# Achievement and Self-Regulation in Pre-Kindergarten Classrooms: Effects of the Tools of the Mind Curriculum

Dale C. Farran, Ph.D.
Sandra Jo Wilson, Ph.D.

7/27/2014

PEABODY research institute

VANDERBILT PEABODY COLLEGE

**Director**
Mark W. Lipsey, Ph.D.

**Senior Associate Director**                                    **Associate Director**
Dale C. Farran, Ph.D.                                            Sandra Jo Wilson, Ph.D.

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

**Contact Us:**

Phone: 615.322.8540
Fax: 615.322.0293

Mailing Address:                                    Delivery Address:

Peabody Research Institute                          Peabody Research Institute
230 Appleton Place                                  1930 South Drive
PMB 181                                             Room 410A
Nashville, TN 37203-5721                            Nashville, TN 37212

Achievement and Self-Regulation in Pre-Kindergarten Classrooms:

Effects of the *Tools of the Mind* Curriculum

Dale C. Farran and Sandra Jo Wilson

Vanderbilt University

*Submitted to Child Development, July 27, 2014*

Author Note

Abstract

Results from a cluster randomized trial of the impact of the Tools of the Mind pre-kindergarten curriculum on preschoolers' achievement, self-regulation, and social behavior are reported. Sixty classrooms with four-year-old children (*N*=877) were assigned to implement either the Tools curriculum (*N*=32 classrooms; 498 children) or the usual pre-k curriculum (*N*=28 classrooms; 379 children). In spite of fairly high implementation fidelity, Tools did not produce significant gains on assessments of achievement, self-regulation, or teacher ratings at the end of pre-k. Negative effects were found on Letter Word Identification, Quantitative Concepts, and Corsi Backward Span in kindergarten and on Spelling, Copy Design, and the composite self-regulation score in 1st grade. Several reasons for these unexpected findings are explored.

Achievement and Self-Regulation in Pre-Kindergarten Classrooms:

Effects of the *Tools of the Mind* Curriculum

**Importance of Prekindergarten Curricula**

Most states currently offer some form of voluntary universal pre-kindergarten (pre-k) for children from low-income families, while Florida, Georgia, and Oklahoma offer universal pre-k for all 4-year-old children. According to the latest *Condition of Education* report (Kena et al., 2014), almost two thirds of 3-5 year olds were in preschool in 2012, 60% of whom attended full day programs. As pre-k programs expand in number, the issues concerning the type, format, and effectiveness of these programs become critically important, particularly with respect to curriculum choices and teacher professional development.

There is evidence that participation in formal pre-k improves some aspects of school readiness at kindergarten entry (Gormley, Gayer, Phillips, & Dawson, 2005; Lamy, Barnett, & Jung, 2005), but evidence for longer-term effects is mixed and a matter of some debate (Hustedt, Barnett, & Jung, 2008; Lipsey, Hofer, Dong, Farran, & Bilbrey, 2013; Magnuson, Ruhm, & Waldfogel, 2007). Other research suggests that, while pre-k programs may improve basic pre-reading skills, their influence on complex language skills, mathematics, self-regulation, and social skills is less clear (Gormley et al., 2005; Jackson et al., 2007; U.S. Department of Health and Human Services, 2005, 2010). Apparent from the mixed research results is that early childhood educators do not yet know how to configure pre-k instruction to reliably promote both school readiness and long-term school success.

These issues make the question of curriculum effectiveness an important and timely one. At present, there is little evidence that different pre-k curricula produce significantly different effects (see Darrow, 2013). The most rigorous and comprehensive evaluation of pre-k curricula to date is the PCER project, which launched 14 randomized trials around the country (Preschool Curriculum

Evaluation Research Consortium, 2008). Most of the curricula tested had a literacy or general

developmental focus (with one focused on math). Overall, 10 of these curricula showed no

statistically significant impacts on any of the student-level outcomes of reading, phonological

awareness, language, or mathematics. None of the pre-k curricula had statistically significant positive

impacts on social skills or problem behaviors. No curriculum outperformed the control classrooms

on all child outcomes; only two showed significant differences on even one skill measured in

kindergarten. And, no curriculum stood out as notably more effective than any of the others.

Despite the lack of evidence differentiating these curricula, many school systems continue to

implement specific curricula, presumably based on philosophical preferences about the nature of

early childhood learning. State-funded pre-k programs often require systems to choose a curriculum

to implement from a list of approved options. One of the currently popular pre-k curricula is *Tools of

the Mind* (Bodrova & Leong, 2007), one of a few curricula recommended for facilitating self-

regulation as well as academic skills (Diamond & Lee, 2011; Hughes, 2011).

**School Readiness Skills**

Successful transition into formal schooling for young children, and subsequent academic

success, requires a variety of competencies, including most obviously the early literacy and numeracy

skills that provide the foundation for reading and mathematics. Another critical area of competency

is the ability to engage in and benefit from the kinds of learning tasks intrinsic to school-based

instruction, including attending to speech that conveys information, completing exercises that

require planning, problem solving, application of knowledge and practice of acquired skills, and

remembering and following rules and instructions (Cooper & Farran, 1988; Howse, Lange, Farran,

& Boyles, 2003; McClelland & Morrison, 2003). These latter skills enable children to focus on, and

benefit from, the educational material and learning opportunities provided in school settings (Blair

& Razza, 2007). Longitudinal studies (e.g., Bodovski, & Farkas, 2009, Duncan et al., 2007; Moffitt et

al., 2011) have demonstrated that these self-regulation skills have an independent relationship with long term academic success, separate from early academic skills.

**The Development of Self-Regulation**

Self-regulation is an umbrella term that refers to skills such as inhibitory control, working memory, sustained attention, and persistence, which are essential for adapting to a formal school environment (Hughes, 2011). Unfortunately, there is limited research available to clarify how pre-k students, especially those who come from poverty, develop the necessary self-regulation skills to be prepared for school. Self-regulation skills show rapid improvement in the preschool years (Carlson, 2005; Garon, Bryson, & Smith, 2008), but children from low-income homes often lag behind their middle- to high-income peers in both language skills and self-regulation (e.g., Howse et al., 2003; Noble, McCandliss, & Farah, 2007; Noble, Norman, & Farah, 2005).

Current educational theory suggests that self-regulation skills are critically correlated with the development of early academic skills (e.g., Bull, Espy, Wiebe, Sheffield, & Nelson, 2011; Duncan et al., 2007; Fuhs, Nesbitt, Farran, & Dong, 2014; Welsh, Nix, Blair, Bierman, & Nelson, 2010) as well as other positive life outcomes (e.g., Caspi, Wright, Moffitt, & Silva, 1998; Kern & Friedman, 2008; Moffitt et al., 2011). Thus, promoting self-regulation skills has been identified as a potentially fruitful target for intervention for children who are at-risk for academic failure (Ursache, Blair, & Raver, 2012). Emerging research suggests that not only are self-regulation skills at school-entry important for the development of academic skills, but *growth* in self-regulation skills may be associated with *growth* in academic skills (e.g., Fuhs et al., 2014; McClelland et al., 2007; Welsh et al., 2010). Although it appears that self-regulation is malleable and can be affected by pre-k experiences, little is known currently about whether a curriculum implemented on a broad basis could facilitate the development of self-regulation.

*Tools of the Mind* **Curriculum**

Tools of the Mind is based on an interactive sequence of change (shown in Figure 1) whereby teachers use assessment and scaffolding to tailor their use and modeling of specific tactics. Developed from a Vygotskian framework, Tools of the Mind (Bodrova & Leong, 2007) focuses on equipping children with cognitive "tools" for learning that they can then apply to the task of acquiring and sustaining academic knowledge and skills as well as behavioral competencies. These skills are commonly referred to as self-regulation skills, or skills associated with higher-order cognitive thinking that facilitate planning and goal-directed behavior. This approach follows from socio-cultural perspectives on child development that emphasize how children acquire skills and "cultural tools" (e.g., spoken and written language, pretend play, the use of numbers, diagrams and maps) in collaboration with knowledgeable others (e.g., Behne et al., 2008; Rogoff, Correa-Chavez, & Cotuc, 2005; Tomasello & Rakoczy, 2003). In the Tools approach, teachers model and use tactics such as concrete mediators (e.g., pictures or symbols), language (both speech and writing), and shared activities to scaffold children's learning. In the Tools approach, however, the tactics, mediators, forms of talk, and activities teachers use to foster learning are designed to themselves be part of what the student learns. Tools emphasizes that teachers use scaffolding techniques that will help children internalize the learning tools at the center of the curriculum; that is, to use the mediators introduced by the teacher and then create their own, to apply self-talk and writing, and to use shared activities and dramatic play in ways that help them attend, self-monitor, solve problems, plan, and remember.

The Tools curriculum was first implemented in pre-k classrooms in 1993 and has undergone substantial revision on the basis of field experience over the past 20 years. Tools is both a curriculum and a professional development program for teachers. As a curriculum, the focus has grown from 40 original activities to 60 or more Vygotskian-based activities designed to promote children's self-regulatory skills and cognitive development. In room arrangement, materials, and a balance between

whole group, small group and center-based activities, *Tools* is similar to a constructivist curriculum. However, *Tools* differs from other constructivist approaches in the prescribed and intentional role of the teacher in the classroom. The teacher's role is specifically prescribed for each major type of activity during the day (e.g., morning meeting, storybook reading, center-based time) through a series of delineated steps to be followed.

Tools is not a curriculum that can be "taken off the shelf" and implemented. Effective use of the curriculum depends on a depth of teacher understanding of the principles of children's learning and development as well as a reconceptualization of the teacher's role in facilitating children's self-regulation development. The *Tools* developers therefore strongly recommend two years of professional development together with in-classroom coaching, and their training packet is set up for this level of teacher contact.

The concepts behind the *Tools* approach are appealing to early childhood educators, especially those who have been concerned about the didactic nature of many early childhood classrooms (Hirsch-Pasek & Golinkoff, 2003; Miller & Almon, 2009). Despite the fact that until a few years ago there was only one small study of the effectiveness of the curriculum (Barnett et al., 2008; Diamond, Barnett, Thomas, & Munro, 2007), *Tools* has received enormous attention in the popular press, featured in the *New York Times*, the *Wall Street Journal*, on National Public Radio, and in the latest popular press book by Tough (2012), to name a few. School systems in Washington, DC, New Jersey, Chicago, and the entire country of Chile have received training from *Tools* staff to implement the approach.

This paper will focus on the results of an experimental evaluation of the *Tools of the Mind* pre-k curriculum, conducted in five school districts in two states. The aim of the *Tools of the Mind* curriculum is to enhance children's learning-related self-regulation skills within an instructional context that promotes the basic academic and social skills that prepare them for kindergarten and

beyond. The tools children learn from the curriculum should equip them to learn in subsequent later grade classrooms. To investigate the effectiveness of *Tools* in achieving these aims, we conducted a longitudinal cluster-randomized experiment to address the following questions:

1. Do children in *Tools of the Mind* classrooms improve more in literacy, math, social skills, and exhibit reduced behavior problems during the preschool year than children in "business as usual" comparison classrooms?

2. Do children in *Tools of the Mind* classrooms show greater gains in learning-related self-regulation than children in the comparison classrooms?

3. Are there differential effects of *Tools of the Mind* associated with characteristics of the children?

4. Do the effects of participating in a *Tools of the Mind* classroom sustain into kindergarten and first grade?

**Method**

**Design and Procedure**

Recruitment for the study occurred in two Southern states. Districts were selected for recruitment based on having an eligible public pre-k program and being willing to participate. Four school districts in one state and one school district in the second state participated in the study. The pre-k programs in the participating schools are funded through grants from their states and/or Title I, and all families had to meet the income guidelines for free or reduced-price lunch in order to enroll their children. Four of the school districts were relatively small ones located in suburban and rural areas surrounding a large city. One district was urban.

A large-scale cluster-randomized block design was employed to test the effectiveness of the *Tools of the Mind* curriculum compared to the typical curricula and practices occurring in the participating school systems. Because it was advantageous for conducting *Tools* professional development if all the pre-k teachers within a school were trained together and encouraged to

support each other during implementation, schools were the unit of randomization. This scheme was also intended to minimize interaction between experimental and comparison teachers that might have compromised the experimental contrast.

The trial was conducted over two years, with randomization occurring during the first year. The four smaller districts served as individual blocks. The 22 schools in the large, urban district were divided into five blocks based on the number of classrooms in each school and the experience of the teachers. Within each block, half the schools were assigned to the *Tools of the Mind* condition and half to the practice as usual comparison condition (with slight variations due to the uneven number of schools and classrooms in some districts). In the four smaller districts, all of the pre-k classrooms in each school then participated in the condition to which the school was assigned; in the large district, some classrooms in a few schools did not participate in either condition. Random assignment of schools to intervention and comparison conditions was performed in the summer of 2009, before the beginning of the 2009-2010 school year. Training of teachers and practice in the curriculum occurred during the 2009-2010 school year. During the first year, the implementation fidelity scheme was developed, but no data were collected on children.

