Experimental Evaluation of the *Tools of the Mind* Pre-K Curriculum

Technical Report

Dale C. Farran, PhD
Sandra J. Wilson, PhD
Deanna Meador, MA
Jennifer Norvell, MEd
Kimberly Nesbitt, PhD

9/1/15

PEABODY research institute

VANDERBILT PEABODY COLLEGE

Director
Mark W. Lipsey, Ph. D.

Senior Associate Director
Dale C. Farran, Ph. D.

Associate Director
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.

Recommended Citation:

Farran, D.C., Wilson, S.J., Meador, D., Norvell, J., & Nesbitt, K. (2015). *Experimental Evaluation of the Tools of the Mind Pre-K Curriculum: Technical Report.* (Working Paper). Peabody Research Institute, Vanderbilt University.

Funding Source:

The research reported here was supported by Grant #R305A090533 from the Institute of Education Sciences, U.S. Department of Education, for the study titled "Experimental Validation of the Tools of the Mind Prekindergarten Curriculum." Grant awarded to Drs. Dale C. Farran, Mark Lipsey, and Sandra Jo Wilson at Vanderbilt University with sub awards to Drs. Deborah Leong and Elena Bodrova at Third Sector New England.

Contact Us:
Phone: 615.322.8540
Fax: 615.322.0293

Mailing Address:

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

Delivery Address:

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

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

PRI Project Staff

Principal Investigators

Dale C. Farran, PhD

Mark W. Lipsey, PhD

Sandra Jo Wilson, PhD

Project Coordinators

Deanna Meador, MA
Jennifer Norvell, MEd
Elizabeth Vorhaus, MEd
Diane Spencer, MEd (North Carolina)
Carolyn Boyles, EdD (North Carolina)

Research Associates

Kimberly Turner Nesbitt, PhD
Karen Anthony, PhD

Research Assistants

Ashley Keene, MS
Jessica Ziegler, MEd
Marianne Reale, MEd

Post-Doctoral Fellows

Mary Wagner Fuhs, PhD
Asha Spivak, PhD

Doctoral Fellows

Sascha Mowrey
Alana Rimmel
Cathy Yun

Project staff also included multiple part-time child assessors and classroom observers in Tennessee and North Carolina.

Curriculum training and coaching supervised by Drs. Deborah Leong and Elena Bodrova.

Table of Contents

Introduction

Description of the Project

The experimental evaluation of the *Tools of the Mind* Pre-K Curriculum described in this report was designed to examine the effectiveness of the *Tools of the Mind* (*Tools*) curriculum for enhancing children's self-regulation skills and their academic preparation for kindergarten when compared to the usual prekindergarten curricula in use in the school system. In order to assess the long-term impacts of *Tools* on student academic achievement and self-regulation outcomes, students were followed into kindergarten and first grade. Participating classrooms were also observed three times during the prekindergarten year using multiple measures designed to capture implementation fidelity as well as child and teacher behaviors in the classroom.

The 5-year research study was funded by a grant from the US Department of Education Institute of Education Sciences (**R305A090533)** to Drs. Dale Farran, Mark Lipsey and Sandra Wilson of the Peabody Research Institute at Vanderbilt University. Training and coaching support for Tools of the Mind was funded through a sub-award to each of the developers, Dr. Deborah Leong at Metropolitan State College of Denver and Dr. Elena Bodrova of the McREL Institute of Denver Colorado.

Early Childhood Curricula and Tools of the Mind

Successful transition into formal schooling for young children and their subsequent academic success require a variety of competencies, including, most obviously, the early literacy and numeracy skills that provide the foundation for reading and mathematics. There are also important social competencies that enable children to interact with peers and adults appropriately in school settings. A third critical area of competency is the ability to engage in and benefit from the kinds of learning tasks and activities intrinsic to school-based instruction, including attending to speech and visual stimuli that convey information, completing exercises that require planning, solving problems, applying knowledge and practicing acquired skills, and remembering and following rules and instructions. These latter skills enable children to focus on, and benefit from, the educational material and learning opportunities provided in school settings. These skills go by different names: self-regulation, executive function, attentional skills, etc., and recent longitudinal studies have demonstrated that they have an independent influence on long term academic success, separate from early academic skills.

