Journal of Educational Psychology

0022-0663/17/$12.00 http://dx.doi.org/10.1037/edu0000203

Learning-Related Cognitive Self-Regulation Measures for Prekindergarten Children: A Comparative Evaluation of the Educational Relevance of Selected Measures

Mark W. Lipsey
Vanderbilt University

Kimberly Turner Nesbitt
University of New Hampshire

Dale C. Farran
Vanderbilt University

Nianbo Dong
University of Missouri-Columbia

Mary Wagner Fuhs
University of Dayton

Sandra Jo Wilson
Abt Associates

Many cognitive self-regulation (CSR) measures are related to the academic achievement of prekindergarten children and are thus of potential interest for school readiness screening and as outcome variables in intervention research aimed at improving those skills in order to facilitate learning. The objective of this study was to identify learning-related CSR measures especially suitable for such purposes by comparing the performance of promising candidates on criteria designed to assess their educational relevance for pre-K settings. A diverse set of 12 easily administered measures was selected from among those represented in research on attention, effortful control, and executive function, and applied to a large sample of pre-K children. Those measures were then compared on their ability to predict achievement and achievement gain, responsiveness to developmental change, and concurrence with teacher ratings of CSR-related classroom behavior. Four measures performed well on all those criteria: Peg Tapping, Head-Toes-Knees-Shoulders, the Kansas Reflection-Impulsivity Scale for Preschoolers, and Copy Design. Two others, Dimensional Change Card Sort and Backwards Digit Span, performed well on most of the criteria. Cross-validation with a new sample of children confirmed the initial evaluation of these measures and provided estimates of test–retest reliability.

Educational Impact and Implications Statement
The ability of prekindergarten children to regulate such cognitive functions as attention and task persistence is related to their learning and academic achievement. This study identified measures of such learning-related cognitive self-regulation especially suitable for screening pre-k children for school readiness and as outcome measures for interventions aimed at improving those skills.

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

The ability of young children to exert control over their cognition and behaviors within educational contexts has been variously labeled approaches to learning (Davoudzadeh, McTernan, & Grimm, 2015; Zimmerman, 1990), learning dispositions (Katz, 1993, 2002), and work-related skills (Cooper & Farran, 1988, 1991; Schmitt, Pratt, & McClelland, 2014). However labeled,

Mark W. Lipsey, Peabody Research Institute, Vanderbilt University; Kimberly Turner Nesbitt, Department of Human Development and Family Studies, University of New Hampshire; Dale C. Farran; Peabody Research Institute, Vanderbilt University; Nianbo Dong, Department of Educational, School, and Counseling Psychology, University of Missouri-Columbia; Mary Wagner Fuhs, Department of Psychology, University of Dayton; Sandra Jo Wilson, Social and Economic Policy Division, Abt Associates.

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 Sciences Postdoctoral Training Grant (R305B100016). 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 concerning this article should be addressed to Mark W. Lipsey, Peabody Research Institute, Vanderbilt University, Nashville, TN 37203. E-mail: mark.lipsey@vanderbilt.edu

ample 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 (Duncan et al., 2007; Li-Grining, Votruba-Drzal, Maldonado-Carreño, & Haas, 2010; McClelland, Morrison, & Holmes, 2000; Morgan, Farkas, & Wu, 2011). The constellation of skills that support this behavior can be referred to broadly as cognitive self-regulation.

Research on cognitive self-regulation (CSR) has been conducted within various conceptual frameworks including attention, executive function, and effortful control. Attentional functions such as conscious detection and sustained focus on a target stimulus are foundational aspects of one's ability to control thoughts and behaviors (Posner & Rothbart, 2000; Rothbart & Ahadi, 1994). Executive function, in turn, is generally defined as a set of cognitive abilities that aid in the completion of goal-directed actions (Hughes & Ensor, 2011; 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). Related to inhibitory control is the construct of effortful control, which involves volitional behavioral regulation related to aspects of temperament (Kochanska, Murray, & Harlan, 2000).

A number of assessments of CSR-related constructs suitable for administration directly to pre-K age children have been developed within these research contexts, and many of them have been shown to be related to concurrent or future academic achievement (Allan & Lonigan, 2011; Blair & Razza, 2007; Gathercole, Brown, & Pickering, 2003; Jacob & Parkinson, 2015; Lan, Legare, Ponitz, Li, & Morrison, 2011) and achievement gains during the pre-K and kindergarten years (Fuhs, Nesbitt, Farran, & Dong, 2014; Matthews, Ponitz, & Morrison, 2009; McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Indeed, evidence indicates that cognitive self-regulation measures are among the strongest predictors of achievement after prior measures of achievement itself (Duncan et al., 2007).

Aside from whatever theoretical insights derive from this research, the relations of such measures to the academic achievement of pre-K children has particular importance in educational contexts. Most immediately, the measures most strongly related to achievement might be used in assessments of school readiness to identify children whose CSR skills may not be sufficient to support effective engagement in the learning opportunities in kindergarten and beyond and thus need help enhancing those skills. Those measures, in turn, would be appropriate targets as outcome variables in intervention research aimed at finding ways those skills can be improved to better facilitate learning in classroom environments.

But, which of the many measures of CSR that can be used with pre-K aged children are especially suitable for these purposes? Most prior studies reporting relations with achievement have focused on only a few measures and typically did not have comparison of those relations as their main objective. Moreover, different studies have used different achievement measures and different samples, features that could themselves influence the magnitude of the relations, thus making it difficult to compare the performance of measures used in different studies. And no study has systematically assessed CSR measures with regard to the multiple attributes that would make them most educationally relevant to pre-K students.

The purpose of the study reported here is to make just such a comparative assessment for a group of candidate measures selected to represent a range of CSR skills while also being easily administered to young children. The aim of this assessment is to identify CSR measures with clear educational relevance for pre-K children; that is, measures that perform especially well when the interplay between CSR and achievement is of interest in pre-K classroom settings. The results, in turn, are intended to provide guidance to pre-K researchers and practitioners seeking measures of CSR for screening or research applications that have sound measurement properties and demonstrated relations to learning and CSR-related classroom behavior.

Criteria for Evaluating CSR Measures

A comparative assessment focused on the educational relevance of CSR measures for pre-K students first requires decisions about the basis for selecting candidate measures and specification of appropriate criteria with which to assess them. To identify promising candidate measures, we did not attempt to apply strict selection criteria but used the informed judgment of our research team to pick measures that represented a range of CSR skills and tasks (described in more detail below) and to favor measures more widely known and used in early childhood research. Further, with practical application and broad utility in mind, we considered only measures that could be easily administered in school settings by school personnel or researchers with limited resources; that is, those that could be completed in a relatively brief period without specialized equipment or online Internet connections. A similar assessment of computer-based CSR measures would be informative, but for this study we chose to focus on readily accessible measures so the findings would be as broadly useful as possible.

To assess the relative performance of the selected CSR measures, we identified a set of attributes we judged to be indicative of their educational relevance for use in pre-K contexts. The most important of these, of course, involved the relation of the measures to academic achievement. Three types of relations were differentiated. First, we examined correlations between the CSR measures administered at the beginning of the pre-K year and later achievement. With our focus on CSR skills related to learning, the most educationally relevant measures are those most directly predictive of achievement. Less predictive measures, by definition, are less closely associated with whatever influence CSR skills have on achievement.

Second, we compared the candidate measures on their ability to predict the gains in achievement made during subsequent periods. Children with better initial CSR skills may show higher subsequent achievement, but that does not necessarily mean they also gain more during that period. They are likely to have higher achievement levels to begin with and may simply maintain their relative position. If we expect children with better CSR skills to be better able to engage in the learning opportunities presented in pre-K classrooms, they should show greater gains in achievement over the pre-K year and, similarly, over later school years. The most educationally relevant CSR measures, therefore, should be those that show the strongest relations to subsequent achievement gains.

Third, we compared the candidate CSR measures on an even more specific kind of relation with achievement. Pre-K experience is not only expected to affect achievement but may also affect CSR skills themselves such that those skills will improve during the course of the school year. Indeed, learning to pay attention, stay on task, change tasks when asked, and other such CSR-related behaviors are part of the school readiness objectives of many pre-K programs. If CSR skills are related to gains in achievement, then gains in CSR skills should, in turn, be related to further gains in achievement. We therefore compared the candidate measures on the extent to which the CSR gains observed over the pre-K year were related to achievement gains. Those relations are especially informative about the potential of the different CSR measures as outcomes for research on pre-K interventions aimed at enhancing learning-related CSR skills. Such interventions would naturally want to target CSR skills for which there was some assurance that gains on those skills were associated with learning gains.

The CSR measures most on target for use in pre-K settings when their implications for learning and achievement are of primary interest should be those that show the strongest relations of these different kinds. That is, we would expect children with better learning-related CSR skills not only to have higher achievement, but to show greater achievement gains over time, and if those CSR skills improve, to show correspondingly larger gains in achievement. The more relevant measures of these learning-related CSR skills, therefore, are those that best demonstrate these relations.