Children were recruited for the research study beginning in the fall of 2010, during which teachers received additional training and implemented the program. The curriculum test thus occurred in the 2010-2011 school year. All procedures used in this research study were vetted and approved by the Vanderbilt University Institutional Review Board. Informed consent was obtained for all participating teachers who provided information about the children's classroom behaviors in a series of surveys. Parental consent was obtained for all participating children, and assent was obtained for all children at each assessment.

**Participants**

Sixty teachers participated in the study, with 32 assigned to the *Tools* condition and 28 assigned to the business-as-usual comparison condition. The comparison classrooms used a variety of curricula, with the modal one being *Creative Curriculum*. All but one teacher were female. Overall, teachers averaged 12 years of teaching experience, with seven years in preschool classrooms. All teachers were licensed and had at least a Bachelor's degree; over half had completed coursework toward or obtained a Master's degree. In addition, each classroom had at least one assistant.

In the 60 classrooms, 877 children (498 Tools; 379 comparison) were age-eligible for pre-k and consented to participate in the study in the fall of 2010. The consent rate in Tools classrooms was 88%, while the consent rate in comparison classrooms was 76%. Demographics for the 877 consented children are shown in Table 1. Overall, the sample of students was diverse in terms of ethnicity and language background, with multiple minority groups represented. Close to 30% of the students were English-language learners.

Attrition during the study was minimal. No teachers dropped out during the test year. Attrition of students over the course of the study was low and similar across Tools and comparison classrooms. Of the consented children, 866 had pretest scores on one or more direct assessments of achievement or self-regulation; we collected teacher reports of behavior on 862 children. The consented children who did not receive either a pretest or a teacher report in the fall of 2010 had either withdrawn from preschool prior to the assessment period or refused to complete one or more of the assessments. In the spring of 2011, 816 children had at least one direct assessment of achievement or self-regulation and teacher reports were received on 821 children. We obtained follow-up assessments on 810 children and teacher reports on 811 children in the spring of 2012 (when most children were completing kindergarten). In the spring of 2013 (at the end of most children's 1st grade year), we obtained assessments on 778 children and teacher reports on 779 children. There were no statistically significant differences in attrition by condition. Children who

were assessed at the end of pre-k or kindergarten did not differ significantly on any baseline variable from children who were not assessed. At the end 1$^{st}$ grade, children who were assessed had significantly higher baseline scores on one achievement measure (Spelling) and significantly lower baseline scores on another achievement measure (Applied Problems) than children who were not assessed. Sample sizes for each assessment by condition are shown in Tables 2-4.

**Instrumentation**

To assess the effects of the curriculum, we used a battery of standardized child achievement measures, a number of direct assessments of self-regulation, and teacher and assessor behavior rating measures. Achievement measures included seven subtests from the Woodcock Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001). The *Letter-Word Identification* subtest requires children to identify and pronounce letters and words. The *Spelling* subtest assesses children's prewriting skills, such as drawing lines and tracing, writing letters, and spelling orally presented words. The *Oral Comprehension* subtest assesses children's ability to understand a short passage by providing a missing word based on the syntactic and semantic cues provided in the sentence, which is read aloud by the examiner. The *Picture Vocabulary* subtest indexes expressive vocabulary. Children are asked to say aloud the noun corresponding to a picture. The *Academic Knowledge* subtest is given in three sections and measures children's factual knowledge of science, social studies, and humanities. The *Applied Problems* subtest assesses children's ability to solve small numerical and spatial problems presented verbally with accompanying pictures of objects. *Quantitative Concepts* assesses children's ability to point to or state answers to questions on number identification, sequencing, shapes, symbols, etc.; it measures aspects of quantitative reasoning and math knowledge.

Because of the specific focus of *Tools of the Mind* on the development of self–regulation, several direct assessments of self-regulation were selected to capture one or more components of self-regulation, including attentiveness, attention shifting, inhibitory control, persistence, and

working memory. The *Dimensional Change Card Sort* (DCCS; Zelazo, 2006) was used to assess

children's attention shifting capabilities. The task involves asking children to sort picture cards by

features depicted on the cards (i.e., color, shape) and has three levels, each requiring a shift of

attention. The four-level scoring recommended by Zelazo was used. Attention and visual-spatial

skills were assessed with the *Copy Design* task (Osborn, Butler, & Morris, 1984) in which children are

asked to copy eight simple geometric shapes that are increasingly complex. Each design has two

trials and total scores could range from 0 to 16 with higher scores indicating more accurate copies.

Working memory was assessed using the *Corsi* block-tapping task (Corsi, 1972). In this task,

children are asked to recall the order in which an examiner points to a series of blocks on a board in

an irregular order. Both forward and backward memory are tested with this task. The score for each

of the forward and backward versions was the number of blocks a child could correctly repeat.

Inhibitory control was assessed with two tasks: *Peg Tapping* (Diamond & Taylor, 1996) and

*Head-Toes-Knees-Shoulders* (HTKS; Ponitz, McClelland, Matthews, & Morrison, 2009). Both tasks ask

children to respond the opposite to the examiner's demonstration. In *Peg Tapping*, children are asked

to tap a wooden peg once when the examiner taps twice or tap twice when the examiner taps once.

HTKS requires children to respond to two oral prompts, "touch your head" and "touch your toes,"

then do the opposite in response to those prompts (i.e., touch their heads when the assessor says

"touch your toes" and vice versa). Touching knees and shoulders is added in a second trial and all

four parts are combined in the final, most difficult, trial. Scores for the task were the sum of

children's performance on the six practice items and the 20 testing items, with children receiving 0s

for incorrect responses, 2s for correct responses, and 1s for self-corrections (range = 0 to 52).

To capture social skills, self-regulation, and adaptive language as evidenced in the classroom,

teachers reported on children's classroom behaviors, self-regulation, and language ability. The

*Cooper-Farran Behavior Rating Scales* (Cooper & Farran, 1991) is a behaviorally anchored instrument

that indexes interpersonal skills and work-related skills in the classroom. The Interpersonal Skills subscale measures how well children get along with peers and with their teacher. The Work-Related Skills subscale includes items about independent work, compliance with and memory for instructions. The *Adaptive Language Inventory* (Feagans & Farran, 1983; Feagans, Fendt, & Farran, 1995) was also used to gauge children's comprehension and use of language in the classroom. Finally, assessors rated children's self-regulatory behaviors during the assessment sessions using the *Self-Regulation Assessor Rating* (SAR; Smith-Donald, Raver, Hayes, & Richardson, 2007). Three subscales from the SAR (attentiveness, impulsiveness, and concentration) were used in the fall and spring of pre-kindergarten; only the attentiveness scale was used in kindergarten and 1st grade.

## Fidelity of Implementation

The first year of the project was spent working with the *Tools* curriculum developers to create an instrument to measure the fidelity of implementation in the classrooms. *Tools of the Mind* is a dynamic curriculum; there is a timeline to determine when each of the 62 activities should be implemented during the year. Not all of them should be done every day, and they vary on when during the year they should be introduced. Each of the activities has between 3 and 12 steps; the steps prescribe the way the activity is to be implemented, which also changes during the year. For example, as children become familiar with an activity, easier steps may be eliminated and further extensions added.

Research staff attended the curriculum training sessions the first year, had several days-long meetings with the developers and trainers, and visited existing *Tools* classrooms in other locations to see the curriculum implemented by experienced teachers. From these activities, a detailed plan of the year for the curriculum was created that described each activity and its relevant mediators, when it should occur, its varying configurations of steps, and any teacher behaviors that should not occur (called "Should Nots"), all in a format that allowed observers to record the quality of each activity.

This initial work was tested via observations of experienced *Tools* teachers; suggestions and ideas to further revise the measure were discussed after these initial observations. The revised *Tools of the Mind Fidelity Instrument* was then operationalized using FileMaker Pro® database software and tablet computers. The fidelity instrument yielded detailed information about how much of the day *Tools* was implemented, which activities were enacted, and with how many steps.

**Data Collection Procedures**

*Assessment Procedures*

Children were individually assessed by trained and certified assessors in two approximately 20-minute sessions at the beginning and end of preschool and again at the end of kindergarten and first grade. Pre-k teachers rated the children's language, social skills, and classroom behavioral competencies in the fall (after 6 weeks of school) and again at the end of the school year in May. Kindergarten and 1$^{st}$ grade teachers rated the same skills in the late spring.

*Observation Procedures*

Daylong observations took place three times during the implementation year, in the fall, mid-winter, and spring. All classrooms were observed in both conditions. Observers were trained in a weeklong intensive session before the first observation. Two days of re-training took place before each of the subsequent observations. Reliability estimates were calculated from two observers spending a full day in 10% of the classrooms at each time point, with different classrooms being chosen at each time point. Estimates for reliabilities for whether an activity occurred and the use of mediators were calculated using Cohen's kappa and were .95 and .90, respectively.. Intraclass correlations were used to calculate reliabilities for the number of steps carried out correctly in an activity and the number of "should nots" observed, and were .98 and .93 respectively.

**Analytic Plan**

The impact models reported below employ three-level nested regression models, with students at Level 1, classrooms at Level 2, and district randomization blocks at Level 3. All analyses of achievement outcomes used the Woodcock Johnson $W$ scores, which are IRT scaled but not adjusted for age. All other outcomes remained in their raw score form. Each impact model included the following covariates: pretest, age at pretest assessment, interval between assessments, gender, ELL status, and IEP status. The pretest, age, and time interval covariates were grand-mean centered; the gender, ELL, and IEP status covariates were entered as dummy codes. For the subgroup analyses, we tested interactions between pretest, age, gender, ELL, and IEP status and experimental condition. The results for each outcome variable are reported separately. In addition, composite achievement and self-regulation scores were created using the procedures described below.

## Data Reduction

Because concurrent correlations among all seven achievement subtests at each assessment wave were statistically significant ($r$s ranged from .15 to .80), we also created an overall academic achievement composite score across all seven subtests. For each assessment wave, equally-weighted composite scores were created by transforming Woodcock Johnson subtest W scores into standardized z-scores and averaging across the z-scores obtained. Similarly, all correlations among the self-regulation assessments at each of the four measurement periods were statistically significant ($r$s ranged from .13 to .53). The scores on each direct assessment of self-regulation were therefore standardized and then averaged to create the composite.

**Results**

Prior to conducting analyses of treatment impacts, we performed a series of analyses on baseline variables as a randomization check. There were no statistically significant differences between *Tools* and comparison conditions with regard to teachers' level of education, $\chi^2$ (2, $n=60)=3.46$, $p=.18$; years teaching, $t(58)=.83$, $p=.41$; or years teaching pre-k, $t(58)=-.04$, $p=.97$.

The students in the *Tools* and comparison groups were similar on all demographic variables. There were no statistically significant differences between the two groups on gender, ELL status, IEP status, and proportion of students on free or reduced-price lunch. There were, however, small but statistically significant differences on age ($t=3.08$, $p<.05$) and ethnicity ($\chi^2=15.20$, $p<.05$). Students in the comparison group were slightly older than those in the *Tools* condition (by about 2½ weeks). The *Tools* condition classrooms had slightly higher proportions of Black and Asian students, while the comparison condition classrooms had proportionately larger numbers of Hispanic and multi-racial children. All analyses reported below employ age, gender, ELL and IEP status as covariates. We elected not to use ethnicity as a covariate because of concerns about reporting idiosyncrasies in the data obtained from some schools.

Finally, randomization checks were performed to compare *Tools* and comparison conditions on all baseline assessments and teacher ratings (see Supplementary Tables S1-S3). No significant pretreatment differences between the *Tools* and comparison conditions were found on any measure.

**Curriculum Effects**

Descriptives on the cases available for the achievement outcomes, the self-regulation outcomes, and the teacher reports at each measurement wave are presented respectively in Tables 2-4. The results of the impact analyses are shown in Table 5 for the achievement outcomes. The table presents the regression coefficients and standard errors for the treatment effects, as well as effect sizes. These coefficients show the impact of *Tools of the Mind* on achievement gains at the end of preschool, end of kindergarten, and end of 1ˢᵗ grade. Standardized mean difference effect sizes (*d*) were computed using the adjusted means reported in Table 2 and the unadjusted pooled standard deviation of the scores at the respective measurement wave.

Across the different achievement outcomes and the three outcome time points, few significant effects for curriculum condition are evident. Those that are significant favor the

comparison condition. At the end of the pre-k year, students in the comparison condition showed significantly greater gains in Oral Comprehension than students in *Tools of the Mind* classrooms. At the end of kindergarten, students who had been in comparison classrooms in pre-k achieved greater gains on Letter Word Identification and Quantitative Concepts than students in *Tools* classrooms. The composite achievement outcome was also statistically significant at the end of kindergarten, favoring the comparison condition. At the end of first grade, students in the comparison classrooms evidenced greater gains in Spelling than those in *Tools* classrooms but the previously found differences were not sustained.

