-5 format. The resulting procedure is shown below.

	Scores in the Original Metric for Each Measure					
Rescaled Score	Peg Tapping	HTKS	KRISP	DCCS	Copy Design	Backwards Digit Span
0	≤ 5	≤ 7	≤ 25	0	0	0
1	6-7	8-15	26-29	1	1	1
2	8-9	16-23	30-32	1	2	2
3	10-12	24-31	33-36	2	3	3
4	13-14	32-38	37-39	2	4-5	4
5	> 14	> 38	> 39	3	> 5	≥ 5

In the initial sample with which this scheme was constructed, correlations between rescaled scores and those in the original metric ranged from .92 to 1.00 across measures and the Time 1 (beginning of pre-k) and Time 2 (end of pre-k) measurement waves. They also performed well for the Time 3 end of kindergarten measures with correlations from .82 to .98. When applied to the Time 1 and 2 data from the cross-validation sample, the correlations ranged from .91 to 1.00. The total scores produced by summing the rescaled scores across all six items showed correlations from .94 to .99 with the factor scores for Time 1, 2, and 3 in the initial sample, and correlations from .97 to .99 with the Time 1 and 2 factor scores in the cross-validation sample.