We then brought two additional perspectives to the assessment of the educational relevance of the candidate CSR measures. For one of these, we considered the extent to which the measures were responsive to developmental change, that is, showed nontrivial increases as CSR skills improved through maturation and whatever facilitation occurred in school classrooms. CSR measures that show no or limited increases during pre-K and subsequent early grades are thus relatively insensitive to the gains young children are known to make during those periods. Measures that are more sensitive to change will, by their very nature, perform better for assessing change and distinguishing children whose CSR skills differ.

Finally, we considered the relation between the candidate measures and teacher ratings of the CSR-related learning skills they are able to observe in the classroom, including persistence, independence, organization, and participation. Teacher ratings of such learning skills have been found to be predictive of later academic achievement (Bodovski & Farkas, 2007; Davoudzadeh et al., 2015; Schmitt et al., 2014) and reflect how CSR skills are manifest in children's classroom behavior. However, these ratings show distinct differences from the results of direct assessments of children's CSR skills (Fuhs, Farran, & Nesbitt, 2015; Matthews et al., 2009; Schmitt et al., 2014), and thus cannot be assumed to be equivalent measures of the underlying CSR skills of interest. Nonetheless, the candidate CSR measures with the greatest educational relevance in pre-K settings should also show close relations to teacher ratings of the learning skills those teachers observe in the classroom. Such relations help establish the ecological validity of the measures for use in pre-K contexts as well as giving them credibility with teachers who may use them.

To conduct a comparative assessment of the performance of direct assessments of CSR skills on these attributes, we selected a range of candidate measures as described in more detail below and administered them to a large sample of children at the beginning and end of the pre-K year and again at the end of kindergarten. We then used those data to assess each measure for its ability to predict achievement and achievement gain, responsiveness to change over time, and correlation with teacher ratings. The best performing measures identified in those analyses were then administered to a new sample of children before and after the pre-K year to allow cross-validation of the findings from the initial sample and support collection of test–retest reliability data. The procedural details and results are described in the sections that follow.

Method

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 learning-related CSR. The skill domains distinguished for this purpose were

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.

We then reviewed a wide range of CSR-related measures that have appeared in research with pre-K age children (a list of those measures is in Appendix A). We categorized each according to the skill domain that seemed most central to accomplishing the tasks the measure presented, relying heavily on the description of the measure in the associated literature. Of course, none of these are pure measures of the skills indicated by the labels we applied to the respective domains; they all tap into multiple overlapping skills. But sorting them this way and selecting at least one measure from each category ensured that we would end up with a diverse set that collectively should span the full range of CSR skill domains identified in research on this topic. When making these selections, we prioritized measures previously shown to be related to academic achievement and those we judged to be most practical for administration in classroom 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 [RT]), which are described below.

Sustained Attention

For assessment tasks requiring the capacity to maintain focus and attention, we chose Copy Design (Davie, Butler, & Goldstein, 1972; Osborn, Butler, & Morris, 1984) and the Kansas Reflection-Impulsivity Scale for Preschoolers (KRISP; Wright, 1971). For Copy Design, children copy eight geometric designs of increasing difficulty and, for each, the quality of the best attempt is scored 0 or 1 by defined criteria with total scores ranging from 0 to 8. Cronbach alphas for this measure were .79 in the data we collected on our sample (described below) at the beginning of the pre-K year and .75 in the data collected at the end of the year.

The KRISP presents children with a series of drawings for which they must identify the duplicate of a target picture from 4–6 other pictures, all but one different in minor ways. Each of 12 trials is scored for number of errors and RT to selection of the first drawing. Accuracy is scored as the number of errors subtracted from the total errors possible (36). RT is scored as the difference between the mean for the 5 hardest and 7 easiest trials divided by the mean for the hardest ones, thus indexing how much the child slowed down to reflect on the harder items. Cronbach alphas for accuracy were .66 at the beginning of pre-K and .63 at the end.

Attention Shifting

For measures requiring the ability to shift focus from one task to another, we selected the *Dimensional Change Card Sort* (DCCS; Zelazo, 2006). Children sort a set of cards according to one dimension (color), and then according to a different dimension (shape). If they are largely successful with that switch, they are given similar cards with a black border around some and asked to sort by color if the card has a border and by shape if not. Children receive a score of 0 if they do not pass the initial color sort, 1 if they pass the color but not the shape sort, 2 if they pass the shape sort, and 3 if they also pass the border sort. Cronbach alphas for color sorting were .81 for data at the beginning of pre-K and .78 at the end; for shape sorting, .96 at the beginning of pre-K and .92 at the end. Too few children were able to complete the border task to allow alpha values to be computed.

Working Memory

For assessment tasks that require the ability to temporarily store and manage information, we selected Operation Span (Blair & Willoughby, 2006f) and Backwards Digit Span (Davis & Pratt, 1995). For Operation Span, children are shown pictures of houses with animals and colors and asked to name them, then recall the animal in each house on a second display of empty houses. Six trials with two, three, or four items to remember are scored 0 for incorrect and 1 for a correct response, with the sum as the final score (range 0 to 18). Cronbach alphas were .77 at the beginning of pre-K year and .64 at the end.

Backwards Digit Span (Davis & Pratt, 1995) asks children to remember, then reverse a series of numbers presented orally; for example, given 1, 3, the child is to respond 3, 1. Across six trials with increasing numbers of digits, each number recalled correctly in backward sequence is scored 1 with the final score as the sum of digits correctly recalled. In the pre-K year, too few children were able to complete a sufficient number of items for Cronbach's alpha to be computed.

Inhibitory Control. For measures that require the ability to suppress a prepotent response in order to complete a task, we selected Head-Toes-Knees-Shoulders (HTKS; Ponitz et al., 2009), Peg Tapping (Diamond & Taylor, 1996), and Spatial Conflict (Blair & Willoughby, 2006e). HTKS asks a child to respond to oral prompts of "touch your head" and "touch your toes" by doing the opposite for 10 trials. If responses on five or more are correct, two new prompts are added for another 10 trials. Each trial is scored 0 for an incorrect response, 1 for an incorrect motion that was corrected, and 2 for a correct response with the sum across all items as the final score (range 0 to 40). Cronbach alphas for the first 10 trials were .96 at the beginning of pre-K and the same at the end; for the second 10, they were .85 at the beginning and .88 at the end.

The Peg Tapping task asks children to tap once when the examiner taps twice and twice when the examiner taps once (Diamond & Taylor, 1996). Children largely successful in practice trials then have 16 test trials scored 0 for incorrect and 1 for correct responses. Final scores range from −1 to 16, with −1 assigned if the child does not reach criterion in the practice trials. Cronbach alphas were .87 in data at the beginning of pre-K year and .88 at the end.

The Spatial Conflict task (Blair & Willoughby, 2006e) was a paper adaptation of the computer-based version (Gerardi-Caulton, 2000). Children are given a card with one button on the right-hand side and one on the left, and shown a series of arrows that point either left or right. They are asked to touch the button on the side the arrow points to using their right hand for the button on the right and their left hand for the one on the left. A series of congruent trials (arrow on the same side of the page it points to), is followed by 16 mixed congruent and incongruent trials 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, with the total score ranging from 0 to 32. Cronbach alphas were .82 for data from the beginning of pre-K and .77 at the end.

Effortful Control

For assessment tasks that require the ability to suppress impulsive or premature responses, we selected the Whisper and Turtle-Rabbit tasks (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). In the Whisper task children are shown pictures of 12 cartoon characters and asked to whisper their names. The cartoon characters vary in familiarity, providing the opportunity for the child to act impulsively (shout) when a very recognizable one comes up. Each trial is scored 0 for a shout, 1 for a normal voice, 2 for no response, and 3 for a whisper (range 0 to 36). Cronbach alphas were .96 at the beginning of pre-K and .95 at the end.

The Turtle-Rabbit task (Kochanska et al., 1996) presents children with a drawing of a curved path with five bends and they are asked to move toy figures along the path without straying. After baseline trials with neutral figures, they are given two trials with a rabbit they are told is fast, and two with a turtle they are told is slow. Each curve is scored 0 if bypassed, 1 if the figure is above the mat but follows the general curvature, and 2 if the figure stays on the mat and within the path. Time to complete each trial is also recorded. Accuracy is scored as the total for all curves and trials (range 0 to 60). Reaction time is scored as the difference between the mean times for the turtle and rabbit trials. Cronbach alphas for accuracy were .99 for both rabbit and turtle at the beginning of pre-K, and .92 for rabbit and .89 for turtle at the end of pre-K.

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) assesses each child's ability to sustain attention. The eight items on this subscale are rated on a 1 (*hardly ever*) to 7 (*almost always*) scale 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." Cronbach alphas for this subscale were .75 at the beginning of pre-K and .74 at the end.

Distractibility. The Distractibility subscale of the TABC assesses the ability to ignore distractions. The eight items on this subscale are rated as described above and cover such behaviors as "Child is easily drawn away from his/her work by noises . . . etc." and "If other children are talking or making noise while teacher is explaining a lesson, this child remains attentive to the teacher." Cronbach alphas were .89 at the beginning of pre-K and .90 at the end.