The results for the battery of self-regulation assessments are shown in Table 6. Here again, we see few statistically significant effects, and those that are significant favor the comparison condition. There were no significant effects on the self-regulation measures at the end of the preschool year, on either the individual assessments or the composite. At the end of kindergarten, a significant negative effect appears for Corsi Backward Span, indicating that comparison group children experienced significantly greater gains on this measure over students who had participated in *Tools* classrooms. At the end of first grade, comparison students showed significantly greater gains on the Copy Design task and on the composite self-regulation score over *Tools* students.

Results for the teacher ratings of interpersonal skills, work-related skills, and language skills, and the assessors' ratings of attentiveness are shown in Table 7. There were no statistically significant differences between the *Tools* and comparison conditions on any of the teacher or assessor reports at any time point.

**Subgroup Analyses**

To answer our third research question on whether there were differential impacts for certain subgroups, we ran the same series of multi-level regression models described above for our main impact analyses, but included condition by subgroup interaction terms to identify any differential

effects. These analyses produced no consistent findings for any outcome, subgroup, or measurement wave. What is notable about the subgroup analyses is the lack of consistency in findings across the models. To illustrate this, we report the *p*-values for each of the interaction terms across all outcomes and waves in Table 8. Full regression models for the subgroup analyses and a detailed discussion of the results of those analyses are provided in the Supplementary Information (Tables S4-S21).

**Relationship of Fidelity of Implementation to Gains in Pre-K**

According to the three observations of curriculum fidelity taken during the year, there was variation among the teachers in the degree to which they implemented the curriculum. However, virtually all of the *Tools* teachers implemented substantial portions of the curriculum at the appropriate times and chose a variety of easy, medium, and difficult activities (see Table 9). Ambiguity from the curriculum developers about what constitutes full implementation makes it difficult to accurately appraise the level of implementation attained. However, observations of fidelity were consistent with independent ratings of teachers' implementation quality provided by *Tools* developers, trainers, coaches, classroom observers, and the teachers themselves (*r*s = .50, .60, .57, .49, .65, respectively). In addition, *Tools* activities were never observed in comparison classrooms.

Variations in fidelity of implementation measures across the group of 32 *Tools* teachers were not significantly associated with greater gains in achievement or self-regulation at any time point. Based on the manuals, it appears that teachers should be enacting about 22 *Tools* activities daily; some teachers in this study reached that goal. However, higher implementation fidelity by pre-k teachers was actually associated with smaller gains in composite achievement scores at the end of kindergarten (*b*=-.018, *se*=.006, *p*<.05) and with smaller gains in composite self-regulation scores at the end of 1ˢᵗ grade (*b*=-.016, *se*=.005, *p*<.05; see also Table S22 in the Supplementary Information).

In practical terms, these differences were relatively minor, comprising about a 3 standard score point difference on the Woodcock Johnson subtests between high and low implementers and less than a point difference on the raw self-regulation assessments between high and low implementers, but still in the opposite direction from expectations.

**Discussion**

Given the widespread interest and growing adoption of the *Tools of the Mind* curriculum, the curriculum developers and research team at the Peabody Research Institute agreed to partner in a rigorous experimental evaluation of the curriculum. A group of experienced pre-k teachers in five school systems participated in a year of training prior to full implementation, with follow-up training the second year and in-class coaching both years. The developers themselves participated in many of the training sessions; all sessions were led by experienced *Tools* trainers. Coaches were recruited at each site but were supervised through telephone calls and e-mails by the trainers for that site. In spite of what appeared to be fairly high levels of implementation fidelity, *Tools of the Mind* did not produce significant gains on any direct assessments of achievement or self-regulation or any teacher ratings of language, self-regulation, or social behavior at the end of pre-k.

This is not, however, a "no effects" study; *negative* effects were found on various outcomes for students in the *Tools* classrooms through the end of 1st grade. Students in comparison classrooms achieved significantly greater gains at the end of pre-k on Oral Comprehension; in addition, the comparison children saw slightly greater gains on most of the achievement and self-regulation outcomes in pre-k, though these impacts were not statistically significant, and the effect sizes were generally less than 0.10. By the end of kindergarten, however, these small gains in favor of comparison students had increased, with comparison students exhibiting significantly greater gains on the achievement composite, two achievement subtests (Letter Word Identification and Quantitative Concepts) and on Corsi Backward Span, a self-regulation task. Significant differences in

favor of the comparison group were also seen at the end of 1st grade on Spelling and Copy Design, as well as the composite self-regulation score. Subgroup analyses did not help us understand the findings. Treatment impacts were not consistently found across individual subgroups on any similar outcome measures, with several showing opposite impacts for the same subgroup on different outcomes.

How could a curriculum with such high expectations and so much "face validity" in its focus have produced such disappointing results? That *Tools of the Mind* would produce effects no different from business as usual in early childhood classrooms is surprising, but it is more surprising that children in the comparison classrooms consistently across the early grades made more positive gains on many subtests and the composites than *Tools* children. These results require careful consideration.

One of the first possibilities to explore is the structure of the curriculum and its expectations of teachers. When we began this project, the curriculum consisted of 42 activities. By the time our study began, the curriculum had grown to 60 activities, with three new ones added during the course of our two-year implementation. Moreover, each of the activities involved a series of steps to implement, and those steps changed across the year with some expected to drop out and others expected to be added. Teachers were provided four 2-inch thick manuals in which information about the structure of the day, and the activities and steps were described; in addition, teachers were given a separate manual of mediators they were to use. Mediators also were expected to be phased in and out across the year.

This is a dynamic curriculum, the most complex we have ever seen. While we observed most teachers making a good faith effort to implement the curriculum, one wonders how much time would be sufficient to become familiar enough with the activities to carry them out with confidence.

A related problem may be trying to fit the *Tools* expectations into the actual time available in a pre-kindergarten day. By the end of the year, the curriculum expectations are that the teachers will

provide an hour of Make Believe Play *plus* an hour of center time ("free play") in addition to numerous other small and whole group activities focused on literacy, math, and science. In a 6 hour day with approximately half of that time spent in required meals, naps and outdoor time, teachers simply could not implement all the activities the curriculum prescribed. However, we do not believe that our findings are a function of the teachers implementing too few activities. The curriculum does not indicate that all 60 or 63 activities be implemented every day; it seems that about 22 are prescribed at any one point. We found no relationship between the number of activities implemented and gains in any area by the children during over the pre-k year and negative associations with gain at the end of kindergarten and first grade. The activities required much speed and many transitions to implement even a modest number of them in a preschool day, possibly leaving little time for the children or the teachers to reflect and process the experiences.

Finally a broader issue results like these highlight is a further examination of the original theory of change. As Figure 1 shows, teachers' use of mediators, engaging children in writing and drawing, implementing dramatic play (of the *Tools* variety) and buddy activities are hypothesized to lead to the development in children of higher levels of attention, planning and problem solving, among others, which in turn would lead to more growth in children's academic skills of literacy, math and science as well as social skills. The results presented here indicated that teachers were implementing the theorized mechanisms of change but that children did not grow more in their self-regulatory skills nor in their academic or social competencies. The disconnect in the model appears to come between the set of activities teachers are to implement and the changes they are supposed to induce in the children.

The ways in which four year olds learn from instruction have recently been investigated in a series of laboratory and field experiments (though not classrooms) (Bonawitz, Shafto, Gweon, Goodman, Spelke, & Schulz, 2011; Butler & Markman, 2012; Gopnik & Wellman, 2012).

Collectively they demonstrate that four year olds are quite attuned to pedagogical cues to guide

inference and exploration; adults are powerful models for children's learning, so powerful that more

restricted demonstrations can actually inhibit exploration (Bonawitz et al., 2011). Bonawitz and

colleagues suggest that targeted pedagogy might actually inhibit novel application of skills or

(perhaps) the generalizing of a skill into new areas. The implication of their work is that targeted

pedagogy might not be the way to approach the internalization of skills as the *Tools* model asserts.

Like many early childhood classrooms where adults do more telling than demonstrating and much

of the instruction is often delivered in whole group formats (Phillips, Gormley, & Lowenstein,

2009), many of the activities in the *Tools* curriculum are targeted, delivered in whole group or half

group (i.e., groups of 8-10 with one guided by the teacher and one by the teacher assistant). These

instructional groupings may make it difficult to for children to generalize learning or do anything

other than mimic exactly what has been taught.

Given the positive findings in favor of the comparison classrooms, it is important to

understand what those classrooms were doing that was replaced by the *Tools* activities. One

candidate is differences in the way the two conditions approached center-based activities. During

center time, *Tools* teachers were expected to implement the *Tools* version of pretend play and but

then also allow time for center-based activities in which children were free to choose and explore

their own activities. This was difficult to do in an already limited school day, and our data indicate

that little time was spent in *Tools* classrooms in free play or center based activities. Pretend play was

done instead and might seem to be a good candidate for the kind of activity that enhances learning

and particularly the development of self-regulation, as the *Tools* curriculum model proposes. An

overview by Lillard and colleagues (Lillard et al., 2013), however, concluded that 40 years of research

examining pretend play provides little support that it has a crucial role in development. Lillard et al.

note the "unusual requirement" in the *Tools* version of pretend play specifically that children must

dictate ahead of time what role they will carry out in the play, and must not deviate from that role. As we observed, teachers, in fact, are encouraged to use the play plan as a management tool, reminding children what they were supposed to be doing if they got off track. Such scripted requirements may have inhibited children from internalizing some of the skills that were hypothesized to be developed via the pretend play activities. In comparison classrooms, by contrast, children were allowed much more time to explore materials freely chosen and on their own, an activity several reviews have asserted is beneficial to children's learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2010; Weisberg, Hirsh-Pasik, & Golinkoff, 2013).

A meta-analytic review of rigorous evaluations of various early childhood curricula has not established either that one curriculum is better than another or that having a particular curriculum produces more positive effects than business as usual (Darrow, 2013). Unfortunately, our cluster randomized control trial of an additional curriculum supports these conclusions. Hill, Beisiegel, and Jacob (2013) argue that the field of early childhood education needs a new approach to professional development with much more work in the early stages of development establishing that the elements presumed to be effective – inquiry-oriented learning approaches, collaborative learning, deep content focus – can be carried out and produce the effects intended. The field might further benefit from additional work on understanding how children internalize and translate skills acquired in one setting or application to others and developing effective ways that teachers might encourage such processes. Curricula that find a way to actively engage young children in the learning process, particularly with materials that lend themselves to exploration, should be more effective (Weisberg et al., 2013). Given the needs of poor children in the U.S. and the hope that prekindergarten experiences can address them, it seems that we should start the process of determining the elements that work as quickly as possible.

**References**

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2010). Does discovery-based

    instruction enhance learning? *Journal of Educational Psychology*, *103,* 1-18. doi: 10.1037/a0021017

Barnett, W. S., Kwanghee, J., 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* (3), 299-213. doi:10.1016/j.ecresq.2008.03.001

Behne, T., Carpenter, M., Grafenhain, M., Liebal, K., Liszkowski, U., Moll, H., . . . Wyman, E.

    (2008). Cultural learning and cultural creation. In N. Budwig, B. Sokos, U. Muller, & J.

    Carpendale (Eds.). *Social life and social knowledge: Toward a process account of development.* (pp. 65-

    101). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief

    understanding to emerging math and literacy ability in kindergarten. *Child Development, 78,* 647–

    663. doi:10.1111/j.1467-8624.2007.01019.x

Bodovski, K. & Farkas, G. (2009). Mathematics growth in early elementary school: The roles of

    beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108,*

    115-130. doi: 10.1086/525550

Bodrova, E., & Leong, D. J. (2007). *Tools of the Mind: The Vygotskian approach to early childhood education*

    (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N., Spelke, E., & Schulz, L. (2011). The double-

    edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition,*

    *120,* 322-330. doi: 10.1016/j.cognition.2010.10.001

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory

    factor analysis to understand executive control in preschool children: Sources of variation in

emergent mathematic achievement. *Developmental Science, 14,* 679–692. doi:10.1111/j.1467-7687.2010.01012.x

Butler, L. & Markman, E. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development, 83,* 1416-1428. doi: 10.1111/j.1467-8624.2012.01775.x

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

Caspi, A., Wright, B., Moffitt, T.E., & Silva, P.A. (1998). Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood. *American Sociological Review, 63*: 424-451. doi:10.2307/2657557

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk factors in kindergarten. *Early Childhood Research Quarterly, 3,* 1-19. doi:10.1016/0885-2006(88)90026-9

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

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Doctoral dissertation, McGill University). Retrieved from http://digitool.Library.McGill.CA:80/R/-?func=dbin-jumpfull&object_id=93903&silo_library=GEN01

Darrow, C. (2013). The effectiveness and precision of intervention fidelity measures in preschool intervention research. *Early Education and Development, 24*, 1137-1160. doi: 10.1080/10409289.2013.765786

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387-1388. doi:10.1126/science.1151148

Diamond, A. & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333,* 959-964. doi: 10.1126/science.1204529

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of

the abilities to remember what I said and to "Do as I say, not as I do". *Developmental*

*Psychobiology, 29,* 315–334. doi:10.1002/(SICI)1098-2302(199605)29:4_315:: AID-

DEV2_3.0.CO;2-T

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al.