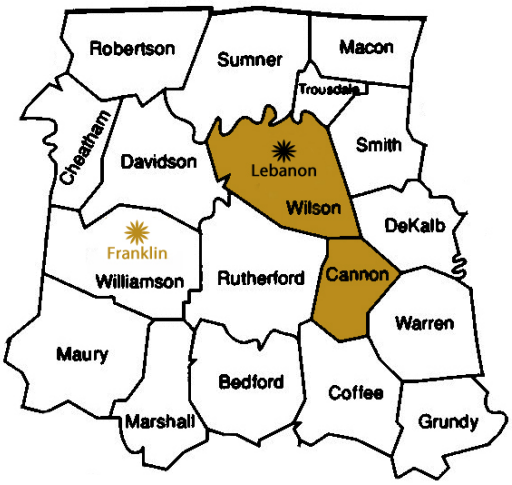
Attempting to influence the development of this third area of competency is associated with a Vygotskian approach, which argues that this emphasis will provide children with sustainable skills they can then apply to the changing array of academic and social tasks they experience as they navigate through formal schooling. This is the theoretical underpinning of the approach taken by *Tools of the Mind* (Bodrova & Leong, 2007). This perspective focuses on equipping children with cognitive "tools" for learning that they can apply to the task of acquiring and sustaining academic knowledge and skills as well as social competencies. At the time of this evaluation, the approach had

not been examined with a rigorous, experimental design that compared its effectiveness to the other alternatives for early childhood programs available to school systems. Four school systems in Tennessee and two in North Carolina agreed to partner with us to conduct this first experimental evaluation of the curriculum.

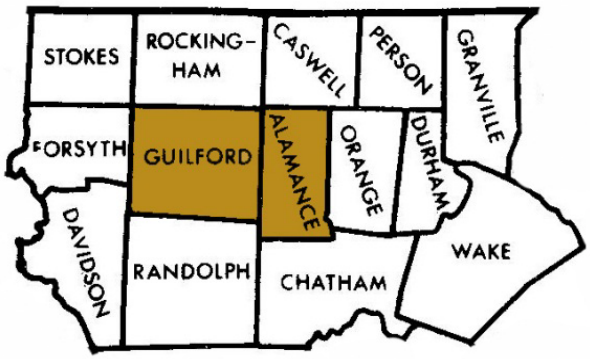
Participants

The project was fortunate to have participants from Franklin Special School District, Lebanon Special School District, Wilson County School District, and Cannon County School District in Tennessee as well as Guilford County Schools and Alamance-Burlington School System in North Carolina.

Tennessee



North Carolina



The evaluation involved two cohorts of children. Cohort 1 (2010-2011) included the four Tennessee school systems and Guilford County Schools in North Carolina and involved children from 60 classrooms (*Tools*=32) in 45 schools (*Tools*=25). Cohort 2 (2011-2012) included Alamance-Burlington School System in North Carolina with children from 20 classrooms (*Tools*=10) in 12 schools (*Tools*=5).

Research Design

To investigate the effectiveness of *Tools*, 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?

In addition, an extensive battery of observational measures was employed to examine implementation fidelity and other classroom processes that might have mediated the curriculum effects.

A large-scale cluster-randomized block design was employed with each cohort 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.

Each cohort's trial was conducted over two years, with randomization occurring during the first year. For cohort 1, the four smaller districts in Tennessee served as individual blocks. The 22 schools in Guilford County, North Carolina 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 2009-2010 school year, the implementation fidelity scheme was developed, but no data were collected on children.