Impulsivity. This was assessed with the Impulsivity subscale of the Children's Behavioral Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001). CBQ items are rated from 1 (*extremely untrue of student*) to 7 (*extremely true*). The 13 items cover such behavior as "sometimes interrupts others when they are speaking" and "usually stops and thinks things over before deciding to do something." Cronbach alphas were .87 at the beginning of pre-K and .88 at the end.

Attention shifting. The CBQ Attention Shifting subscale was used for this dimension. Twelve items are also rated as described above and include 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." Cronbach alphas were .87 at the beginning of pre-K and .89 at the end.

Work-related skills. A scale that spanned a variety of children's CSR skills as observed in the classroom was also 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. Items are rated from 1 to 7 using behavioral anchors distinctive to each item. Cronbach alphas were .95 at the beginning and end of pre-K.

Academic Achievement Measures

Achievement was measured with five subscales from the Woodcock Johnson III achievement battery (Woodcock, McGrew, & Mather, 2001) widely used in early childhood education research. These included two math subtests: Applied Problems (numerical and spatial problems) and Quantitative Concepts (numbers, sequencing, shapes, and symbols). Language and literacy skills were assessed with Letter-Word Identification (identify and pronounce letters and read words), Picture Vocabulary (name objects in pictures and point to the picture that goes with a word), and Oral Comprehension (complete an orally presented passage by providing the appropriate missing word). Data analysis used 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 pre-K 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 32 schools/centers across four school systems and five community childcare centers in middle Tennessee. The consent rate was 60% (range 13% to 100% across classrooms). Consented children identified 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, 5 did not assent, and 32 moved before the study ended, leaving 535 children in the final analytic sample.

Participating schools/centers were in urban, suburban, and rural settings and provided a racially and economically diverse sample of children. Although information about race and economic status was not available for individual children, aggregate data for the schools/centers showed proportions of African American children that ranged from 0% to 87% ($M = 16\%$), Hispanic children from 2% to 34% ($M = 11\%$), and non-Hispanic White children from 13% to 95% ($M = 71\%$). Economic diversity was indicated by a range of children qualifying for free or reduced price lunch programs from 16% to 100% ($M = 55\%$). The children in the analytic sample ranged in age from 3.8 to 5.4 ($M = 4.6$) at the beginning of pre-K and 52% were male.

Procedure. Children were assessed twice during the pre-K year—near the beginning (early September through October) and the end (mid-March to early May), referred to as Time 1 and Time 2, respectively. They were assessed again at the end of kindergarten (mid-March to early May; Time 3). Time 1 and 2 assessments were administered in three sessions of 20–30 min with nearly all sessions occurring within 10 or fewer weeks. 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 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 near the beginning and 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 2 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%). To accommodate limited resources for individual testing, only 10 consented children were randomly selected from classrooms with more than 10. This procedure produced a sample of 416 children, but 21 did not pass the Pre-LAS screen for English proficiency, four did not assent to the assessments, 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 varied, but the measures were administered in fixed order at each session. 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). The majority (74%) of these teachers had also participated in the initial phase of this study.

Results

Analysis of the data described above was organized to compare the 12 candidate CSR measures with regard to their performance in the three areas described earlier that we judged to be especially pertinent to applications in pre-K settings where relevance to academic achievement is a major concern: (a) their predictive ability for academic achievement, (b) responsiveness to developmental change, and (c) concurrence with teacher ratings.

Predictive Ability for Academic Achievement

The most important consideration for our purposes in assessing the CSR measures was their ability to predict 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. The most direct answer to the question of the relative strength of the relation between each of the selected CSR measures and later achievement is obtained by comparing their correlations at each time of measurement with achievement measured at a later time. To address this question we first standardized the WJ composite achievement measure and each of the CSR measures separately for each time of testing so that the magnitude of the respective relations could be easily compared. We then constructed multilevel regression models in which each CSR measure in turn was used as the sole predictor of achievement at a later time. Multilevel analysis was necessary to respect the structure of the data and ensure that standard errors were properly estimated; it was conducted with SPSS 23 Mixed Models with children nested in classrooms, classrooms in schools (three levels), and both classrooms and schools treated as random effects. All the time intervals available in our data were examined: predicting from Time 1 (beginning of pre-K) to Time 2 (end of pre-K) and Time 3 (end of kindergarten), and predicting from Time 2 to Time 3.

Table 1 reports the standardized regression coefficients estimated in each of these analyses. Because all the variables were standardized and there was only one predictor in each analysis, these coefficients can be read as zero-order product-moment correlation coefficients. All these correlations were statistically significant with the largest 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), the correlations for those CSR measures ranged from .37 to .56. From the end of pre-K (Time 2) to the end of kindergarten, they ranged from .38 to .55.

CSR predicting achievement gain. The analyses reported above show that children with better initial skills on the CSR measures show higher achievement levels at a later time, but those children also have higher achievement to begin with—the concurrent CSR-achievement correlations at Time 1 and Time 2 for the best CSR measures in Table 1 ranged from .42 to .59. The next set of analyses therefore addressed the further question of the relative ability of the CSR measures to predict the achievement gains made over a subsequent period. Our interest is in achievement gains associated with the experiences children have over a school year, not the portion predictable from their initial achievement levels prior to those experiences. For these analyses, we used the same 3-level regression models described above, but with the Time 1 WJ composite variable included as a covariate in each analysis along with the respective CSR measure. The CSR measures in these analyses, therefore, were predicting residual gain in achievement; that is, later achievement with initial achievement held constant (Cronbach & Furby, 1970).

Table 1
Standardized Regression Coefficients Between Each of the Cognitive Self-Regulation (CSR) Measures and Later Academic Achievement for the Initial and Cross-Validation Samples

	Initial sample ($n = 535$)			Cross-validation sample ($n = 356$)
CSR measure	Time 1 CSR & Time 2 Achievement	Time 1 CSR & Time 3 Achievement	Time 2 CSR & Time 3 Achievement	Time 1 CSR & Time 2 Achievement
Backwards Digit Span	.42	.37	.47	.46
Copy Design	.41	.40	.38	.40
DCCS	.45	.44	.42	.50
HTKS	.50	.49	.55	.52
KRISP Accuracy	.48	.50	.43	.46
KRISP Reaction Time	.25	.23	.21	—[a]
Operation Span	.26	.27	.21	—
Peg Tapping	.56	.51	.52	.58
Spatial Conflict	.29	.27	.18	—
Turtle-Rabbit Accuracy	.22	.23	.18	—
Turtle-Rabbit Reaction Time	.32	.31	.26	—
Whisper Task	.37	.36	.25	—
CSR factor score				.79

Note. $N = 488$ at Time 3. All correlations are statistically significant at $p < .01$ in multilevel analysis. 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. The CSR factor score is based on the six individual CSR measures shown for the cross-validation sample.
[a] Measure not included in cross-validation.

The first three columns of Table 2 show standardized regression coefficients from these analyses. It is not surprising that they are relatively small given the strong relation between initial and later achievement. Nonetheless, many of the CSR measures had statistically significant predictive relations with achievement gain from the beginning to the end of pre-K and to the end of kindergarten, as well as from the end of pre-K to the end of kindergarten. The measures with significant positive predictive relations for all three intervals, at least at $p < .10$, were Backwards Digit Span, Copy Design, HTKS, KRISP Accuracy, and Peg Tapping.

Table 2
Standardized Regression Coefficients for the Relation Between Each Cognitive Self-Regulation (CSR) Measure and Residual Gain on the Academic Achievement Composite for the Initial and Cross-Validation Samples

	Initial sample ($n = 535$)					Cross-validation sample ($n = 356$)	
CSR measure	Time 1 CSR & T1-T2 Ach Gain	Time 1 CSR & T1-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	Time 1 CSR & T1-T2 Ach Gain	T1-T2 CSR Gain & T1-T2 Ach Gain
Backwards Digit Span	.06*	.05†	.08*	.12*	.14*	.05	.06†
Copy Design	.12*	.12*	.05*	.07*	.06*	.05	.10*
DCCS	.07*	.10*	.04	.10*	.06*	.11*	.10*
HTKS	.10*	.08*	.13*	.09*	.14*	.06†	.08*
KRISP Accuracy	.09*	.17*	.10*	.09*	.08*	.09*	.09*
KRISP Reaction Time	.09*	.09*	.02	.05*	.05†	—[a]	—
Operation Span	.07*	.09*	.01	.05*	.02	—	—
Peg Tapping	.09*	.09*	.05†	.11*	.07*	.10*	.15*
Spatial Conflict	.06*	.06*	.03	.05*	.05†	—	—
Turtle-Rabbit Accuracy	.03	.05†	−.02	.08*	.03	—	—
Turtle-Rabbit Reaction Time	.03	.05†	.07*	.01	.04	—	—
Whisper Task	.06*	.07*	−.05*	.09*	−.01	—	—
CSR factor score						.18*	.16*

Note. N = 488 at T3. Ach = Achievement; 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 = beginning of pre-K; Time 2 = end of pre-K; Time 3 = end of kindergarten. The CSR factor score is based on the 6 individual CSR measures shown for the cross-validation sample.
[a] Measure not included in cross-validation.
† $p < .10$. * $p < .05$.