(2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428-1446.

doi:10.1037/0012-1649.43.6.1428

Feagans, L. V., & Farran, D. C. (1983). *The Adaptive Language Inventory.* Chapel Hill, NC: Frank Porter

Graham Child Development Center.

Feagans, L., Fendt, K., & Farran, D.C. (1995). The effects of day care intervention on teachers'

ratings of the elementary school discourse skills in disadvantaged children. *International Journal*

*of Behavioral Development, 18,* 243-261 doi:10.1177/016502549501800204.

Fuhs, M., Nesbitt, K., Farran, D., & Dong, N. (2014). Longitudinal associations between executive

functioning and academic skills across content areas. *Developmental Psychology, 50,* 1698-1709.

doi:10.1037/a0036633

Garon, N., Bryson, S.E., & Smith, I.M. (2008). Executive function in preschoolers: A review using

an integrative framework. *Psychological Bulletin, 134*, 31-60. doi:10.1037/0033-2909.134.1.31

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian

learning mechanisms, and the theory theory. *Psychological Bulletin.* 138, 1085-1108. doi:

10.1037/a0028044

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on

cognitive development. *Developmental Psychology, 41,* 872-884. doi:10.1037/0012-1649.41.6.872

Hill, H., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus,

crossroads, and challenges. *Educational Researcher, 42,* 476-487. doi:10.3102/0013189x13512674

Hirsh-Pasek, K. & Golinkoff, R. (2003) *Einstein never used flash cards.* Emmaus, PA: Rodale Press.

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. doi:10.1080/00220970309602061

Hughes, C. (2011). Changes and challenges in 20 years of research into the development of

  executive functions. *Infant and Child Development, 20,* 251-271. doi: 10.1002/icd.736

Hustedt, J. T., Barnett, W. S., & Jung, K. (2008). *Longitudinal effects of the Arkansas Better Chance*

  *program: Findings from kindergarten and first grade.* Rutgers University, NJ: The National Institute

  for Early Education Research. Retrieved from http://nieer.org/pdf/

Jackson, R. McCoy, A. Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., Ross C., Schochet, P.,

  & Swank, P. (2007). *National evaluation of Early Reading First: Final report*, U.S. Department of

  Education, Institute of Education Sciences, Washington, DC: U.S. Government Printing

  Office.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., . . . Kristapovich, P. (2014). *The*

  *condition of education 2014* (NCES 2014-083). U.S. Department of Education, National Center

  for Education Statistics. Washington, DC. Retrieved May 24, 2014 from

  http://nces.ed.gov/pubsearch.

Kern, M. & Friedman, H. (2009). Early educational milestones as predictors of lifelong academic

  achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology, 30,*

  419-430.

Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of the Michigan School Readiness Program on young*

  *children's abilities at kindergarten entry.* Rutgers University, NJ: The National Institute for Early

  Education Research. Retrieved September 8, 2008, from

  http://nieer.org/resources/research/multistate/mi.pdf.

Lillard, A., Lerner, M., Hopkins, E., Dore, R., Smith, E., & Palmquist, C. (2013). The impact of

    pretend play on children's development: A review of the evidence. *Psychological Bulletin, 139*, 1–

    34. doi: 10.1037/a0029321

Lipsey, M., Hofer, K., Dong, N., Farran, D.C., & Bilbrey, C. (2013). *Evaluation of the Tennessee*

    *Voluntary Prekindergarten Program: Kindergarten and first grade follow-up results from the randomized*

    *control design.* Working paper available online (106 pages).

    http://peabody.vanderbilt.edu/research/pri/TN_VPK_Evaluation_Research_Report2_Appe

    ndices_August_2013.pdf

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do

    subsequent classroom experiences matter? *Early Childhood Research Quarterly, 22,* 18-38.

    doi:10.1016/j.ecresq.2006.10.002

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J.

    (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math

    skills. *Developmental Psychology, 43,* 947–959. doi:10.1037/0012-1649.43.4.947

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in

    preschool children. *Early Childhood Research Quarterly, 18*, 206-224. doi:10.1016/S0885-

    2006(03)00026-7

Miller, E., & Almon, J. (2009). *Crisis in the kindergarten: Why children need to play in school.* College Park,

    MD: Alliance for Childhood Press.

Moffitt, T., Arseneault, L., Belsky, D., Dickson, N., Hancox, R., Harrington, H., … Caspi, A. (2011).

    A gradient of childhood self-control predicts health, wealth, and public safety. *PNAS, 108,*

    2693-2698.

Noble, K. G., McCandliss, B. D., & Farah, M. (2007). Socioeconomic gradients predict individual

    differences in neurocognitive abilities. *Developmental Science, 10*, 464-480.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic

    status in kindergarten children. *Developmental Science, 8* (1), 74-87.

Osborn, A. F., Butler, N. R., & Morris, A. C. (1984). *The social life of Britain's five-year-olds: A report of the*

    *child health and education study*. London: Routledge & Kegan Paul.

Phillips, D., Gormley, W., & Lowenstein, A. (2009). Inside the pre-kindergarten door: Classroom

    climate and instructional time allocation in Tulsa's pre-k programs. *Early Childhood Research*

    *Quarterly, 24,* 213-228.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation

    of behavioral regulation and its contributions to kindergarten outcomes. *Developmental*

    *Psychology, 45,* 605–619. doi:10.1037/a0015365

Preschool Curriculum Evaluation Research Consortium (2008). *Effects of Preschool Curriculum Programs*

    *on School Readiness* (NCER 2008-2009). Washington, DC: National Center for Education

    Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC:

    U.S. Government Printing Office.

Rogoff, B., Correa-Chavez, M., Cotuc, M. (2005). A cultural/historical view of schooling in human

    development. In S. White & Pillemer, D. (Eds.), *Developmental psychology and social change: Research,*

    *history and policy* (pp. 225-263*)*. New York, NY: Cambridge University Press.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and

    concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based

    research. *Early Childhood Research Quarterly, 22* (2), 173-187. doi: 10.1016/j.ecresq.2007.01.002

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to

    shared to collective intentionality. *Mind & Language, 18,* 121-147.

Tough, P. (2012). *How children succeed*. New York: Houghton-Mifflin.

Ursache, A., Blair, C., & Raver, C. (2012). The promotion of self-regulation as a means of enhancing

school readiness and early achievement in children at risk for school failure. *Child Development Perspectives, 6,* 122-128. doi: 10.1111/j.1750-8606.2011.00209.x

U.S. Department of Health and Human Services, Administration for Children and Families (May 2005). *Head Start impact study: First year findings.* Washington, DC.

U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). *Head Start Impact Study. Final Report.* Washington, DC.

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. doi:10.1037/a0016738

Weisberg, D., Hirsh-Pasek, K., & Golinkoff, R. (2013). Guided play: Where curricular goals meet a playful pedagogy. *Mind, Brain, and Education, 7,* 104-112 doi: 10.1111/mbe.12015

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

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

Zill, N., & Resnick, G. (2005). Emergent literacy of low-income children in Head Start: Relationships with child and family characteristics, program factors and classroom quality. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of Early Literacy Research Vol. II* (pp. 347-371). New York: Guilford.

Table 1

*Demographics for Tools of the Mind and Comparison Children*

|  | Tools of the Mind | | Business as Usual | |
|---|---|---|---|---|
|  | *n* | *%* | *n* | *%* |
| Male | 261 | .53 | 218 | .58 |
| White | 192 | .39 | 157 | .41 |
| Black | 145 | .29 | 86 | .23 |
| Hispanic | 118 | .24 | 95 | .25 |
| Asian | 32 | .06 | 21 | .06 |
| Multi-racial | 4 | .01 | 16 | .04 |
| Other Minority | 7 | .01 | 4 | .01 |
| ELL (=yes) | 140 | .28 | 117 | .31 |
| IEP (=yes) | 68 | .14 | 58 | .15 |
| FRPL (=yes) | 329[a] | .86 | 293[a] | .88 |
|  | $\bar{x}$ | *sd* | $\bar{x}$ | *sd* |
| Age (months) at pretest[b] | 54.1 | 3.6 | 54.6 | 3.7 |
| Age (months) at posttest[c] | 61.5 | 3.5 | 62.0 | 3.7 |
| n students | 498 |  | 379 |  |
| n classrooms | 32 |  | 28 |  |

[a]Missing for 116 Tools children and 46 comparison children; percentages reflect percent of non-missing cases.
[b]n=494 Tools; 372 comparison.
[c]n=467 Tools; 349 comparison.

Table 2

*Standard Score Means, Standard Deviations, and Sample Sizes for the Woodcock Johnson Achievement Test Outcomes by Condition at Each Assessment*

| | Fall Pre-K | | Spring Pre-K | | Spring of K | | Spring 1st Grade | |
|---|---|---|---|---|---|---|---|---|
| | *N* | Mean (sd) | *N* | Mean (sd) | *N* | Mean (sd) | *N* | Mean (sd) |
| **Tools Condition** | | | | | | | | |
| Letter Word | 492 | 91.6 (12.9) | 465 | 100.0 (10.9) | 459 | 107.6 (11.3) | 443 | 108.1 (11.3) |
| Spelling | 492 | 80.2 (12.2) | 465 | 88.6 (13.9) | 459 | 99.4 (13.9) | 443 | 99.0 (15.0) |
| Academic Knowledge | 492 | 86.2 (19.4) | 465 | 92.7 (14.6) | 459 | 94.6 (12.5) | 443 | 94.5 (11.5) |
| Oral Comprehension | 492 | 90.0 (13.2) | 465 | 93.9 (13.9) | 459 | 97.1 (12.9) | 443 | 98.6 (11.7) |
| Picture Vocabulary | 492 | 91.9 (20.7) | 465 | 95.6 (13.8) | 459 | 94.9 (11.5) | 443 | 95.7 (10.5) |
| Applied Problems | 492 | 92.8 (15.5) | 465 | 98.6 (12.1) | 459 | 100.2 (12.7) | 443 | 99.6 (13.1) |
| Quantitative Concepts | 492 | 85.7 (11.9) | 465 | 92.3 (13.2) | 459 | 96.7 (11.3) | 443 | 94.9 (11.5) |
| Passage Comprehension | | | | | | | 443 | 96.2 (12.0) |
| **Comparison Condition** | | | | | | | | |
| Letter Word | 369 | 90.0 (13.2) | 348 | 100.3 (11.8) | 351 | 108.2 (12.2) | 335 | 108.1 (12.8) |
| Spelling | 369 | 78.0 (12.6) | 348 | 86.6 (15.2) | 351 | 100.3 (14.7) | 335 | 100.2 (16.0) |
| Academic Knowledge | 369 | 85.2 (19.2) | 348 | 92.5 (15.0) | 351 | 93.4 (13.5) | 335 | 94.2 (12.8) |
| Oral Comprehension | 369 | 89.1 (13.0) | 348 | 93.9 (15.0) | 351 | 96.4 (14.3) | 335 | 97.7 (12.5) |
| Picture Vocabulary | 369 | 91.5 (20.2) | 348 | 95.9 (13.9) | 351 | 94.6 (11.0) | 335 | 94.9 (10.9) |
| Applied Problems | 369 | 91.8 (14.7) | 348 | 97.8 (12.9) | 351 | 100.6 (12.5) | 335 | 98.7 (12.9) |
| Quantitative Concepts | 369 | 83.9 (12.0) | 348 | 91.7 (13.2) | 351 | 97.8 (12.2) | 335 | 93.9 (12.4) |
| Passage Comprehension | | | | | | | 335 | 95.4 (13.2) |

Table 3

*Means, Standard Deviations and Sample Sizes for the Self-Regulation Direct Assessment by Condition at Each Assessment*

| | Fall Pre-K | | Spring Pre-K | | Spring K | | Spring 1st Grade | |
|---|---|---|---|---|---|---|---|---|
| | *N* | Unadjusted Mean (sd) | *N* | Adjusted Mean (sd) | *N* | Adjusted Mean (sd) | *N* | Adjusted Mean (sd) |
| **Tools Condition** | | | | | | | | |
| Forward Span | 492 | 2.5 (1.3) | 465 | 3.0 (1.2) | 459 | 3.9 (1.1) | 443 | 4.6 (1.1) |
| Backward Span | 492 | 1.2 (1.2) | 465 | 1.6 (1.3) | 459 | 2.7 (1.4) | 443 | 3.7 (1.3) |
| DCCS | 492 | 1.3 (0.6) | 465 | 1.6 (0.6) | 459 | 1.9 (0.6) | 443 | 2.5 (0.9) |
| Copy Design | 492 | 1.1 (1.6) | 465 | 5.3 (2.8) | 459 | 7.8 (2.9) | 443 | 9.2 (3.1) |
| HTKS | 492 | 10.5 (13.6) | 464 | 21.8 (17.2) | 459 | 36.3 (13.7) | 443 | 43.8 (9.7) |
| Peg Tapping | 493 | 4.4 (5.8) | 465 | 9.4 (5.6) | 459 | 13.3 (4.0) | 443 | 14.7 (2.7) |
| **Comparison Condition** | | | | | | | | |
| Forward Span | 370 | 2.5 (1.3) | 348 | 3.1 (1.1) | 351 | 4.0 (1.1) | 335 | 4.7 (1.1) |
| Backward Span | 369 | 1.2 (1.1) | 348 | 1.6 (1.4) | 351 | 2.9 (1.3) | 335 | 3.8 (1.3) |
| DCCS | 371 | 1.3 (0.6) | 348 | 1.7 (0.6) | 351 | 2.0 (0.6) | 335 | 2.6 (0.9) |
| Copy Design | 369 | 1.0 (1.5) | 348 | 4.8 (2.8) | 351 | 7.7 (2.9) | 335 | 9.6 (2.9) |
| HTKS | 369 | 9.6 (12.2) | 348 | 22.1 (17.1) | 351 | 36.7 (14.3) | 335 | 44.7 (8.1) |
| Peg Tapping | 369 | 4.3 (5.8) | 348 | 9.3 (6.0) | 351 | 13.2 (4.2) | 335 | 15.0 (2.0) |

*Note.* DCCS=Dimensional Change Card Sort; HTKS=Head-Toes-Knees-Shoulders.