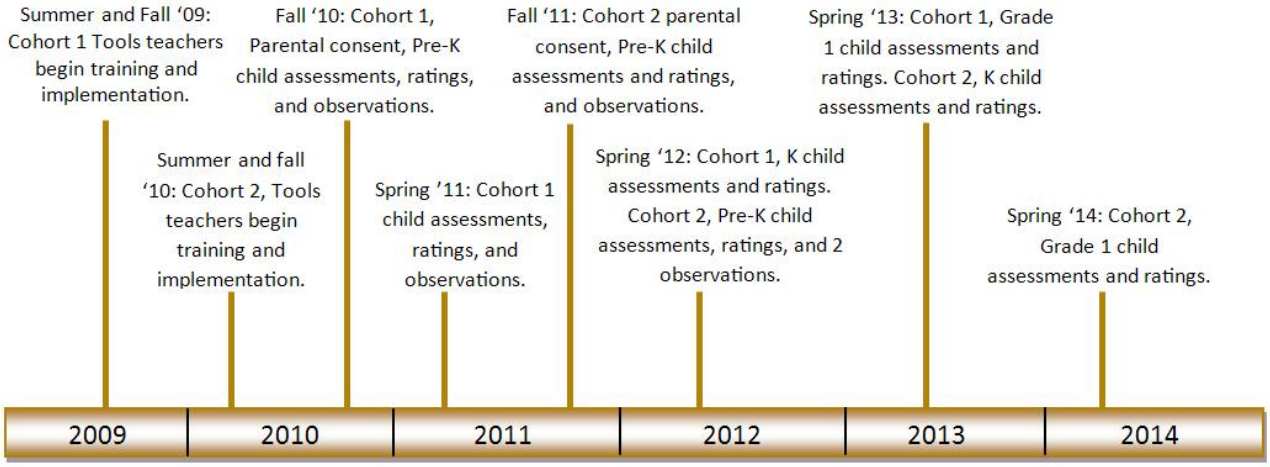
For cohort 2, randomization occurred in the summer prior to the 2010-2011 school year. The twelve schools were divided into three blocks and schools within blocks were randomly assigned to *Tools* (n=5 schools) or the comparison condition (n=7 schools). All of the prekindergarten classrooms in the schools participated in the condition to which they were assigned. As with cohort 1, training of teachers and practice in the curriculum occurred during the 2010-2011 school year, but no data were collected on children.

Children were recruited for the research study beginning in the fall of 2010 for cohort 1 and the fall of 2011 for cohort 2, during which teachers received additional training and implemented the program. The curriculum tests thus occurred in the 2010-2011 school year for cohort 1 and the 2011-2012 school year for cohort 2. Following the treatment year, for the next two years, children were located in their kindergarten and first grade classroom. Individual assessments on the same

battery used in pre-k supplemented by a reading readiness measures were administered to all children. Kindergarten and first grade teachers were surveyed in the spring and rated various aspects of children's classroom interactions. 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.

Project Timeline

Below is the timeline for project activities from 2009 to 2014. This timeline shows the assessments, behavioral ratings, and classroom observations that students and teachers participated in for both cohorts.



Organization of this Report

What follows is a full technical summary of the data collection activities and main analyses.

Description of Measures

The goal of the *Experimental Evaluation of the Tools of the Mind Pre-K Curriculum* is to determine if the *Tools* curriculum is more effective in enhancing children's self-regulation and academic preparedness for kindergarten when compared to other "business as usual" preschool curricula. To achieve this goal, we examined child achievement measures as well as a number of direct assessment and behavior rating measures using a variety of self-regulation tasks. The academic achievement measures are described below. After that, we provide descriptions of the self-regulation direct assessment measures and the teacher ratings also being used in the study.

Woodcock-Johnson III Tests of Achievement (WJ-III)

The areas of assessment were chosen by the developers as ones they believed the curriculum would affect. Certified individuals assessed children individually in fall and spring.

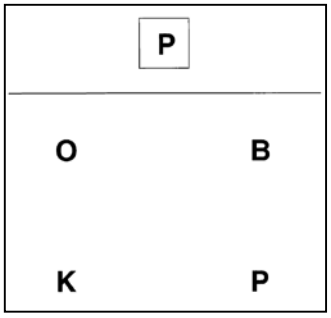
WJ-III standard scores are reported, which are normed to a representative sample of American youth. Standard scores have a mean of 100 and a standard deviation of 15. A score of 100, therefore, is considered average. Higher scores on the measures reflect better academic performance. An increase in *standard scores* from fall to spring indicates learning at a faster rate than the children had previously demonstrated. Analyses were conducted with *W* scores.