CSR gain predicting achievement gain. The last set of analyses addressing CSR-achievement relations compared the CSR measures with regard to the extent to which the CSR skill gains they showed during the pre-K year, and between the end of pre-K and end of kindergarten, were correlated with achievement gains made over those same periods. For this gain-with-gain analysis, we first used the same three-level regression models described earlier to estimate residual gain for each CSR measure over the respective periods by predicting later CSR scores from the initial (Time 1) values on the same CSR measure. The residuals from those analyses, representing the changes in the CSR measure that cannot be predicted from their initial status, are residual gain scores for the CSR measures. Those residual gain scores were then used as independent variables in a second series of multilevel regression analyses in which each CSR residual gain score was used to predict later achievement with initial achievement controlled, the analysis model used above to examine residual gain on achievement.

The fourth and fifth columns of Table 2 report the standardized regression coefficients from these analyses. As in the previous analysis, these coefficients are relatively small because the much larger relations between prepost CSR and prepost achievement have been adjusted out of the results. The relations of CSR residual gain during pre-K with residual achievement gain during that same year, and with residual 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, as indicated by the pattern of statistical significance, were Backwards Digit Span, Copy Design, DCCS, HTKS, KRISP Accuracy, and Peg Tapping.

Responsiveness to Developmental Change

The last set of analyses reported above demonstrated that residual gain on some of the CSR measures was significantly related to residual gain on the achievement measures. However, those analyses do not directly address the question of how much change there is on each CSR measure during the pre-K year. As noted earlier, the most educationally relevant CSR measures are those capable of showing the most growth during the pre-K year. To examine the responsiveness of the measures to developmental change, children's scores on each CSR measures at the beginning of pre-K were compared to their scores at the end of the year. These analyses were conducted with four-level regression 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. The CSR scores were not standardized for this analysis, allowing estimation of the mean scores at Time 1 (time = 0) and Time 2 (time = 1) in the original metric. Table 3 shows the means and the standard deviations for each CSR measure. The difference between children's performance at Time 1 and Time 2, indexed by the regression coefficient on the time dummy code, was statistically significant for all the CSR measures except Turtle-Rabbit Accuracy. Pre-post standardized mean difference effect sizes are also shown in Table 3, computed as the Time 2 mean minus the Time 1 mean divided by the pooled standard deviation. These effect sizes for all the measures other than Turtle-Rabbit accuracy 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 3 also shows the zero-order product-moment 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 were not so large as to indicate that only stable individual differences are reflected in these CSR measures with no room for influence from differential experiences in and out of the classroom during this period.

Concurrence With Teacher Ratings

To investigate the relation between the CSR measures and teacher's ratings of CSR-related behavior in the classroom, we examined the correlations between each CSR measure and each of the five teacher rating scales (CFBR Work Related Skills,

Table 3
Change in Scores on the Cognitive Self-Regulation (CSR) Measures from the Beginning (Time 1) to End of Pre-K (Time 2)

CSR measure	Time 1: *M* (*SD*)	Time 2: *M* (*SD*)	T1 — T2 effect size	T1 — T2 correlation
Backwards Digit Span	1.31 (1.20)	2.05 (2.13)	.43	.46
Copy Design	1.40 (1.43)	2.27 (1.70)	.55	.59
DCCS	1.47 (.57)	1.75 (.52)	.51	.38
HTKS	8.91 (11.89)	15.51 (14.11)	.51	.61
KRISP Accuracy	28.94 (4.09)	31.44 (3.13)	.69	.56
KRISP Reaction Time	.15 (.34)	.30 (.26)	.50	.12
Operation Span	8.57 (3.87)	9.67 (3.18)	.31	.38
Peg Tapping	6.99 (6.01)	10.21 (5.48)	.56	.62
Spatial Conflict	20.86 (6.82)	22.82 (6.06)	.30	.31
Turtle-Rabbit Accuracy	54.18 (9.96)	54.22 (6.62)	.00	.20
Turtle-Rabbit Reaction Time	5.84 (8.33)	10.69 (15.43)	.39	.54
Whisper Task	30.04 (8.13)	32.82 (6.04)	.39	.35

Note. $N = 535$. 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. DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers.

TABC Distractibility, TABC Persistence, CBQ Attention Shifting, and CBQ Impulsivity) and a composite scale created by summing z-scores computed for each teacher rating scale. These correlations were estimated as standardized regression coefficients in 3-level regression models in which the respective CSR measure at either Time 1 or Time 2 was the sole predictor of a teacher rating obtained at the corresponding time. The correlations of each CSR measure with the composite scale and with each individual teacher rating scale are reported in Table 4 for the beginning and end of pre-K.

As Table 4 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. The correlations were substantially similar for ratings at the beginning and end of pre-K. The correlations with individual teacher rating scales showed similar patterns, though lower for the CBQ scales.

Summary of Findings on the Selected Criteria

Table 5 summarizes the comparative findings reported above for the performance of the candidate CSR measures by identifying the top performers in each analysis based on the magnitude of the parameter estimates and/or statistical significance. 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 for predicting achievement, though it was not among the top performers in the other analyses. The most notable feature of this summary is the consistency of the CSR measures that performed well—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.

Performance of the Top CSR Measures in Combination

As the summary in Table 5 indicates, there were six CSR measures that performed best in the comparative analyses. With those results in hand, we then undertook an exploration of the relations of those six measures to achievement when taken altogether to determine which showed the strongest independent relations relative to the others and to assess the potential value of a composite of multiple measures. For that purpose, another series of three-level 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 relations to the different dependent variables of interest. This was done by first fitting the models with the six CSR measures omitted to obtain an estimate of the total unconditional variance (residual variance when the achievement pretest was a necessary covariate) across all levels on the respective dependent variables. We then ran the same models with all six measures included as predictors and obtained the total conditional variance from those analyses. The difference between the total unconditional variance without the six CSR measures and the total conditional variance with them in the model represents the amount of the total between-student

Table 4
Concurrent Correlations Between Child Cognitive Self-Regulation (CSR) Measures and Teacher Rating Scales at the Beginning (Time 1) and End of Pre-K (Time 2) for Initial and Cross Validation Samples

	Initial sample (n = 535)												Cross-validation sample (n = 356)	
	CFBR–work related skills		TABC–distractibility		TABC–persistence		CBQ–attention shifting		CBQ–impulsivity		Teacher rating composite		Teacher rating total score	
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	Time 1	Time 2
Backwards Digit Span	.24	.20	.22	.15	.22	.17	.14	.14	.04	.05	.21	.17	.30	.27
Copy Design	.34	.32	.33	.28	.31	.32	.18	.20	.11	.08	.31	.29	.42	.47
DCCS	.27	.25	.29	.24	.27	.25	.21	.17	.13	.12	.28	.25	.36	.30
HTKS	.36	.39	.35	.39	.28	.40	.28	.29	.10	.18	.34	.39	.39	.41
KRISP Accuracy	.38	.39	.35	.35	.32	.38	.28	.28	.12	.24	.36	.40	.41	.38
KRISP RT	.22	.20	.19	.13	.21	.16	.16	.13	.01	.03	.19	.16	—[a]	—
Operation Span	.23	.19	.23	.14	.15	.16	.14	.13	.06	.08	.20	.17	—	—
Peg Tapping	.43	.39	.42	.36	.36	.36	.34	.28	.16	.19	.42	.38	.45	.41
Spatial Conflict	.18	.15	.24	.17	.19	.20	.17	.13	.16	.12	.23	.19	—	—
Turtle-Rabbit Accuracy	.24	.23	.27	.26	.20	.25	.23	.23	.13	.18	.26	.28	—	—
Turtle-Rabbit RT	.21	.17	.19	.13	.16	.13	.12	.08	.03	.06	.17	.14	—	—
Whisper Task	.27	.17	.27	.19	.19	.20	.22	.10	.03	.14	.24	.20	—	—

Note. Correlations greater than .09 are statistically significant at $p < .05$ in multilevel analysis. DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers; RT = reaction time.
[a] Measure not included in cross-validation.

Table 5
Summary of the Performance of the Candidate Cognitive Self-Regulation (CSR) Measures on the Attributes Examined

CSR measure	Predicting achievement			Developmental change[d]	Concurrence with teacher ratings[e]
	T1 & T2 CSR & later achievement[a]	T1 & T2 CSR & achievement gains[b]	PreK CSR gains & achievement gains[c]		
Copy Design	X	X	X	X	X
HTKS	X	X	X	X	X
KRISP Accuracy	X	X	X	X	X
Peg Tapping	X	X	X	X	X
DCCS	X		X	X	X
Backwards Digit Span	X	X	X		
Turtle-Rabbit Accuracy					X
KRISP Reaction Time					
Operation Span					
Turtle-Rabbit Reaction Time					
Spatial Conflict					
Whisper Task					

Note. The better performing 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] Correlations for T1 predicting to T2 and T3 achievement, and T2 predicting to T3 achievement are $\geq .35$ and significant at $p \leq .05$. [b] 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. [c] Correlations for T1-T2 gain predicting T1-T2 and T1-T3 achievement gain are significant at $p \leq .05$. [d] Effect size for change from T1 to T2 is $\geq .50$. [e] T1 and T2 correlations with the Teacher Rating Composite are $\geq .25$ and significant at $p \leq .05$.

variance accounted for by the CSR measures, essentially an R-squared value when represented as a proportion. The square root of that estimate was taken as the multiple correlation of interest. In addition, the standardized regression coefficient for each CSR measure indicated the independent contribution that measure made to predicting the respective dependent variable.