Table 4

*Means, Standard Deviations and Sample Sizes for the Teacher and Assessor Ratings by Condition at Each Assessment*

| | Fall Pre-K | | Spring Pre-K | | Spring K | | Spring 1st Grade | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Unadjusted | | Adjusted | | Adjusted | | Adjusted | |
| | N | Mean (sd) | N | Mean (sd) | N | Mean (sd) | N | Mean (sd) |
| Tools Condition | | | | | | | | |
| CFBRS Interpersonal Skills Scale | 492 | 5.2 (1.1) | 472 | 5.5 (1.1) | 459 | 5.7 (1.0) | 442 | 5.7 (1.1) |
| CFBRS Work-Related Skills Scale | 492 | 4.5 (1.2) | 472 | 5.0 (1.2) | 459 | 5.0 (1.2) | 442 | 4.9 (1.2) |
| Adaptive Language Inventory | 492 | 2.9 (0.8) | 472 | 3.3 (0.8) | 459 | 3.2 (0.8) | 442 | 3.2 (0.8) |
| Assessor Ratings of Attention | 494 | 2.4 (0.6) | 467 | 2.5 (0.6) | 459 | 2.7 (0.5) | 443 | 2.7 (0.5) |
| Comparison Condition | | | | | | | | |
| CFBRS Interpersonal Skills Scale | 370 | 5.4 (1.1) | 349 | 5.4 (1.1) | 352 | 5.6 (1.0) | 337 | 5.7 (1.0) |
| CFBRS Work-Related Skills Scale | 370 | 4.7 (1.1) | 349 | 5.0 (1.1) | 352 | 4.9 (1.2) | 337 | 4.9 (1.3) |
| Adaptive Language Inventory | 370 | 3.0 (0.8) | 349 | 3.2 (0.9) | 352 | 3.2 (0.8) | 337 | 3.2 (0.8) |
| Assessor Ratings of Attention | 371 | 2.4 (0.6) | 349 | 2.5 (0.6) | 351 | 2.7 (0.5) | 335 | 2.7 (0.5) |

*Note.* CFBRS=Cooper-Farran Behavior Rating Scales.

Table 5

*Impact of Tools of the Mind on Academic Achievement at the End of Pre-K, Kindergarten, and 1ˢᵗ Grade*

| Variable | End of Pre-K | | | End of Kindergarten | | | End of First Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| | *b* | *se* | *d* | *b* | *se* | *d* | *b* | *se* | *d* |
| Letter Word | -2.87 | 1.83 | -0.12 | -4.40* | 1.64 | -0.17 | -3.09 | 1.86 | -0.11 |
| Spelling | 0.74 | 2.20 | 0.03 | -3.64 | 1.82 | -0.16 | -4.10* | 1.73 | -0.17 |
| Academic Knowledge | -0.83 | 0.96 | -0.05 | 0.42 | 0.85 | 0.03 | -0.23 | 0.88 | -0.02 |
| Oral Comprehension | -1.60* | 0.74 | -0.10 | -0.59 | 0.83 | -0.04 | 0.41 | 0.77 | 0.03 |
| Picture Vocabulary | -1.08 | 0.65 | -0.07 | -0.23 | 0.58 | -0.02 | 0.08 | 0.65 | 0.01 |
| Applied Problems | -0.28 | 1.11 | -0.01 | -1.50 | 0.95 | -0.09 | 0.25 | 1.04 | 0.01 |
| Quantitative Concepts | -1.12 | 1.00 | -0.08 | -2.67* | 0.87 | -0.21 | -0.40 | 0.82 | -0.03 |
| Composite Achievement | -0.33 | 0.27 | -0.06 | -0.81* | 0.29 | -0.15 | -0.37 | 0.29 | -0.07 |
| Passage Comprehension | | | | | | | -0.72 | 1.43 | -0.04 |

*Note.* Coefficients in the table are unstandardized regression coefficients from multi-level regression models. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. The *d* column shows the standardized mean difference effect size for the impact estimate.
* $p < .05$.

Table 6

*Impact of Tools of the Mind on Self-Regulation at the End of Pre-K, Kindergarten, and 1ˢᵗ Grade*

| Variable | End of Pre-K | | | End of Kindergarten | | | End of First Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| | *b* | *se* | *d* | *b* | *se* | *d* | *b* | *se* | *d* |
| Forward Span | -0.09 | 0.08 | -0.08 | -0.08 | 0.08 | -0.07 | -0.07 | 0.08 | -0.06 |
| Backward Span | -0.08 | 0.09 | -0.06 | -0.22* | 0.10 | -0.16 | -0.08 | 0.09 | -0.06 |
| DCCS | -0.03 | 0.04 | -0.05 | -0.05 | 0.05 | -0.09 | -0.05 | 0.08 | -0.06 |
| Copy Design | 0.43 | 0.22 | 0.15 | 0.06 | 0.20 | 0.02 | -0.39* | 0.20 | -0.13 |
| HTKS | -0.21 | 1.08 | -0.01 | -0.41 | 0.97 | -0.03 | -0.89 | 0.69 | -0.10 |
| Peg Tapping | 0.10 | 0.41 | 0.02 | 0.19 | 0.27 | 0.05 | -0.30 | 0.17 | -0.13 |
| Composite Self-Regulation | -0.09 | 0.23 | -0.02 | -0.24 | 0.21 | -0.06 | -0.59* | 0.24 | -0.15 |

*Note.* Coefficients in the table are unstandardized regression coefficients from multi-level regression models. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. The *d* column shows the standardized mean difference effect size for the impact estimate.
* $p<.05$.

Table 7

*Impact of Tools of the Mind on Teacher and Assessor Ratings at the End of Pre-K, Kindergarten, and 1ˢᵗ Grade*

| Variable | End of Pre-K | | | End of Kindergarten | | | End of First Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| | *b* | *SE* | *d* | *b* | *SE* | *d* | *b* | *SE* | *d* |
| CFBRS Interpersonal Skills Scale | 0.11 | 0.09 | 0.10 | 0.04 | 0.10 | 0.04 | 0.02 | 0.10 | 0.02 |
| CFBRS Work-Related Skills Scale | 0.09 | 0.11 | 0.07 | 0.10 | 0.09 | 0.09 | -0.03 | 0.10 | -0.03 |
| Adaptive Language Inventory | 0.09 | 0.08 | 0.11 | 0.03 | 0.06 | 0.04 | 0.02 | 0.07 | 0.03 |
| Self-Regulation Assessor Ratings | 0.00 | 0.05 | -0.01 | -0.03 | 0.04 | -0.05 | -0.04 | 0.04 | -0.09 |

*Note.* Coefficients in the table are unstandardized regression coefficients from multi-level regression models. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. The *d* column shows the standardized mean difference effect size for the impact estimate. CFBRS=Cooper-Farran Behavior Rating Scales.
* $p<.05$.

Table 8

*Subgroup Analysis: p-values from Tests of Interactions*

| Variable | End of Pre-K | | | | | End of Kindergarten | | | | | End of 1st Grade | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Condition by | | | | | Condition by | | | | | Condition by | | | |
| | Pretest | Gender | ELL | IEP | Age | Pretest | Gender | ELL | IEP | Age | Pretest | Gender | ELL | IEP | Age |
| Letter Word | 0.660 | 0.078 | 0.900 | 0.532 | *0.012* | 0.692 | 0.484 | *0.043* | 0.861 | 0.500 | 0.930 | 0.394 | 0.184 | 0.691 | 0.530 |
| Spelling | 0.274 | 0.308 | 0.353 | 0.357 | 0.243 | 0.811 | 0.911 | 0.211 | 0.347 | 0.312 | 0.852 | 0.192 | 0.832 | 0.603 | 0.508 |
| Academic Knowledge | 0.176 | 0.897 | 0.236 | 0.831 | 0.858 | 0.891 | 0.565 | 0.815 | 0.693 | 0.622 | 0.198 | 0.989 | 0.335 | 0.168 | 0.445 |
| Oral Comprehension | 0.183 | 0.949 | 0.872 | 0.493 | 0.156 | *0.003* | 0.206 | 0.830 | 0.284 | 0.108 | 0.147 | 0.291 | 0.843 | 0.567 | 0.860 |
| Picture Vocabulary | 0.078 | 0.183 | 0.280 | 0.378 | 0.428 | 0.850 | 0.540 | 0.230 | 0.320 | 0.165 | 0.254 | *0.039* | 0.993 | 0.423 | 0.190 |
| Applied Problems | 0.819 | 0.326 | 0.892 | 0.358 | 0.692 | 0.943 | 0.739 | 0.378 | 0.551 | 0.922 | 0.335 | 0.516 | 0.939 | 0.867 | 0.963 |
| Quantitative Concepts | 0.616 | 0.202 | 0.845 | 0.630 | 0.245 | 0.793 | 0.539 | 0.742 | 0.907 | 0.911 | 0.564 | 0.329 | 0.474 | 0.973 | 0.441 |
| Passage Comprehension | | | | | | | | | | | 0.612 | 0.267 | 0.749 | 0.559 | 0.391 |
| Forward Span | 0.733 | 0.906 | *0.012* | 0.827 | 0.979 | 0.120 | 0.893 | 0.100 | 0.547 | 0.207 | 0.543 | 0.156 | 0.610 | 0.247 | 0.497 |
| Backward Span | 0.882 | *0.037* | 0.082 | 0.655 | 0.402 | 0.070 | 0.883 | 0.689 | 0.388 | 0.274 | 0.885 | 0.353 | 0.388 | 0.342 | 0.453 |
| DCCS | 0.668 | *0.041* | 0.855 | 0.214 | 0.671 | *0.027* | 0.180 | 0.248 | 0.155 | 0.556 | 0.395 | 0.605 | 0.985 | *0.044* | 0.635 |
| Copy Design | 0.072 | 0.578 | 0.356 | 0.785 | 0.267 | 0.795 | 0.119 | 0.843 | *0.036* | 0.321 | 0.369 | 0.449 | 0.238 | 0.266 | 0.810 |
| HTKS | 0.536 | 0.105 | 0.420 | 0.093 | 0.970 | 0.641 | 0.995 | 0.994 | 0.322 | 0.645 | 0.159 | 0.237 | 0.933 | 0.639 | 0.606 |
| Peg Tapping | 0.305 | 0.877 | 0.102 | 0.219 | 0.528 | 0.339 | 0.908 | *0.021* | 0.808 | 0.981 | 0.429 | 0.719 | 0.299 | 0.259 | 0.468 |
| Interpersonal Skills | *0.003* | 0.651 | 0.374 | *0.037* | 0.827 | 0.114 | 0.658 | 0.449 | 0.838 | 0.155 | 0.576 | 0.309 | 0.340 | 0.675 | 0.119 |
| Work-related Skills | 0.164 | 0.566 | 0.550 | *0.038* | 0.270 | 0.715 | 0.785 | 0.321 | 0.146 | 0.236 | 0.889 | 0.588 | 0.305 | 0.191 | 0.241 |
| Adaptive Language | 0.326 | 0.399 | 0.587 | 0.905 | 0.310 | 0.716 | 0.079 | 0.525 | 0.460 | 0.942 | 0.387 | 0.079 | 0.353 | 0.584 | 0.993 |
| Assessor Ratings | 0.881 | 0.721 | 0.913 | 0.591 | 0.827 | 0.566 | 0.599 | 0.059 | 0.531 | 0.141 | 0.840 | 0.470 | 0.923 | 0.138 | 0.688 |

*Note.* Table shows the p-values associated with the coefficients for the condition x subgroup interactions from multi-level models. See the online supplementary materials for full details on the subgroup analyses.
DCCS=Dimensional Change Card Sort; HTKS=Head-Toes-Knees-Shoulders.