The same measures were used in follow up assessments.

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

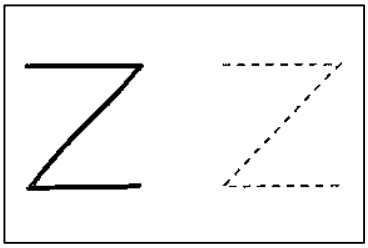
WJ-III Literacy Measures

Letter Word Identification



- Letter Word Identification assesses children's letter and word identification ability. Items include identifying and pronouncing presented letters and pronouncing presented words.
- Sample Script: *This is the letter "P." Find the "P" down here.*

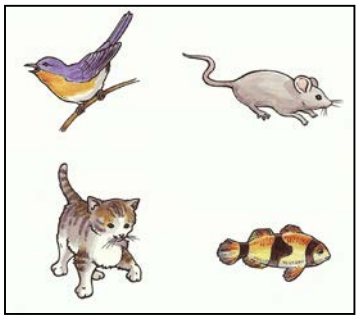
Spelling



- Spelling measures the ability to write orally presented letters and words correctly beginning with tracing simple shapes.
- Sample Script: *Watch Me.* [Trace "Z" on left. Hand pencil to child, point to "Z" on right] *Now you make one just like I did. Stay on the line.*

WJ-III Language Measures

Academic Knowledge

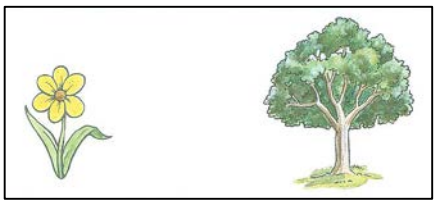


- Academic Knowledge is given in three subtests measuring factual knowledge of science, social studies, and humanities.
- Sample Script: *Look at the pictures. Put your finger on the one that flies.*

Oral Comprehension

- Oral Comprehension assesses children's ability to understand a short passage by providing a missing word based on the syntactic and semantic cues of the sentence.
- Sample Script: *Water looks blue and grass looks* _____ (pause expectantly).

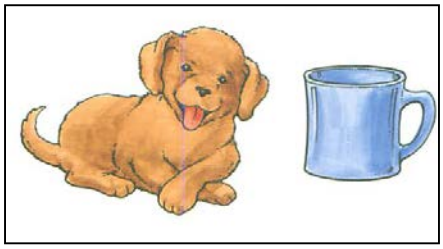
Picture Vocabulary



- Picture Vocabulary assesses children's expressive language and word knowledge at the single word level. After the initial items, children must say the name of the picture.
- Sample Script of initial item: *Put your finger on the flower.*

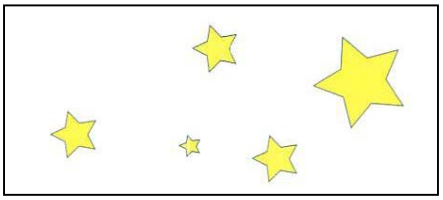
WJ-III Mathematics Measures

Applied Problems

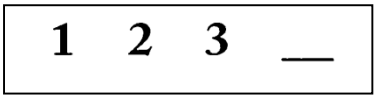


- Applied Problems assesses children's ability to solve mathematics problems. The items in the scale measure children's ability to identify information necessary to solve problems and to determine an appropriate strategy to solve the problem.
- Sample Script: *How many dogs are there in this picture?*

Quantitative Concepts



- Quantitative Concepts is a measure given in two parts. The first part assesses children's knowledge of mathematical concepts, including vocabulary, numbers, shapes, and symbols. The second part measures sequencing of numbers, with increasing.
- Sample Script A: *Point to the largest star. Now point to the smallest star.*
- Sample Script B: *Look at these numbers and tell me the number that belongs in the blank space.*



Self-Regulation

Children were assessed individually in two sessions in the fall and spring. The following assessments were used. Additional detail on the self-regulation measures, including scripts and score sheets, can be found here: https://my.vanderbilt.edu/toolsofthemindevaluation/resources/child-assessment-measures/.