The results of these analyses are summarized in the upper portion of Table 6. The first panel reports the collective relation of the six CSR measures to composite achievement measured later. The multiple correlations, ranging from .68 to .72, can be compared with the analogous correlations for the individual measures shown in Table 1, 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 second panel of Table 6 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 2, all of which are notably smaller. The regression coefficients in Table 6 indicate that KRISP Accuracy and HTKS have the strongest independent relations to achievement gain, followed by Copy Design and Backwards Digit Span.

The third panel in Table 6 reports the results for the most important predictive relations with achievement—those between residual gain on the CSR measures and residual 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 2, ranged from .32 to .39. The individual measures making the strongest independent contributions were Backwards Digit Span and Peg Tapping.

The last panel of Table 6 shows the multiple correlations that index the concurrent relations 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 4, all of which are smaller. The standardized regression coefficients in Table 6, in turn, indicate that Peg Tapping, KRISP Accuracy, and HTKS made the strongest independent contributions to those relations.

The results in Table 6 show, unsurprisingly, that a combination of the six top performing individual CSR measures has greater predictive relations with achievement and more concurrence with teacher ratings than any single measure. Moreover, in most instances the improvement in the magnitude of the respective relations is great enough to indicate that a composite of measures holds more promise as a general measure of CSR for pre-K children than any one of them used alone. Among the six measures, the strongest independent contributions were made by Peg Tapping, KRISP Accuracy, and HTKS, which would thus be the leading candidates for the most efficient composite measure. In addition, Backwards Digit Span had an especially strong influence in the relation between CSR gain and achievement gain and thus would deserve some consideration as well.

Cross-Validation

As described above, six of the candidate CSR child assessments performed well in our comparative analyses. However, the large number of analyses conducted to identify those six allow ample opportunity for chance factors in the particular sample of children and the data they provided to influence the results. In the follow-up cross-validation study, therefore, we administered those six mea-

Table 6
Multiple Correlation and Regression Coefficients for the Top Six Cognitive Self-Regulation (CSR) Measures Together Predicting Achievement and Concurring With Teacher Ratings in the Initial Sample (Top Panel) and Cross-Validation Sample (Bottom Panel)

IVs and DV	Multiple correlation	Standardized regression coefficients for CSR measures: Copy Design	HTKS	KRISP accuracy	Peg tapping	DCCS	Backwards Digit Span
Initial sample (N = 535)							
IVs: T1 CSR measures DV: T2 achievement	.72*	.10*	.17*	.19*	.24*	.17*	.19*
IVs: T1 CSR measures DV: T3 achievement	.68*	.10*	.11*	.25*	.19*	.19*	.17*
IVs: T2 CSR measures DV: T3 achievement	.71*	.07*	.25*	.17*	.19*	.12*	.23*
IVs: T1 CSR measures DV: T1–T2 achievement gain	.29*	.08*	.05*	.05*	.04	.04†	.05†
IVs: T1 CSR measures DV: T1–T3 achievement gain	.30*	.06*	.02	.14*	.03	.07*	.04
IVs: T2 CSR measures DV: T2–T3 achievement gain	.25*	.02	.10*	.08*	.01	.01	.06*
IVs: T1–T2 CSR gain DV: T1–T2 achievement gain	.38*	.04*	.06*	.06*	.08*	.07*	.10*
IVs: T1–T2 CSR gain DV: T1–T3 achievement gain	.32*	.04	.11*	.05*	.04	.04	.12*
IVs: T1 CSR measures DV: T1 teacher ratings	.56*	.16*	.09*	.15*	.25*	.08†	.02
IVs: T2 CSR measures DV: T2 Teacher ratings	.57*	.11*	.24*	.22*	.19*	.02	−.04
Cross-Validation Sample (N = 356)							
IVs: T1 CSR measures DV: T2 achievement	.73*	.02	.13*	.22*	.25*	.22*	.19*
IVs: T1 CSR measures DV: T1–T2 achievement gain	.26*	−.01	.00	.08*	.06†	.10*	.04
IVs: T1–T2 CSR Gain DV: T1–T2 achievement gain	.37*	.08*	.05†	.06*	.11*	.07*	.04
IVs: T1 CSR measures DV: T1 teacher ratings	.60*	.19*	.10†	.19*	.21*	.08	.04
IVs: T2 CSR measures DV: T2 teacher ratings	.62*	.30*	.16*	.18*	.11†	.10*	.03

Note. IV = independent variable; DV = dependant variable; 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.
† $p < .10$. * $p < .05$.

sures[1] to a new sample to check the stability of the key features that favored them in the initial analyses. We also used this new sample to estimate the test–retest reliability of the selected measures.

The six selected CSR measures and the WJ-III achievement 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. 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 a subset of 20 teacher ratings items from the initial phase: 10 items from CFBR Work-Related Skills, 3 from CBQ (2 Impulsivity, 1 Attention Shifting), and 7 from TABC (3 Persistence, 4 Distractibility). The 20 selected items were those that had the largest loadings with the common factor identified by a principal components analysis of the original 57 items. Factor loadings greater than .70 for data collected at both the beginning and end of pre-K indicated that these 20 items efficiently represented the principal factor underlying the original 57 items and they were therefore used in the cross-validation to reduce the response burden on the teachers. These items were all rated on 7-point scales and showed a high level of internal consistency (Cronbach's alpha of .98 at both Time 1 and 2). A total score was computed as the mean of the 20 items.

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 was estimated using multilevel regression to account for the effect of the nesting of children within classrooms and classrooms within schools. For each mea-

[1] Scores for the Backward Digit Span measure in the cross-validation reflect the longest span correctly recalled (range = 1–8) based on administration procedures from the Wechsler Intelligence Score for Children, 4th 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.

sure, 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 Peg Tapping, .80 (.03); HTKS, .78 (.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, yielding a reliability coefficient of .89 (.02).

Predictive relations with achievement. To assess the ability of the CSR measures to predict academic achievement 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 4, Table 1). 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 1). The correlation with achievement for a composite score that combined all six CSR measures is also shown in Table 1 and demonstrates that the combined set of items performs notably better than any one item.

The ability of each CSR measure to predict the gain children in the cross-validation sample made in achievement over the pre-K year was also 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 6, Table 2), but showed some modest inconsistencies with the initial sample results (column 1, Table 2) for Copy Design (.05 vs. .12), DCCS (.11 vs. .07), and HTKS (.06 vs. .11). The coefficients for the more revealing relations 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 7 in Table 2). The strongest relations 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 2) for some measures, specifically Backwards Digit Span (.06 vs. .12), Copy Design (.10 vs. .07), and Peg Tapping (.15 vs. .11). The predictive coefficients for the composite score that combined all six CSR measures are also shown in Table 2 and here also the combined set of items performs better than any one item.

The final series of analyses for the predictive relations with achievement in the cross-validation sample investigated the independent contribution of each of the six CSR measures relative to the others when they were used simultaneously as independent variables in multilevel regressions. These analyses were conducted using the same models and procedures described earlier for the analogous analyses with the initial data. The results are reported in Table 6 (bottom panel) along with those for the initial sample (top panel). Across all outcomes, Peg Tapping, DCCS, and KRISP Accuracy showed the largest independent relations to achievement 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 relation 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, the coefficients most similar on statistical significance and magnitude across the achievement outcomes were for KRISP Accuracy, though Peg Tapping, DCCS, and HTKS also showed relatively good replication.

Concurrent relations with teacher ratings. The total score of the 20 teacher rating items was used as the dependent variable in multilevel regression models with each CSR measure in turn as the sole independent variable, as in the analogous analysis with the initial sample. The standardized regression coefficients that represent the correlations between each CSR measure and the teacher rating total score are reported in the columns on the far right in Table 4. They ranged from .27 to .47 and all were statistically significant. The largest correlations were found for Copy Design, Peg Tapping, HTKS, and KRISP Accuracy. Compared with the analogous values from the initial sample (also shown in Table 4), all but two of these correlations are larger and those two are close to the prior values. A broader view is provided by the correlations between the total teacher rating scores at Times 1 and 2 and the composite score that combined all six CSR measures. These were .54 and .60, respectively (not shown in Table 4), again showing stronger relations than any of the individual measures in that composite.