Table 9

*Means, Standard Deviations, and Ranges for Fidelity of Implementation by Observation*

|  | Observation 1 | | | Observation 2 | | | Observation 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *Min* | *Max* | *Mean (SD)* | *Min* | *Max* | *Mean (SD)* | *Min* | *Max* | *Mean (SD)* |
| Activities (#) | 5 | 22 | 13.97 (3.57) | 4 | 22 | 14.91 (3.77) | 6 | 20 | 14.84 (3.34) |
| Steps (#) | 16 | 78 | 53.66 (16.18) | 11 | 95 | 61.81 (18.27) | 15 | 91 | 62.00 (17.13) |
| Mediators (#) | 12 | 46 | 30.75 (7.96) | 12 | 48 | 32.31 (7.84) | 7 | 44 | 32.47 (8.21) |
| Should Nots(#) | 0 | 8 | 4.78 (2.57) | 0 | 9 | 3.19 (3.06) | 0 | 12 | 4.44 (2.91) |

*Note.* N = 32 *Tools of the Mind* Classrooms; Correlation among activities, steps, and mediators aggregated across the three observations were high, rs > .92. The should nots were not related to the other three fidelity variables, rs< .18. While the table provides descriptives by time point, these scores were aggregated for the analysis of fidelity impacts shown in Table 10, quantified as the number of activities implemented. The correlations across the three observations ranged from .53 to .72 for activities, .61 to .78 for steps, .58 to .74 for mediators, and .38 to .50 for should nots.

**Tools of the Mind Curriculum
Theory of Change**

**Target Skills**

**Developmental Functioning for "Target Skill X"**

Content: Literacy, Math/Science, Social Competence

Teacher Tactics
- External mediators
- Language
- Shared activities

Tools of the Mind
- Mediators
- Language
  - Self-talk
  - Writing, drawing
  - Decontextualized speech
- Shared activities
  - Dramatic play
  - Buddy activities

**Dynamic Assessment Scaffolding**

Learning-related Self-regulation
- Attending
- Planning
- Thinking
- Remembering
- Practicing
- Problem solving

Language and Literacy
- Oral language
- Decontextualized speech
- Phonemic awareness
- Phonics
- Vocabulary
- Listening comprehension
- Print awareness

Math and Science
- Numbers and operations
- Geometry
- Measurement
- Data analysis
- Algebra
- Science

Social/Emotional Competence
- Social skills
- Emotion regulation
- Perspective taking
- Self-control
- Social problem solving skills

Figure 1. Theory of Change for the Tools of the Mind Pre-K Curriculum

Supplementary Information

Achievement and Self-Regulation in Pre-Kindergarten Classrooms:

Effects of the *Tools of the Mind* Curriculum

Table S1

*Baseline Equivalence on the Woodcock Johnson Achievement Measures*

| Parameter | LWW | SPW | OCW | PVW | AKW | APW | QCW |
|---|---|---|---|---|---|---|---|
| Fixed Effects | | | | | | | |
| Intercept | 314.56* | 337.07* | 438.96* | 450.33* | 426.60* | 380.18* | 403.12* |
| | (-1.91) | (-1.97) | (-2.16) | (-3.07) | (-3.06) | (-3.76) | (-1.35) |
| **Condition** | **1.7** | **1.29** | **-0.42** | **-1.23** | **-0.5** | **-1.09** | **0.62** |
| | (-2.5) | (-1.92) | (-1.33) | (-2.21) | (-2.2) | (-2.62) | (-1.07) |
| Variance Components | | | | | | | |
| Residual | 623.56* | 502.27* | 207.60* | 526.21* | 440.74* | 855.51* | 150.85* |
| | (-31.15) | (-25.04) | (-10.38) | (-26.26) | (-22.02) | (-42.63) | (-7.51) |
| District Block | 6.96 | 19.06 | 34.03 | 62.58 | 62.25 | 96.2 | 11.23 |
| | (-10.27) | (-13.78) | (-19.02) | (-36.7) | (-37.26) | (-56.92) | (-7.19) |
| Classroom | 46.31* | 17.37 | 10.53* | 32.80* | 37.99* | 38.03* | 5.65 |
| | (-18.31) | (-10.42) | (-5.24) | (-14.08) | (-14.2) | (-19.3) | (-3.18) |
| n Tools | 492 | 492 | 492 | 492 | 492 | 492 | 492 |
| n Comparison | 369 | 369 | 369 | 369 | 369 | 369 | 369 |

*Note.* The top portion of the table shows unstandardized regression coefficients with standard errors in parentheses below them; the bottom portion of the table shows variance component estimates with standard errors in parentheses. LWW=Letter-Word; SPW=Spelling; OCW=Oral Comprehension; PVW=Picture Vocabulary; AKW=Academic Knowledge; APW=Applied Problems; QCW=Quantitative Concepts. All analyses used W scores. *p<.05.

Table S2

*Baseline Equivalence on the Self-Regulation Assessments*

| Parameter | DCCS | Copy Design | Forward Span | Backward Span | Peg Tap | HTKS |
|---|---|---|---|---|---|---|
| Fixed Effects | | | | | | |
| Intercept | 1.30* | 1.09* | 2.54* | 1.15* | 4.33* | 10.17* |
| | (-0.05) | (-0.07) | (-0.09) | (-0.07) | (-0.62) | (-1.28) |
| **Condition** | **0.01** | **0.09** | **-0.0004** | **-0.04** | **-0.09** | **0.48** |
| | (-0.04) | (-0.11) | (-0.1) | (-0.09) | (-0.46) | (-0.98) |
| Variance Components | | | | | | |
| Residual | 0.32* | 2.43* | 1.52* | 1.26* | 30.66* | 159.21* |
| | (-0.02) | (-0.12) | (-0.08) | (-0.06) | (-1.53) | (-7.98) |
| District Block | 0.02 | 0.00 | 0.03 | 0.02 | 2.49 | 10.55 |
| | (-0.01) | (0.00) | (-0.03) | (-0.02) | (-1.53) | (-6.78) |
| Classroom | 0.00 | 0.00 | 0.03 | 0.03 | 0.82 | 2.48 |
| | (0.00) | (0.00) | (-0.03) | (-0.02) | (-0.62) | (-3.03) |
| n Tools | 492 | 492 | 492 | 492 | 493 | 492 |
| n Comparison | 371 | 369 | 370 | 369 | 369 | 369 |

*Note.* The top portion of the table shows unstandardized regression coefficients with standard errors in parentheses below them; the bottom portion of the table shows variance component estimates with standard errors in parentheses. DCCS=Dimensional Change Card Sort; HTKS=Head-Toes-Knees-Shoulders.
*p<.05.

Table S3

*Baseline Equivalence on the Teacher and Assessor Ratings*

| Parameter | IPS | WRS | ALI | SAR |
|---|---|---|---|---|
| Fixed Effects | | | | |
| Intercept | 5.17* | 4.47* | 2.87* | 2.41* |
| | (-0.08) | (-0.09) | (-0.07) | (-0.07) |
| **Condition** | **-0.19** | **-0.21** | **-0.14** | **0.01** |
| | (-0.12) | (-0.13) | (-0.09) | (-0.07) |
| Variance Components | | | | |
| Residual | 1.00* | 1.16* | 0.55* | 0.34* |
| | (-0.05) | (-0.06) | (-0.03) | (-0.02) |
| District Block | 0.00 | 0.00 | 0.01 | 0.03 |
| | (0.00) | (0.00) | (-0.01) | (-0.02) |
| Classroom | 0.13* | 0.18* | 0.09* | 0.04* |
| | (-0.04) | (-0.05) | (-0.03) | (-0.01) |
| n Tools | 492 | 492 | 492 | 494 |
| n Comparison | 370 | 370 | 370 | 371 |

*Note.* The top portion of the table shows unstandardized regression coefficients with standard errors in parentheses; the bottom portion of the table shows variance component estimates with standard errors in parentheses. IPS=Cooper-Farran Interpersonal Skills; WRS=Cooper-Farran Work-related Skills; ALI=Adaptive Language Inventory; SAR=Self-Regulation Assessor Rating.
*$p < .05$.

**Subgroup Analyses**

Complete results for the subgroup analyses are reported in Tables S4-S21. Across the 17 outcomes and three measurement waves, three pretest by intervention condition interactions were statistically significant, for Oral Comprehension and DCCS at kindergarten and for teacher-rated Interpersonal Skills at the end of pre-k. Comparison group children with low pretests achieved greater gains in Oral Comprehension and teacher ratings of Interpersonal Skills than *Tools* children, while *Tools* children with low pretests on DCCS achieved greater gains than comparison children. Three gender by condition interactions were statistically significant, on Corsi Backward Span and DCCS at the end of prekindergarten and on Picture Vocabulary at the end of 1st grade. Girls in *Tools* classrooms experienced significantly smaller gains on Corsi Backward Span than girls in comparison classrooms. Separate treatment impact estimates for boys and girls on the DCCS and Picture Vocabulary were not significant; the significant interactions were largely due to gender differences observed within *Tools* classrooms on the DCCS and in the comparison classrooms on Picture Vocabulary.

Three ELL by condition interactions were significant, on Corsi Forward Span at the end of prekindergarten and on Letter Word Identification and Peg Tapping at the end of kindergarten. At the end of prekindergarten, ELL students in comparison classrooms exhibited significantly greater gains in Corsi Forward Span than ELL students in *Tools* classrooms. Native English speaking comparison students achieved significantly greater gains in Letter Word Identification scores at the end of kindergarten than native English speaking students who had been in *Tools* classrooms. On Peg Tapping, the interaction was in the opposite direction. Native English speaking *Tools* students achieved significantly greater gains in Peg Tapping scores at the end of kindergarten than native English speaking students who had been in comparison classrooms.

Four IEP by condition interactions were significant, on Interpersonal and Work-related Skills at the end of pre-k, on Copy Design at the end of kindergarten, and on DCCS at the end of 1st grade. Separate treatment impact estimates for students with and without IEPs on both Interpersonal Skills and Work-related Skills were not statistically significant; the significant interactions on these teacher ratings were largely due to differences between the subgroups observed within *Tools* classrooms. On the Copy Design test, IEP students in comparison classrooms achieved greater gains at the end of kindergarten that IEP students who had been in *Tools* classrooms. Separate treatment impact estimates for students with and without IEPs on the DCCS were not significant; the significant interaction was largely due to differences between subgroups observed within comparison classrooms.

Finally, one age by condition interaction was significant, on Letter Word Identification at the end of pre-k, with older students in *Tools* classrooms achieving greater gains than older students in comparison classrooms and younger students in comparison classrooms achieving greater gains than younger students in *Tools* classrooms.

Table S4

*Results of Subgroup Analyses for Letter Word*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 349.48* | 1.77 | 401.21* | 2.21 | 442.90* | 2.22 |
| Condition=Tools | -3.59 | 2.38 | -3.13 | 2.49 | -0.97 | 2.93 |
| Pretest | 0.52* | 0.04 | 0.53* | 0.05 | 0.45* | 0.05 |
| Gender=Male | -2.37 | 1.99 | -1.43 | 2.36 | 3.79 | 2.88 |
| ELL=Yes | 3.61 | 2.24 | 2.78 | 2.66 | 3.42 | 3.18 |
| IEP=Yes | -2.50 | 2.80 | -7.51* | 3.33 | -15.97* | 4.01 |
| Age at Pretest | -0.42 | 0.26 | 0.12 | 0.32 | 0.11 | 0.38 |
| Interval from Pretest | -1.99 | 1.69 | 3.46* | 1.65 | 4.15 | 2.20 |
| Condition*Pretest | 0.02 | 0.05 | 0.02 | 0.06 | 0.01 | 0.07 |
| Condition*Gender | -4.57 | 2.59 | -2.16 | 3.09 | -3.19 | 3.75 |
| Condition*ELL | -0.38 | 3.06 | 7.16* | 3.53 | 5.64 | 4.24 |
| Condition*IEP | -2.34 | 3.74 | -0.77 | 4.41 | 2.11 | 5.31 |
| Condition*Age | 0.89* | 0.35 | 0.29 | 0.42 | 0.32 | 0.51 |

* $p<.05$

Table S5

*Results of Subgroup Analyses for Spelling*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 372.20* | 2.25 | 424.72* | 2.04 | 455.09* | 2.06 |
| Condition=Tools | -1.23 | 2.84 | -5.98* | 2.56 | -2.87 | 2.65 |
| Pretest | 0.59* | 0.05 | 0.32* | 0.05 | 0.33* | 0.06 |
| Gender=Male | -6.17* | 2.40 | -3.61 | 2.35 | -0.09 | 2.58 |
| ELL=Yes | 7.72* | 2.59 | 5.07* | 2.56 | 2.57 | 2.78 |
| IEP=Yes | -3.17 | 3.31 | -9.26* | 3.25 | -11.23* | 3.56 |
| Age at Pretest | 0.11 | 0.32 | 0.28 | 0.32 | 0.32 | 0.34 |
| Interval from Pretest | -2.09 | 2.14 | 2.58 | 1.60 | -0.27 | 2.04 |
| Condition*Pretest | -0.08 | 0.07 | -0.02 | 0.07 | -0.01 | 0.07 |
| Condition*Gender | -3.16 | 3.10 | -0.34 | 3.05 | -4.36 | 3.34 |
| Condition*ELL | -3.24 | 3.49 | -4.25 | 3.40 | 0.78 | 3.66 |
| Condition*IEP | -4.06 | 4.41 | -4.07 | 4.33 | 2.45 | 4.71 |
| Condition*Age | 0.50 | 0.43 | 0.43 | 0.42 | 0.31 | 0.46 |

*p<.05.