Peg Tapping

- Children are instructed to tap once with a wooden dowel when the examiner taps twice and to tap twice when the examiner taps once.
- The Peg Tapping task is a measure of inhibitory control; a child must inhibit the most powerful immediate response of imitating the examiner.
- Each item is scored 0 if the child gives the incorrect number of taps and 1 if the child gives the correct number of taps. Scores on the items are summed and converted to a proportion correct out of a possible score of 16. Larger scores on the task reflect greater inhibitory control.
- For more information see: Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the ability to remember what I said and to "do as I say, not as I do." *Developmental Psychobiology*, *29*, 315-334.

Head Toes Knees Shoulders (HTKS)

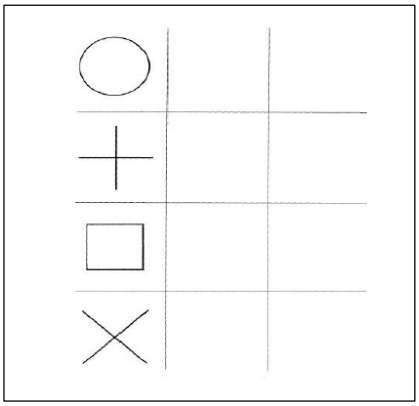
- Children are asked to play a game in which they must do the opposite of what the examiner says. The examiner instructs children to touch their head (or their toes), but instead of following the command, the children are supposed to do the opposite and touch their toes (or their head). If children pass the head/toes part of the task, they complete an advanced trial where the knees and shoulders commands are added.
- The HTKS task is a measure of inhibitory control; a child must inhibit the dominant response of imitating the examiner.
- Each response is scored with the following system: 0 = incorrect response, 1 = any motion to an incorrect response, but self-corrected to the correct response, and 2 = correct response. Scores on the first six practice items and the 20 test items are summed and converted to a proportion correct out of a possible score of 52. Larger scores on the task reflect greater inhibitory control.

- For more information see: 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.

Dimensional Change Card Sort (DCCS)

- Children are required to sort picture cards first according to one dimension (e.g., color) and then according to another dimension (e.g., shape). If they can make this switch, children are then asked to complete an advanced version of the DCCS that adds a third sorting rule, sorting by the borders on the cards (e.g., the presence of a border means one rule, no border means another rule).
- The DCCS is a measure of attention shifting. To complete the task, children must shift their attention to a different dimension of the card – from the color of the object to the shape of the object (e. g. focus on the shape and not the color of the shape). To complete the advanced phase, children must shift their focus from one dimension to another from card to card.
- The task is scored as follows, using a system developed by Zelazo. Scores were converted to a proportion correct out of 3. Larger scores on the task reflect greater ability to shift attention with task demands and less perseveration.
  - 0 = Sorted by color on fewer than 5/6 cards
  - 1 = Sorted by color on at least 5/6 cards, but sorted by shape on fewer than 5/6 cards
  - 2 = Sorted by color and shape on at least 5/6 cards, but sorted fewer than 9/12 cards correctly on advanced version
  - 3 = Sorted by color and shape on at least 5/6 card and sorted at least 9/12 cards correctly on advanced version.
- For more information see: Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297-301.

Copy Design



- Children are asked to copy 8 simple geometric designs. Children are given two attempts to draw each of the 8 designs. The attempts are scored to indicate if the child was able to properly replicate the design.
- The Copy Design task is a measure of persistence and sustained attention during a difficult task.
- Each design is given a score of 1 if at least one attempt is correct, 2 if both attempts are correct, and 0 if both attempts are incorrect or are not attempted. Scores on the items are

summed and converted to a proportion correct out of a possible score of 16. Larger scores on the task indicate greater attention and sustained focus.

- For more information see: 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.

Corsi Blocks

- Children are asked to point to a series of blocks as indicated by the examiner. Children are first asked to repeat the pattern exactly as the examiner did (i.e., forwards) then they are asked to reverse the pattern given by the examiner (i.e., backwards). Task difficulty increases by asking children to repeat increasingly longer block patterns. The child gets two attempts at each pattern and continues until the recalled pattern is no longer correct.
- Corsi Blocks is a measure of working memory.
- The task is scored as the largest pattern span that the child is able to reproduce. The maximum score possible is 9 for forward span and 7 for backward span. Larger scores indicate a greater working memory.
- For more information see: Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, *38*, 317-338.