Useful Information

As shown in analyses with both the initial and cross-validation samples, none of the individual top performing CSR measures did nearly as well in our tests as a composite of all six of them. Moreover, each of the six measures made an independent contribution to the predictive strength of the composite, so no more efficient subset of fewer than all six measures would perform quite as well. Our procedure for administering those measures with the cross-validation sample demonstrated that it was feasible to include them in a single assessment session of 35–45 min. Further information about these six measures and how they are administered can be obtained from the corresponding author. It might be tempting to shorten the battery by omitting Copy Design and Backwards Digit Span, but analyses not reported here across both samples showed that this produced a notable decrement in the performance of the composite for predicting achievement. The six measures are scored on quite different scales, however, complicating the integration of them into a single composite measure. For research purposes, computing standardized z-scores for each, then summing them provides a straightforward way to create such a composite measure. Such standardization, however, makes the scoring dependent on the means and standard deviation of the particular sample on which the data were collected. Those values may not be well estimated in small samples and, in any event, such sample dependence undermines comparability across samples and studies. For more 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. Appendix B describes the rationale, procedure, and results of this rescaling.

Use of any of the CSR measures identified in this study as outcome variables in research on the effects of interventions with pre-K students will likely involve cluster-randomized trials with students nested within classrooms and schools. The intraclass correlations (ICCs; also known as intracluster correlations) that characterize the proportions of total variance that are between schools and between classrooms within schools are critical for estimating statistical power during the planning stage and influence the standard errors in multilevel analysis. The multilevel structure of the samples used in the present study allows ICCs to be estimated for classroom and school clusters. These estimates are reported in Table 7 for the initial sample, the larger of the two available, at the beginning of pre-K. They were estimated in three-level unconditional models with each of the CSR measures in turn as the dependent variable. The respective ICC values were computed as the proportion of the total variance associated with each of the levels in the multilevel structure. The between-school and between-classroom-within-school ICCs were relatively modest for this pre-K sample with the between-classroom value virtually zero for several measures.

Discussion

The objective of this study was to identify direct assessment measures of CSR for pre-K children that are closely linked to their academic achievement, that is, learning-related cognitive self-regulation (LRCSR), and that perform well on other criteria that make them educationally relevant for research and practical applications. In pursuing this objective, we evaluated existing measures in a comparative fashion, choosing candidate measures with attention to the aspect of CSR most salient in the tasks the measure presented to children, prior evidence about their association with academic achievement, and the ease with which they could be administered in classroom settings. The most important consideration for identifying the best performing of the selected measures was their ability to predict children's subsequent academic achievement and achievement gains. Because an important use of such measures is as outcomes for research on interventions aimed at improving LRCSR, we also considered the extent to which the measures showed change over the pre-K year, thus demonstrating their ability to respond to increases in children's LRCSR skills. Finally, we attended to the concurrent relations between the candidate measures and ratings by pre-K teachers of children's LRCSR-related behavior in the classroom as a further indication of their educational relevance.

Table 7
Intraclass Correlation Coefficients Associated With Students Nested Within Classrooms and Classrooms Nested Within Schools in the Initial Sample

CSR measure	Between schools	Between classrooms within schools	Between students within classrooms
Backwards Digit Span	.020	.000	.980
Copy Design	.049	.017	.934
DCCS	.028	.038	.934
HTKS	.006	.036	.958
KRISP Accuracy	.013	.000	.987
Peg Tapping	.035	.000	.965
CSR Composite Score	.027	.020	.953

Note. $N = 535$. CSR = cognitive self-regulation; DCCS = Dimensional Change Card Sort; HTKS = Head Toes Knees Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers. The CSR Composite Score combines the six individual CSR measures shown.

Of the 12 candidate measures evaluated, analyses with the initial pre-K sample identified six that performed especially well against these criteria. Cross-validation with a new sample confirmed the stability of those findings. The best performing measures overall 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.

The findings reported here complement and extend the body of research on the psychometric characteristics of CSR measures for pre-K age children. While same day test-retest reliability for DCCS has been documented (Beck, Schaefer, Pang, & Carlson, 2011) and normed performance standards have been established (Weintraub et al., 2013), the present study adds to the validation of this measure for use in preschool settings to assess learning-related CSR by demonstrating its relation with concurrent teacher ratings of CSR in situ and with later academic achievement. Also, with regard to research with the HTKS task, the present work adds to the somewhat mixed results from attempts to demonstrate associations between teacher ratings of classroom behavioral regulation and academic achievement (Graziano et al., 2015; McClelland et al., 2007; Ponitz et al., 2009) by demonstrating test–retest reliability in both CSR tasks and teacher ratings. Similarly, this study contributes test-retest and internal consistency reliability estimates and supportive validity data for preschool applications for all six of the measures that performed best in our comparative evaluation.

Although sound psychometric characteristics are fundamental for any CSR measure that will be used for practical or research purposes, the unique contribution of this study is the head-to-head comparison of the selected measures on a range of probing performance indicators related to their educational relevance for pre-K children in classroom settings. The results provide a firm empirical basis for the use of any of the top performing measures for either of the applications that motivated this study. The better performing measures showed close relations to achievement and achievement gains, sensitivity to developmental change, and reasonable congruence with the CSR-related behavior teachers observed in the classroom. These characteristics make them especially suitable as screening measures to identify children whose CSR skills may be low enough to impair their learning in pre-K contexts and to monitor improvement in those skills during the pre-K year. Further, those characteristics make the top performing measures suitable choices as outcomes for intervention research aimed at improving those CSR skills that have sufficiently close relations to learning that such improvement may, in turn, boost academic achievement. It is especially fortuitous in this regard that the best performing individual CSR measures are among the easiest to administer and score. Most notably, peg tapping and HTKS performed very well by the criteria we applied and both can be administered quickly and easily without special equipment or materials and without extensive training. Each could thus be used on a stand-alone basis with the results reported here providing assurance of their educational relevance for pre-K students.

Similar to Willoughby et al. (2016), however, we found that the best performing CSR measures work better as a composite. It is hardly surprising that a combination of related measures performs better than any individual measure by itself. What was somewhat unexpected was that each of the six made significant independent contributions to at least some of the relations examined with them in combination. The best composite measure based on these results, therefore, would include all six individual measures. However, there were differences in the independent contribution each measure made to the performance of that composite, with Peg Tapping, KRISP Accuracy, and HTKS demonstrating the strongest independent relations, and gain on Backwards Digit Span showing an especially strong independent relation to achievement gain. A more efficient composite measure incorporating the three top performers in this analysis, possibly with Backward Digit Span included as a fourth, therefore, would also perform better than any single measure while not requiring data collection on all six.

It is notable that the six measures contributing to the full composite measure represented a mix of CSR skills—two primarily emphasizing sustained attention (Copy Design, KRISP), two emphasizing inhibitory control (Peg Tapping, HTKS), one emphasizing attention shifting (DCCS), and one emphasizing working memory (backward digit span). Although the skills in this mix were interrelated and all were found to be relevant to learning, no one skill dominated so strongly that the others were irrelevant. This grouping of tasks as indicators of CSR aligns with prior work supporting a multidimensional view of executive function (Miyake et al., 2000).

Recognition of the advantages of an ensemble of measures to fully represent various forms of CSR is not unique to the present study. The work on the NIH Toolbox of brief measures for executive function (Weintraub et al., 2013; Zelazo & Bauer, 2013) and the program of research by Willoughby and colleagues (Willoughby, Blair, Wirth, & Greenberg, 2012; Willoughby, Wirth, & Blair, 2011) also takes this approach. Moreover, Willoughby et al. (2016) identified a cluster of measures of executive function that are associated with academic achievement. Similarly, the Chicago School Readiness Project has developed a brief direct assessment battery of CSR measures appropriate for field-based settings (Smith-Donald, Raver, Hayes, & Richardson, 2007). The measures in their battery have shown relations to classroom learning behaviors (Denham, Warren-Khot, Bassett, Wyatt, & Perna, 2012) and concurrent and future academic achievement (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009) in young children.

The present work extends these efforts in several ways. First, it introduces a particular mix of easy to administer LRCSR measures especially suitable for use in early childhood education settings. Additionally, it broadens the scope of the data supporting the relations of these LRCSR measures to learning in those settings. As with the measures in the Chicago School Readiness Project battery, the measures we have identified are related to academic achievement and teacher reported LRCSR, relations that have not been demonstrated for many CSR measures appropriate for use with pre-K age children. In addition, however, this study has further explored relations with achievement gains, assessed developmental change over time, and demonstrated both test-retest and internal consistency reliability for the better performing measures that emerged from our comparative evaluation.

There are, of course, limitations to the research presented here. For practical reasons, it was not possible to collect data and conduct comparative analyses for all the CSR measures that have been used with pre-K age children. The selections we made may have omitted some measures that would have performed as well or better than those chosen. In particular, we would expect some of the computer-based measures to perform well on our criteria. Nonetheless, we excluded them to focus on measures that did not require computer support or Internet connections, which we believe makes them more accessible and easily used in pre-K classroom settings, especially for potential use by teachers.