Table S6

*Results of Subgroup Analyses for Academic Knowledge*

|  | End of Pre-K | | End of Kindergarten | | End of First Grade | |
| --- | --- | --- | --- | --- | --- | --- |
| Parameter | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 443.11* | 1.14 | 455.82* | 1.05 | 467.95* | 1.01 |
| Condition=Tools | -1.43 | 1.31 | 0.79 | 1.29 | 1.12 | 1.26 |
| Pretest | 0.62* | 0.03 | 0.42* | 0.03 | 0.40* | 0.03 |
| Gender=Male | -1.85 | 1.02 | -0.18 | 1.12 | 0.12 | 1.02 |
| ELL=Yes | 0.17 | 1.43 | 0.26 | 1.56 | 1.54 | 1.44 |
| IEP=Yes | -3.10* | 1.46 | -3.46* | 1.61 | -4.98* | 1.46 |
| Age at Pretest | 0.01 | 0.13 | -0.02 | 0.15 | -0.06 | 0.14 |
| Interval from Pretest | 2.09* | 1.00 | 0.95 | 0.79 | 1.89* | 0.91 |
| Condition*Pretest | -0.05 | 0.04 | 0.01 | 0.04 | -0.05 | 0.04 |
| Condition*Gender | 0.17 | 1.33 | 0.85 | 1.47 | -0.02 | 1.33 |
| Condition*ELL | -2.31 | 1.95 | 0.49 | 2.11 | 1.86 | 1.93 |
| Condition*IEP | -0.42 | 1.95 | 0.84 | 2.13 | 2.66 | 1.93 |
| Condition*Age | 0.03 | 0.18 | -0.10 | 0.20 | 0.14 | 0.18 |

*p<.05.

Table S7

*Results of Subgroup Analyses for Oral Comprehension*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 449.78* | 1.11 | 463.25* | 1.17 | 474.39* | 0.99 |
| Condition=Tools | -2.08 | 1.22 | -1.28 | 1.32 | 0.12 | 1.30 |
| Pretest | 0.74* | 0.05 | 0.72* | 0.05 | 0.58* | 0.05 |
| Gender=Male | 0.38 | 1.13 | 0.13 | 1.18 | 1.19 | 1.18 |
| ELL=Yes | -5.20* | 1.57 | -0.41 | 1.64 | 2.48 | 1.65 |
| IEP=Yes | -0.58 | 1.65 | -2.14 | 1.69 | -3.26 | 1.69 |
| Age at Pretest | -0.08 | 0.15 | -0.23 | 0.16 | -0.09 | 0.15 |
| Interval from Pretest | 0.27 | 0.89 | -1.07 | 0.84 | -1.93* | 0.96 |
| Condition*Pretest | -0.08 | 0.06 | -0.18* | 0.06 | -0.09 | 0.06 |
| Condition*Gender | -0.09 | 1.48 | -1.95 | 1.54 | -1.62 | 1.53 |
| Condition*ELL | 0.33 | 2.07 | 0.47 | 2.17 | 0.43 | 2.16 |
| Condition*IEP | -1.48 | 2.16 | -2.40 | 2.24 | -1.27 | 2.23 |
| Condition*Age | 0.28 | 0.20 | 0.33 | 0.21 | 0.04 | 0.21 |

*p<.05.

Table S8

*Results of Subgroup Analyses for Picture Vocabulary*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 462.86* | 0.79 | 469.19* | 0.73 | 476.72* | 0.99 |
| Condition=Tools | -1.20 | 1.06 | -0.39 | 0.98 | 0.67 | 1.03 |
| Pretest | 0.41* | 0.03 | 0.27* | 0.02 | 0.26* | 0.02 |
| Gender=Male | 0.75 | 0.98 | 1.39 | 0.91 | 2.45* | 0.91 |
| ELL=Yes | -4.23* | 1.34 | -4.17* | 1.26 | -2.44 | 1.30 |
| IEP=Yes | 0.67 | 1.40 | -0.97 | 1.29 | -2.63* | 1.31 |
| Age at Pretest | -0.03 | 0.13 | -0.10 | 0.12 | -0.01 | 0.12 |
| Interval from Pretest | 1.61* | 0.66 | 1.33* | 0.55 | -0.08 | 0.84 |
| Condition*Pretest | 0.06 | 0.03 | 0.01 | 0.03 | -0.04 | 0.03 |
| Condition*Gender | -1.70 | 1.28 | -0.73 | 1.19 | -2.46* | 1.19 |
| Condition*ELL | 2.00 | 1.85 | 2.07 | 1.73 | 0.02 | 1.75 |
| Condition*IEP | -1.64 | 1.86 | -1.71 | 1.72 | 1.38 | 1.72 |
| Condition*Age | 0.14 | 0.17 | 0.22 | 0.16 | 0.21 | 0.16 |

*p<.05.

Table S9

*Results of Subgroup Analyses for Applied Problems*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 406.39* | 1.66 | 433.93* | 1.17 | 453.09* | 1.45 |
| Condition=Tools | 0.95 | 1.80 | -0.49 | 1.57 | 0.45 | 1.70 |
| Pretest | 0.49* | 0.03 | 0.36* | 0.03 | 0.32* | 0.03 |
| Gender=Male | 1.69 | 1.72 | -0.35 | 1.50 | 2.10 | 1.59 |
| ELL=Yes | 1.75 | 2.11 | 5.32* | 1.82 | 6.17* | 1.97 |
| IEP=Yes | -4.29 | 2.47 | -5.04* | 2.14 | -6.03* | 2.28 |
| Age at Pretest | 0.07 | 0.23 | 0.07 | 0.20 | 0.09 | 0.21 |
| Interval from Pretest | 1.27 | 1.34 | 0.66 | 0.89 | 2.40 | 1.38 |
| Condition*Pretest | -0.01 | 0.04 | 0.00 | 0.04 | -0.04 | 0.04 |
| Condition*Gender | -2.20 | 2.24 | 0.65 | 1.96 | -1.35 | 2.07 |
| Condition*ELL | 0.38 | 2.79 | 2.13 | 2.41 | -0.20 | 2.58 |
| Condition*IEP | 3.01 | 3.27 | 1.70 | 2.85 | 0.50 | 3.00 |
| Condition*Age | 0.12 | 0.30 | -0.03 | 0.27 | -0.01 | 0.28 |

*p<.05.

Table S10

*Results of Subgroup Analyses for Quantitative Concepts*

|  | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| Parameter | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 421.68* | 1.04 | 447.99* | 1.03 | 461.87* | 1.12 |
| Condition=Tools | -0.72 | 1.39 | -2.85* | 1.29 | -0.61 | 1.30 |
| Pretest | 0.75* | 0.05 | 0.57* | 0.05 | 0.49* | 0.05 |
| Gender=Male | 1.39 | 1.21 | 0.94 | 1.20 | 2.09 | 1.28 |
| ELL=Yes | 3.68* | 1.38 | 4.09* | 1.38 | 4.50* | 1.48 |
| IEP=Yes | -3.02 | 1.71 | -4.36* | 1.68 | -4.41* | 1.79 |
| Age at Pretest | 0.09 | 0.16 | -0.14 | 0.16 | -0.04 | 0.17 |
| Interval from Pretest | 2.39* | 0.95 | -1.53 | 0.81 | 1.13 | 1.10 |
| Condition*Pretest | 0.03 | 0.07 | -0.02 | 0.07 | 0.04 | 0.07 |
| Condition*Gender | -2.01 | 1.58 | -0.97 | 1.57 | -1.63 | 1.67 |
| Condition*ELL | -0.37 | 1.88 | -0.61 | 1.85 | -1.39 | 1.94 |
| Condition*IEP | 1.10 | 2.28 | -0.26 | 2.24 | 0.08 | 2.36 |
| Condition*Age | 0.26 | 0.22 | 0.02 | 0.22 | 0.18 | 0.23 |

*p<.05.

Table S11

*Results of Subgroup Analyses for Passage Comprehension*

| Parameter | End of First Grade | |
|---|---|---|
| | *b* | *se* |
| Intercept | 457.88* | 1.78 |
| Condition=Tools | -1.13 | 2.22 |
| Pretest (composite) | 2.25* | 0.22 |
| Gender=Male | 0.44 | 1.97 |
| ELL=Yes | 11.38* | 2.56 |
| IEP=Yes | -5.62* | 2.79 |
| Age at Pretest | -0.30 | 0.26 |
| Interval from Pretest | -0.29 | 1.73 |
| Condition*Pretest | -0.15 | 0.29 |
| Condition*Gender | -2.85 | 2.56 |
| Condition*ELL | 1.10 | 3.42 |
| Condition*IEP | -2.15 | 3.67 |
| Condition*Age | 0.31 | 0.36 |

*p<.05.

Table S12

*Results of Subgroup Analyses for Forward Span*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 3.19* | 0.10 | 3.94* | 0.11 | 4.59* | 0.11 |
| Condition=Tools | -0.19 | 0.12 | -0.10 | 0.12 | 0.05 | 0.12 |
| Pretest | 0.37* | 0.05 | 0.22* | 0.05 | 0.27* | 0.05 |
| Gender=Male | 0.04 | 0.11 | 0.18 | 0.11 | 0.04 | 0.12 |
| ELL=Yes | 0.32* | 0.12 | 0.23 | 0.12 | 0.27* | 0.13 |
| IEP=Yes | -0.01 | 0.16 | -0.26 | 0.16 | -0.33 | 0.17 |
| Age at Pretest | 0.02 | 0.02 | 0.00 | 0.02 | 0.01 | 0.02 |
| Interval from Pretest | 0.06 | 0.09 | -0.10 | 0.08 | 0.10 | 0.10 |
| Condition*Pretest | -0.02 | 0.06 | 0.09 | 0.06 | -0.04 | 0.06 |
| Condition*Gender | -0.02 | 0.15 | 0.02 | 0.15 | -0.22 | 0.15 |
| Condition*ELL | -0.41* | 0.16 | -0.27 | 0.16 | 0.09 | 0.17 |
| Condition*IEP | -0.05 | 0.22 | 0.13 | 0.21 | 0.26 | 0.22 |
| Condition*Age | 0.00 | 0.02 | 0.03 | 0.02 | 0.01 | 0.02 |

*p<.05.

Table S13

*Results of Subgroup Analyses for Corsi Backward Span*

|  | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| Parameter | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 1.50* | 0.12 | 2.71* | 0.14 | 3.70* | 0.13 |
| Condition=Tools | -0.06 | 0.14 | -0.16 | 0.15 | -0.04 | 0.15 |
| Pretest | 0.20* | 0.06 | 0.11 | 0.06 | 0.23* | 0.06 |
| Gender=Male | -0.35* | 0.14 | 0.00 | 0.14 | -0.09 | 0.14 |
| ELL=Yes | -0.32* | 0.15 | 0.01 | 0.16 | 0.38* | 0.16 |
| IEP=Yes | -0.24 | 0.20 | -0.47* | 0.20 | -0.46 | 0.20 |
| Age at Pretest | 0.04* | 0.02 | 0.08* | 0.02 | 0.04 | 0.02 |
| Interval from Pretest | 0.35* | 0.11 | 0.20 | 0.10 | 0.11 | 0.12 |
| Condition*Pretest | 0.01 | 0.08 | 0.15 | 0.08 | 0.01 | 0.08 |
| Condition*Gender | 0.39* | 0.19 | 0.03 | 0.19 | 0.17 | 0.18 |
| Condition*ELL | 0.35 | 0.20 | -0.08 | 0.21 | -0.18 | 0.20 |
| Condition*IEP | -0.12 | 0.27 | 0.23 | 0.27 | 0.25 | 0.26 |
| Condition*Age | 0.02 | 0.02 | -0.03 | 0.03 | 0.02 | 0.02 |

*p<.05.