Behavior Rating Scales (collected from teachers)

Teachers rated the children in their classes 6 weeks after school began and again at the end of the year.

Cooper-Farran Behavior Rating Scales

The Cooper-Farran Behavior Rating Scale is composed of 37 items in two subscales. The Interpersonal Skills subscale (IPS) includes 21 items and the Work-Related Skills (WRS) subscale includes 16 items. The IPS subscale measures how well children get along with peers and the teacher. The WRS subscale includes items about independent work, compliance with instructions, and memory for instructions. Items are rated on a 1-7 scale with descriptive phrases to "anchor" points 1, 3, 5, and 7.

Example item for Interpersonal Skills:

EFFECT ON OTHER CHILDREN

1	2	3	4	5	6	7
Does not purposefully annoy anyone		Teases others but stops short of actual annoyance		Occasionally tries to get attention by playful but annoying behavior		Repeatedly irritates others by hostile touching, poking, verbal insulting, etc.

Example item for Work-Related Skills:

RELEVANT PARTICIPATION IN GROUP DISCUSSIONS

1	2	3	4	5	6	7
Often contributes original ideas; relevant and responsive to others' comments and interests		Makes an occasional relevant comment; attentive		Inattentive to others; quiet but uninvolved		Makes irrelevant remarks; interrupts the flow

- For more information see: Cooper, D., & Farran, D. C. (1988). Behavioral risk in kindergarten. *Early Childhood Research Quarterly*, *3*, 1-20.

Adaptive Language Inventory (ALI)

- The ALI focuses on children's comprehension and use of language in classroom settings in comparison to their peers and has been used both at the preschool and elementary levels. The measure consists of 18 items that focus on comprehension, production, rephrasing, spontaneity, listening, and fluency. Children are rated on 1-5 scale.
- Sample items: Responds to questions asked of him/her in a thoughtful logical way. Listens carefully when the teacher is giving instructions to the class.
- For more information see: 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.

Self-Regulation Assessor Rating (SAR)

Assessors rated children's self-regulatory behaviors during the assessment sessions. 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. Data from the attentiveness scale is used in this report.

- Sample Item:

  A1. Pays attention during instructions and demonstrations

  3. Child looks closely at pictures to distinguish between them; child attends to and complies with interviewer
  2. Child's attention occasionally drifts, particularly at the end of activities, but is responsive to prompt
  1. Child's attention frequently drifts and examiner provides frequent prompts
  0. Child spends most of time off-task, inattentive

- For more information see: 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

Classroom Observation Measures

In addition to child assessments and teacher ratings, data were also collected through a series of classroom observations and observer ratings in both *Tools* and comparison classrooms. Classrooms were observed three times during the fall and spring semesters, with one observation in the fall and two in the spring semester. Two observers visited the classroom to note all instructional classroom activities during the day. One observer completed the *Narrative Record* and *Tools of the Mind* fidelity measure as well as an environmental scan of the classroom materials. The second observer completed the *Teacher Observation in Preschools (TOP)* and the *Child Observation in Preschool (COP).* At the conclusion of the observation, each observer collaborated to complete the *Post Observation Rating Scale (PRS)*. The classroom observation measures are described below. Full manuals and additional information can be found here: https://my.vanderbilt.edu/toolsofthemindevaluation/resources/.