We also acknowledge the uncertain generalizability of the findings based on the sample of pre-K children that provided the data for this study. Though the initial sample included more than 500 children drawn from a relatively large number of schools and community childcare centers, it was of necessity limited to children whose parents consented to their participation. We have no data for the 40% of the children in those classrooms whose parents did not return consent forms (only a very few actively declined to consent) and thus have no basis for determining if they were systematically different from those consented in ways that might have affected our findings. And, though the sample was diverse with regard to gender, race, and economic status, it was fundamentally a convenience sample, not a probability sample of a defined population of pre-K children. Because of the span of schools, childcare centers, and classrooms, we have some confidence that this sample represented fairly typical pre-K age children in the middle Tennessee region, but no assurance that similarly constructed samples in other parts of the country would have produced comparable findings.

We also must emphasize that the fact that the LRCSR measures identified in this study are predictive of later achievement and achievement gains does not mean that they represent causal factors for those outcomes. Our purpose in this study was not to attempt to establish causal relations but, among other objectives, to identify measures that might be especially appropriate as outcome variables in research that does investigate causal influences. With conceptually relevant and responsive measures in hand, a key question for future research is what practical interventions or teacher practices are capable of increasing pre-K children's LRCSR skills. There is some evidence using one or another of the measures identified here that such effects are possible (e.g., Bierman et al., 2008; Fuhs, Farran, & Nesbitt, 2013; Raver et al., 2011), but also some less encouraging findings (e.g., Barnett et al., 2008).

Assuming that LRCSR can be boosted, an even more important question is whether doing so for pre-K children will, in turn, lead to greater learning and increased academic achievement. With regard to that question, we believe the concurrence between the LRCSR measures identified in this study and teacher ratings of LRCSR-related behaviors in the classroom is especially important. A very plausible theory of action for the potential effects on academic achievement of interventions that increase LRCSR is that they are mediated by the kinds of LRCSR-related behaviors teachers observe in the classrooms, for example, engagement in learning activities, persistence in completing tasks, attentiveness to teachers' instructions, and the like. Testing such causal and mediational relations is best done via randomized experiments and goes beyond the scope of the present study but is a promising area for further research aimed at improving the effectiveness of pre-K instruction. Based on the evidence developed in the present study, the well-performing LRCSR measures we have identified should be quite appropriate for supporting such research.

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.ecresq.2008.03.001

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development, 12,* 169–193. http://dx.doi.org/10.1080/15248372.2011.563485

Beck, L. H., Bransome, E. D., Jr., Mirsky, A. F., Rosvold, H. E., & Sarason, I. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350. http://dx.doi.org/10.1037/h0043220

Bender, L. (1938). *A visual motor Gestalt test and its clinical use* (Research Monograph No. 3). New York, NY: American Orthopsychiatric Association.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi Block-Tapping Task: Methodological and theoretical considerations. *Brain and Cognition, 38,* 317–338. http://dx.doi.org/10.1006/brcg.1998.1039

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, 20,* 821–843. http://dx.doi.org/10.1017/S0954579408000394

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. http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x

Blair, C., & Willoughby, M. T. (2006a). *Measuring executive function in young children: Item Selection*. Chapel Hill, NC: Pennsylvania State University and University of North Carolina.

Blair, C., & Willoughby, M. T. (2006b). *Measuring executive function in young children: Operation Span*. Chapel Hill, NC: Pennsylvania State University and University of North Carolina.

Blair, C., & Willoughby, M. T. (2006c). *Measuring executive function in young children: The Pig Game*. Chapel Hill, NC: Pennsylvania State University and University of North Carolina.

Blair, C., & Willoughby, M. T. (2006d). *Measuring executive function in young children: Silly Sounds Game*. Chapel Hill, NC: Pennsylvania State University and University of North Carolina.

Blair, C., & Willoughby, M. T. (2006e). *Measuring executive function in young children: Spatial Conflict*. Chapel Hill, NC: Pennsylvania State University and University of North Carolina.

Blair, C. B., & Willoughby, M. T. (2006f). *Measuring executive function in young children: Spatial Conflict II: Arrows*. 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. http://dx.doi.org/10.1086/525550

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24,* 337–349. http://dx.doi.org/10.1016/j.ecresq.2009.06.001

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

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053. http://dx.doi.org/10.1111/1467-8624.00333

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk factors in kindergarten. *Early Childhood Research Quarterly, 3,* 1–19. http://dx.doi.org/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.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—or should we? *Psychological Bulletin, 74,* 68–80. http://dx.doi.org/10.1037/h0029382

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

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47,* 25–31. http://dx.doi.org/10.1080/00049539508258765

Davoudzadeh, P., McTernan, M., & Grimm, K. (2015). Early school readiness predictors of grade retention from kindergarten through eighth grade: A multilevel discrete-time survival analysis approach. *Early Childhood Research Quarterly, 32,* 183–192. http://dx.doi.org/10.1016/j.ecresq.2015.04.005

Denham, S. A., Warren-Khot, H. K., Bassett, H. H., Wyatt, T., & Perna, A. (2012). Factor structure of self-regulation in preschoolers: Testing models of a field-based assessment for predicting early school readiness. *Journal of Experimental Child Psychology, 111,* 386–404. http://dx.doi.org/10.1016/j.jecp.2011.10.002

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. http://dx.doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28,* 689–729. http://dx.doi.org/10.1207/s15326942dn2802_7

Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cognitive deficits in children treated early and continuoulsly for PKU. *Monographs of the Society for Research in Child Development, 62*(4), Serial No.252.

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.

Dibner, A. S., & Korn, E. J. (1969). Group administration of the Bender-Gestalt test to predict early school performance. *Journal of Clinical Psychology, 25,* 263–268. http://dx.doi.org/10.1002/1097-4679(196907)25:3<263::AID-JCLP2270250311>3.0.CO;2-D

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. http://dx.doi.org/10.1037/0012-1649.43.6.1428

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

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2013). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *School Psychology Quarterly, 28,* 347–359. http://dx.doi.org/10.1037/spq0000031

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive function skills and achievement gains: Comparing direct assessments and teacher ratings. *Journal of Educational Psychology, 107,* 207–221. http://dx.doi.org/10.1037/a0037366

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology, 50,* 1698–1709. http://dx.doi.org/10.1037/a0036633

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. http://dx.doi.org/10.1111/1467-7687.00134

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a Stroop-like day-night test. *Cognition, 53,* 129–153. http://dx.doi.org/10.1016/0010-0277(94)90068-X

Graziano, P. A., Slavec, J., Ros, R., Garb, L., Hart, K., & Garcia, A. (2015). Self-regulation assessment among preschoolers with externalizing behavior problems. *Psychological Assessment, 27,* 1337–1348. http://dx.doi.org/10.1037/pas0000113

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education, 71,* 151–174. http://dx.doi.org/10.1080/00220970309602061

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34,* 1326–1339. http://dx.doi.org/10.1037/0012-1649.34.6.1326

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology, 108,* 663–676. http://dx.doi.org/10.1016/j.jecp.2010.06.005

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85,* 512–552. http://dx.doi.org/10.3102/0034654314561338

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology, 20,* 573–591. http://dx.doi.org/10.1207/S15326942DN2003_2

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science, 6,* 498–504. http://dx.doi.org/10.1111/1467-7687.00307

Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs, 78*(1), No. 578.

Katz, L. G. (1993). Dispositions: Definitions and implications for early childhood practices. *Perspectives From EECE: A Monograph Series*. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.

Katz, L. (2002). "Not all dispositions are desirable": Implications for assessment. *Assessment in Education: Principles, Policy & Practice, 9,* 53–54. http://dx.doi.org/10.1080/09695940220119175

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. http://dx.doi.org/10.1037/0012-1649.36.2.220

Kochanska, G., Murray, K., 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. http://dx.doi.org/10.2307/1131828

Koppitz, E. M. (1973). Bender Gestalt Test performance and school achievement: A 9-year study. *Psychology in the Schools, 10,* 280–284. http://dx.doi.org/10.1002/1520-6807(197307)10:3<280::AID-PITS2310100302>3.0.CO;2-6

Kirkorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: A standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology, 16,* 840–850. http://dx.doi.org/10.1080/01688639408402697

Kuhl, J., & Kraska, K. (1993). Self-regulation: Psychometric properties of a computer-aided instrument. *German Journal of Psychology, 17,* 11–24.

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. http://dx.doi.org/10.1016/j.jecp.2010.11.001

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46,* 1062–1077. http://dx.doi.org/10.1037/a0020066

Luria, A. R., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of the beahvior disturbance produced by a left frontal arachnoidal endothelioma (meningioma). *Neuropsychologia, 2,* 257–280. http://dx.doi.org/10.1016/0028-3932(64)90034-X

MacCoby, E., Dowley, E., Hagen, J., & Degerman, R. (1965). Activity level and intellectual functioning in normal preschool children. *Child Development, 36,* 761–770. http://dx.doi.org/10.2307/1126921

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. http://dx.doi.org/10.1037/a0014240

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/S0885-2006(00)00069-7

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. http://dx.doi.org/10.1006/cogp.1999.0734

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44,* 472–488. http://dx.doi.org/10.1177/0022219411414010

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, UK: Routledge and Kegan Paul.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia, 20,* 249–262. http://dx.doi.org/10.1016/0028-3932(82)90100-2

Pickering, S. J., & Gathercole, S. E. (2001). *Working memory test battery for children*. London: Psychological Corporation.