Table S14

*Results of Subgroup Analyses for the Dimensional Change Card Sort*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 1.59* | 0.06 | 1.86* | 0.06 | 2.44* | 0.10 |
| Condition=Tools | 0.03 | 0.06 | 0.04 | 0.07 | 0.09 | 0.11 |
| Pretest | 0.33* | 0.05 | 0.16* | 0.05 | 0.24* | 0.09 |
| Gender=Male | -0.04 | 0.06 | 0.00 | 0.06 | 0.08 | 0.10 |
| ELL=Yes | -0.11 | 0.07 | -0.14* | 0.07 | 0.02 | 0.12 |
| IEP=Yes | -0.15 | 0.08 | -0.28* | 0.09 | -0.40* | 0.14 |
| Age at Pretest | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 |
| Interval from Pretest | 0.07 | 0.05 | 0.02 | 0.05 | 0.09 | 0.09 |
| Condition*Pretest | -0.03 | 0.07 | 0.16* | 0.07 | 0.10 | 0.11 |
| Condition*Gender | -0.16* | 0.08 | -0.11 | 0.08 | -0.07 | 0.13 |
| Condition*ELL | -0.02 | 0.09 | 0.11 | 0.10 | 0.00 | 0.15 |
| Condition*IEP | 0.14 | 0.11 | 0.17 | 0.12 | 0.38* | 0.19 |
| Condition*Age | 0.00 | 0.01 | -0.01 | 0.01 | -0.01 | 0.02 |

*p<.05.

Table S15

*Results of Subgroup Analyses for Copy Design*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 4.90* | 0.26 | 7.71* | 0.25 | 9.47* | 0.24 |
| Condition=Tools | 0.46 | 0.30 | -0.36 | 0.30 | -0.74* | 0.31 |
| Pretest | 0.65* | 0.08 | 0.64* | 0.10 | 0.57* | 0.10 |
| Gender=Male | -0.35 | 0.26 | -0.69* | 0.29 | -0.43 | 0.31 |
| ELL=Yes | 1.00* | 0.29 | 0.70* | 0.32 | 0.95* | 0.34 |
| IEP=Yes | -0.35 | 0.36 | -0.59 | 0.41 | -0.89* | 0.43 |
| Age at Pretest | 0.13* | 0.03 | 0.03 | 0.04 | 0.04 | 0.04 |
| Interval from Pretest | -0.01 | 0.23 | 0.25 | 0.19 | -0.05 | 0.25 |
| Condition*Pretest | 0.20 | 0.11 | 0.03 | 0.13 | 0.12 | 0.14 |
| Condition*Gender | 0.19 | 0.34 | 0.60 | 0.38 | 0.31 | 0.41 |
| Condition*ELL | 0.35 | 0.38 | 0.08 | 0.42 | -0.52 | 0.44 |
| Condition*IEP | -0.13 | 0.48 | -1.14* | 0.54 | -0.63 | 0.57 |
| Condition*Age | -0.05 | 0.05 | 0.05 | 0.05 | 0.01 | 0.06 |

*p<.05.

Table S16

*Results of Subgroup Analyses for Head-Toes-Knees-Shoulders*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 18.58 | 1.33 | 33.38 | 1.14 | 42.86 | 0.79 |
| Condition=Tools | 2.09 | 1.66 | 0.56 | 1.50 | -0.50 | 1.04 |
| Pretest | 0.64 | 0.07 | 0.30 | 0.06 | 0.10 | 0.04 |
| Gender=Male | -0.41 | 1.61 | -2.71 | 1.46 | -1.01 | 0.98 |
| ELL=Yes | -3.91 | 1.75 | -2.45 | 1.61 | 0.20 | 1.10 |
| IEP=Yes | -7.62 | 2.26 | -8.09 | 2.02 | -5.54 | 1.36 |
| Age at Pretest | 0.52 | 0.21 | 0.19 | 0.19 | 0.18 | 0.13 |
| Interval from Pretest | 3.91 | 1.19 | -0.26 | 0.90 | -0.88 | 0.81 |
| Condition*Pretest | -0.05 | 0.08 | 0.04 | 0.08 | 0.07 | 0.05 |
| Condition*Gender | -3.40 | 2.09 | -0.01 | 1.90 | -1.51 | 1.27 |
| Condition*ELL | 1.88 | 2.33 | 0.02 | 2.11 | 0.12 | 1.45 |
| Condition*IEP | 5.03 | 2.99 | 2.66 | 2.68 | 0.84 | 1.80 |
| Condition*Age | -0.01 | 0.28 | -0.12 | 0.25 | 0.09 | 0.17 |

*p<.05.

Table S17

*Results of Subgroup Analyses for Peg Tapping*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 9.14* | 0.44 | 12.66* | 0.32 | 14.82* | 0.23 |
| Condition=Tools | -0.59 | 0.58 | -0.17 | 0.43 | -0.59* | 0.28 |
| Pretest | 0.57* | 0.05 | 0.25* | 0.04 | 0.08* | 0.02 |
| Gender=Male | -0.77 | 0.52 | -0.26 | 0.42 | -0.25 | 0.27 |
| ELL=Yes | 0.10 | 0.58 | 0.52* | 0.47 | 0.36 | 0.31 |
| IEP=Yes | -0.58 | 0.74 | -1.68 | 0.59 | -0.68 | 0.38 |
| Age at Pretest | 0.06 | 0.07 | -0.01 | 0.06 | -0.01 | 0.04 |
| Interval from Pretest | 0.30 | 0.40 | -0.47 | 0.25 | -0.12 | 0.23 |
| Condition*Pretest | -0.06 | 0.06 | -0.05 | 0.05 | 0.02 | 0.03 |
| Condition*Gender | 0.10 | 0.68 | -0.06 | 0.55 | -0.13 | 0.35 |
| Condition*ELL | -1.28 | 0.78 | -1.43* | 0.62 | -0.41 | 0.40 |
| Condition*IEP | -1.21 | 0.98 | -0.19 | 0.79 | -0.56 | 0.50 |
| Condition*Age | 0.06 | 0.09 | 0.00 | 0.07 | 0.03 | 0.05 |

*$p<.05$.

Table S18

*Results of Subgroup Analyses for Interpersonal Skills*

|  | End of Pre-K | | End of Kindergarten | | End of First Grade | |
| --- | --- | --- | --- | --- | --- | --- |
| Parameter | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 5.46* | 0.08 | 5.74* | 0.10 | 5.75* | 0.10 |
| Condition=Tools | 0.00 | 0.10 | 0.00 | 0.12 | 0.07 | 0.13 |
| Pretest | 0.85* | 0.03 | 0.58* | 0.04 | 0.49* | 0.05 |
| Gender=Male | 0.01 | 0.07 | -0.14 | 0.09 | -0.29* | 0.10 |
| ELL=Yes | 0.22* | 0.08 | 0.43* | 0.11 | 0.29* | 0.12 |
| IEP=Yes | 0.05 | 0.10 | 0.09 | 0.13 | 0.07 | 0.14 |
| Age at Pretest | -0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 |
| Interval from Pretest | 0.05 | 0.08 | -0.10 | 0.08 | 0.15 | 0.11 |
| Condition*Pretest | -0.13* | 0.04 | -0.09 | 0.06 | -0.04 | 0.07 |
| Condition*Gender | -0.04 | 0.09 | 0.05 | 0.12 | 0.14 | 0.14 |
| Condition*ELL | -0.09 | 0.10 | -0.11 | 0.14 | 0.15 | 0.16 |
| Condition*IEP | -0.27* | 0.13 | -0.04 | 0.18 | 0.08 | 0.19 |
| Condition*Age | 0.00 | 0.01 | -0.02 | 0.02 | -0.03 | 0.02 |

*p<.05.

Table S19

*Results of Subgroup Analyses for Work-related Skills*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | $b$ | $se$ | $b$ | $se$ | $b$ | $se$ |
| Intercept | 5.02* | 0.09 | 5.03* | 0.09 | 4.99* | 0.11 |
| Condition=Tools | -0.04 | 0.13 | -0.04 | 0.13 | 0.11 | 0.13 |
| Pretest | 0.79* | 0.04 | 0.51* | 0.05 | 0.51* | 0.06 |
| Gender=Male | -0.13 | 0.08 | -0.26* | 0.11 | -0.37* | 0.12 |
| ELL=Yes | 0.28* | 0.09 | 0.48* | 0.12 | 0.41* | 0.14 |
| IEP=Yes | 0.02 | 0.12 | 0.02 | 0.16 | -0.18 | 0.17 |
| Age at Pretest | 0.02* | 0.01 | 0.03* | 0.01 | 0.02 | 0.02 |
| Interval from Pretest | -0.03 | 0.10 | 0.05 | 0.08 | 0.14 | 0.12 |
| Condition*Pretest | -0.07 | 0.05 | -0.02 | 0.07 | 0.01 | 0.07 |
| Condition*Gender | 0.06 | 0.11 | 0.04 | 0.15 | 0.09 | 0.16 |
| Condition*ELL | -0.07 | 0.12 | -0.16 | 0.16 | 0.19 | 0.18 |
| Condition*IEP | -0.32 | 0.16 | -0.31 | 0.21 | 0.30 | 0.23 |
| Condition*Age | -0.02 | 0.01 | -0.02 | 0.02 | -0.03 | 0.02 |

*p<.05.

Table S20

*Results of Subgroup Analyses for the Adaptive Language Inventory*

|  | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| Parameter | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 3.17* | 0.07 | 3.12* | 0.07 | 3.14* | 0.08 |
| Condition=Tools | 0.09 | 0.09 | -0.01 | 0.09 | 0.09 | 0.10 |
| Pretest | 0.82* | 0.05 | 0.46* | 0.05 | 0.52* | 0.06 |
| Gender=Male | -0.09 | 0.06 | -0.06 | 0.07 | 0.00 | 0.08 |
| ELL=Yes | 0.09 | 0.07 | 0.22* | 0.09 | 0.19* | 0.10 |
| IEP=Yes | -0.21* | 0.09 | -0.23* | 0.11 | -0.18 | 0.12 |
| Age at Pretest | 0.02* | 0.01 | 0.02 | 0.01 | 0.00 | 0.01 |
| Interval from Pretest | -0.09 | 0.07 | 0.11 | 0.06 | 0.13 | 0.09 |
| Condition*Pretest | -0.06 | 0.06 | -0.03 | 0.07 | -0.07 | 0.08 |
| Condition*Gender | 0.07 | 0.08 | -0.17 | 0.10 | -0.19 | 0.11 |
| Condition*ELL | -0.05 | 0.10 | -0.07 | 0.12 | 0.12 | 0.13 |
| Condition*IEP | 0.01 | 0.12 | -0.11 | 0.15 | 0.09 | 0.16 |
| Condition*Age | -0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 |

*p<.05.

Table S21

*Results of Subgroup Analyses for the Assessor Ratings of Attention*

| Parameter | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Intercept | 2.39* | 0.05 | 2.62* | 0.09 | 2.65* | 0.04 |
| Condition=Tools | -0.02 | 0.07 | -0.02 | 0.05 | -0.09 | 0.05 |
| Pretest | 0.49* | 0.04 | 0.37* | 0.04 | 0.29* | 0.04 |
| Gender=Male | -0.01 | 0.05 | -0.03 | 0.05 | -0.06 | 0.05 |
| ELL=Yes | 0.10 | 0.06 | -0.08 | 0.05 | 0.03 | 0.05 |
| IEP=Yes | -0.22* | 0.07 | -0.13 | 0.07 | -0.19* | 0.06 |
| Age at Pretest | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 |
| Interval from Pretest | 0.10 | 0.05 | 0.05 | 0.04 | -0.07 | 0.05 |
| Condition*Pretest | -0.01 | 0.06 | -0.03 | 0.05 | 0.01 | 0.05 |
| Condition*Gender | -0.02 | 0.07 | -0.03 | 0.06 | 0.04 | 0.06 |
| Condition*ELL | 0.01 | 0.08 | 0.14 | 0.07 | -0.01 | 0.07 |
| Condition*IEP | -0.05 | 0.10 | -0.06 | 0.09 | -0.13 | 0.08 |
| Condition*Age | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 |

*p<.05.

Table S22

*Impacts of Fidelity of Implementation on Child Outcomes at the End of Pre-K, Kindergarten, and 1ˢᵗ Grade*

| Variable | End of Pre-K | | End of Kindergarten | | End of First Grade | |
|---|---|---|---|---|---|---|
| | *b* | *se* | *b* | *se* | *b* | *se* |
| Composite Achievement | -0.008 | 0.006 | -0.018* | 0.006 | -0.010 | 0.006 |
| Composite Self-Regulation | -0.002 | 0.005 | -0.006 | 0.005 | -0.016* | 0.005 |
| CFBRS Interpersonal Skills Scale | 0.003 | 0.002 | 0.001 | 0.002 | -0.001 | 0.002 |
| CFBRS Work-Related Skills Scale | 0.002 | 0.002 | 0.003 | 0.002 | -0.001 | 0.002 |
| Adaptive Language Inventory | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| Self-Regulation Assessor Ratings | -0.001 | 0.001 | -0.001 | 0.001 | -0.001 | 0.001 |

*Note.* Analysis conducted only *Tools of the Mind* classrooms. Coefficients in the table are unstandardized regression coefficients from multi-level regression models. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest.
* $p<.05$