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

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12,* 427–441. http://dx.doi.org/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' preacademic skills: Self-regulation as a mediating mechanism. *Child Development, 82,* 362–378. http://dx.doi.org/10.1111/j.1467-8624.2010.01561.x

Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory self-control in preschool children. *Merrill-Palmer Quarterly, 30,* 131–147.

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,* 958–972. http://dx.doi.org/10.1037/a0015861

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350. http://dx.doi.org/10.1037/h0043220

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103,* 55–66. http://dx.doi.org/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. http://dx.doi.org/10.1111/1467-8624.00355

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42,* 1029–1040. http://dx.doi.org/10.1016/j.neuropsychologia.2003.12.012

Schmitt, S., Pratt, M., & McClelland, M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development, 25,* 641–660. http://dx.doi.org/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,* 173–187. http://dx.doi.org/10.1016/j.ecresq.2007.01.002

Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon says.". *Child Development, 44,* 849–853. http://dx.doi.org/10.2307/1127737

Ward, H., Shum, D., McKinlay, L., Baker-Tweney, S., & Wallace, G. (2005). Development of prospective memory: Tasks based on the prefrontal-lobe model. *Child Neuropsychology, 11,* 527–549. http://dx.doi.org/10.1080/09297040490920186

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. http://dx.doi.org/10.1111/mono.12031

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102,* 43–53. http://dx.doi.org/10.1037/a0016738

Willoughby, M. T., & Blair, C. B., & the Family Life Project Investigators. (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment, 28,* 319–330. http://dx.doi.org/10.1037/pas0000152

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment, 24,* 226–239. http://dx.doi.org/10.1037/a0025361

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of Experimental Child Psychology, 108,* 414–435. http://dx.doi.org/10.1016/j.jecp.2010.04.007

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, MO: 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. http://dx.doi.org/10.1038/nprot.2006.46

Zelazo, P. D., & Bauer, P. J. (2013). National Institute of Health Toolbox Cognition Battery (NIH Toolbox CB): Validation for children between 3 and 15 years. *Monographs of the Society for Research in Child Development, 78,* 1–172.

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

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25,* 3–17. http://dx.doi.org/10.1207/s15326985ep25012

Appendix A

Candidate Direct Child Assessment Tasks Categorized by the Most Salient CSR Component Skill

Instrument	Task
Sustained attention—attending to and sustaining focus on a task	
1. Copying tasks	
Bender Gestalt Test (Bender, 1938; Dibner & Korn, 1969; Koppitz, 1973)	Children copy nine simple geometric designs exactly.
Copy Design (Davie et al., 1972; Osborn et al., 1984)	Children copy eight simple geometric designs exactly.
2. Matching tasks	
Matching Familiar Figures Test (Kagan et al., 1964)	Children select a picture that matches a target picture; accuracy is a measure of attention, over-fast reaction times assess impulsivity.

(*Appendices continue*)

Appendix A (*continued*)

Instrument	Task
Kansas Reflection-Impulsivity Scale for Preschoolers (Wright, 1971)	Children select a picture that matches a target picture; accuracy is a measure of attention, over-fast reaction times assess impulsivity.
Tower of London Task (Kirkorian et al., 1994; Ward et al., 2005)	Children build a tower to match a picture using blocks and a peg for stacking.
3. Stimulus-response tasks	
Continuous Performance Test (Beck et al., 1956)	Children press a button when a computer-generated target stimulus appears and inhibit responses to non-target stimuli. Correct responses measure of sustained attention; incorrect responses measure impulsivity.
Self-Regulation Test for Children (Howse et al., 2003; Kuhl & Kraska, 1993)	Children press the button that matches the target on a screen; can involve distractors to make the task more difficult.
Attention shifting—shifting focus within or between tasks as situations demand	
Something's the Same/Item Selection (Blair & Willoughby, 2006a)	Children categorize colored pictures first by the object, then switch to sort by color.
Dimensional Change Card Sort (Diamond et al., 2005; Zelazo, 2006) & variants (Wisconsin Card Sorting Task)	Children sort a set of cards by shape, then switch to sort by color. A more difficult version adds a variable cue that indicates the sorting rule.
Flexible Item Selection Task (Jacques & Zelazo, 2001)	Children select a pair of cards that match on one dimension (e.g., shape, color), then must select a different pair that matches on another dimension.
Working memory—active maintenance and manipulation of information in memory	
Operation Span (Blair & Willoughby, 2006b)	Children must recall a series of objects shown inside a picture of a house. A color distractor adds difficulty.
Self-Ordered Pointing (Petrides & Milner, 1982).	Children are presented a set of pages divided into four sections; they must go through and point to a different remembered picture on each page.
Backward Digit Span (Davis & Pratt, 1996; Pickering & Gathercole, 2001).	Children recall a series of orally presented digits backwards.
Corsi Block Tapping (Berch et al., 1998)	Children reproduce the sequence in which the assessor taps a series of blocks with sequences of increasing length.
Inhibitory control—volitional inhibition of a prepotent response to complete a task	
1. Stroop-like tasks	
Silly Sounds Game (Blair & Willoughby, 2006d)	Children meow to pictures of dogs and bark to pictures of cats.
Day/Night (Carlson & Moses, 2001; Gerstadt et al., 1994)	Children say "night" to sun pictures and "day" to moon pictures.
Grass/Snow (Carlson & Moses, 2001)	Children say "green" to snow pictures, and "white" to grass pictures.
2. Stroop-like tasks with motor response	
Bear & Dragon and variants (Jones et al., 2003; Reed et al., 1984)	A bear puppet and dragon puppet give children tasks (touching feet, hopping, etc.); then, they are asked to only perform tasks given by the bear puppet, not those given by the dragon puppet.
Simon Says (Carlson, 2005; Strommen, 1973)	Children perform certain tasks (touching their feet, hopping, etc.) only when the assessor precedes the command with "Simon Says."
Head-to-Toes; Head-Toes-Knees-Shoulders (Ponitz et al., 2009; McClelland et al., 2007)	Children do the opposite of what the assessor requests; e.g., if asked to touch their head, they touch their toes.
Luria Hand Game (Hughes, 1998; Luria et al., 1964)	Similar to the Head-to-Toes task, children do the opposite of what the assessor indicates using hand signals (e.g., holding up a fist vs. one finger).
Peg- or Finger-Tapping (Diamond & Taylor, 1996; Diamond et al., 1997; Smith-Donald et al., 2007)	Children tap a peg, pencil, or finger twice when the experimenter taps once, and vice versa.
Pig Game (Blair & Willoughby, 2006c)	In a series of animal pictures, children press a button when they see animals that aren't pigs, and don't press the button when they do see pigs.
3. Spatial Conflict/Simon Tasks	
Spatial conflict (Blair & Willoughby, 2006e)	Target pictures are presented on the left side of a paper and children point to the target pictures with their right hand, and vice versa.

(*Appendices continue*)

Appendix A (*continued*)

Instrument	Task
Spatial conflict (Gerardi-Caulton, 2000)	Computerized version of the task above; children push a key on one side of the keyboard for target on the opposite side of the computer screen.
4. Flanker Tasks	
Attention Network Task/Flanker Task (Ponitz et al., 2009; Rueda et al., 2004)	Children indicate the direction of a target flanked by same/opposite direction distractors; e.g., feed the central fish by pressing a button corresponding to the direction which the middle fish is swimming when flanked by fish swimming the same or opposite direction.
Effortful control—suppression of impulsive or premature responses when required by a task	
Snack delay; gift delay (Kochanska et al., 2000)	Variety of delay tasks in which children must wait before eating a cookie, open a gift, etc.
Tower: Turn-taking (Kochanska et al., 1996)	Assessor and child take turns placing blocks on a tower; children must wait their turn without reminders.
Whisper (Kochanska, et al., 1996)	Children see pictures of familiar cartoon characters and whisper their names; number whispered vs. shouted or said in normal voice is scored.
Walk-a-Line Slowly/Draw-a-Line Slowly (Maccoby et al., 1965)	Children walk or draw a line at normal speed, then do the same thing slowly; time difference between regular and slow trials is scored.
Turtle and Rabbit (Kochanska et al., 1996)	Children are given "fast" rabbit and a "slow" turtle toys and move them along a path; scored for accuracy in negotiating the path and the time difference between fast and slow trials.

(*Appendices continue*)

Appendix B

Scoring Scheme for the Child Measures of Learning-Related Cognitive Self-Regulation

Rescaled score	Scores in the original metric for each measure					
	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

Note. HTKS = Head-Toes-Knees-Shoulders; KRISP = Kansas Reflection-Impulsivity Scale for Preschoolers; DCCS = Dimensional Change Card Sort. The six learning-related cognitive selfregulation (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 relation 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 above. 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.

Received May 7, 2014
Revision received February 2, 2017
Accepted February 21, 2017 ■