Narrative Record

The *Narrative Record* is an open-ended format for recording descriptive narrative data about classroom activities, as well the amount of time spent in various activities. In addition, ratings of teacher instructional level and student engagement are recorded. This system was used in both *Tools* and comparison classrooms to determine similarities and differences among them. The focus of the *Narrative Record* is the ***whole class***; whatever the class as a whole (defined as at least 75% of the children) is doing is coded. The *Narrative Record* consists of the following items:

- Episodes of Time: Each instructional episode is coded for beginning and ending times. An episode is defined as beginning when there is a change in the method of instruction or a change in the focus of instruction. Detailed notes are kept about each episode.
- Codes for Type of Activity (Learning Setting) during the episode
    - Single setting: Whole Group with or without Teacher (WG), Small Group (SG), Meal, Transition
    - Multiple settings: Small Groups operating with Centers (SGC), Small Group with Teacher during Centers (SGTC), Out of Room (e.g., outdoors, bathroom, or specials, such as library)
- Codes for Content of Instruction (Learning Focus) occurring during the episode (e.g., math, reading, code-based language instruction, science, social studies, art, music & movement, and none)
- Codes for the highest Level of Instruction provided by the teacher during an episode. These range from no instruction (0) to highly inferential instruction (4). Inferential means asking open-ended questions; highly inferential involves several turns with inferential questions and follow up.
- Codes for Engagement Level of Students across an episode. These range from very low engagement (1) to extremely and consistently high engagement (5) across the entire episode.

The *Narrative Record* also tracks the following *Tools*-specific behaviors that could be exhibited in Comparison classrooms as well:

- Positive Behavior Reinforcement by the Teacher or Assistant

- Behavior Reminders by the Teacher or Assistant
- Choral Responses from the Children (children are encouraged to call out answers)
- Teacher Paired Activities (meaning the teacher has assigned pairs of children to interact)
- Individual Scaffolding by the Teacher or Assistant
- Teacher Directed Private Speech (meaning the teacher has directed children to use private speech)
- Intentional Teacher Mistakes

For more information see: Farran, D.C. & Bilbrey, C. (2004). *Narrative Record Observation for Classrooms, Adapted for Tools of the Mind classrooms*. Nashville, TN: Peabody Research Institute, Vanderbilt University.

Teacher Observation in Preschools (TOP)

The *TOP* is a system for observing the teacher and assistant's behaviors in preschool classrooms across a daylong visit when the children are in the classroom. The *TOP* is based on a series of snapshots of the behaviors of both the teacher and assistant across a period of time when children and teachers are in the room. Each snapshot may be, by itself, an unreliable piece of information, but collectively, they combine to provide a picture of how the teacher and assistant are spending their time in a classroom. The teacher's behavior is observed for a 3-second window before scoring. Once scoring has been completed for the teacher, the same procedure is followed for the assistant in the classroom. Teacher and assistant are coded at the beginning of a "sweep;" children are coded immediately afterward. At the end of an observation, a total of 20 sweeps each was collected on the teacher and the assistant. The *TOP* measures:

- How much and to whom the teacher talks and listens
- The types of tasks in which the teacher or assistant is engaged
  - Instruction or Assessment
  - Management including administration, management, monitoring, and personal care
  - Behavior Approving or Disapproving
  - Social
  - None
- The level of ongoing instruction or assessment
  - None, Low, Basic Skills, Some Inferential, and Highly Inferential
- The areas of learning on which the teacher or assistant focuses
  - Specific Learning Focus: math, literacy, science, social studies
  - Other: art, music, fine motor, drama, etc.
  - No Learning Focus: no instruction or assessment
- The tone of the interactions the teacher or assistant has with the class

TOP data are not collected when children are out of the room. Only one sweep is allowed during a meal inside the room. The *TOP* focuses on times when teachers and children could interact in the classroom.

For More Information See: Bilbrey, C., Vorhaus, E., Farran, D. & Shufelt, S. (2007) *Teacher Observation in Preschools* (2008 revision). Tools of the Mind Adaptation (2010). Peabody Research Institute, Nashville, TN.

Child Observation in Preschools (COP)

The *COP* is a system for observing children's behaviors in preschool classrooms across a daylong visit. The *COP* is based on a series of snapshots of children's behaviors across a period of time when children and teachers are in the room. Each snapshot may be, by itself, an unreliable piece of information, but collectively, they combine to provide a picture of how children are spending their time in a classroom (as an aggregate), as well as information about individual differences among children in their activity preferences. A specific child is observed during a 3-second window and then coded across 9 dimensions before the observer moves to the next child. At the end of an observation, a total of 20 sweeps was collected on each child in the classroom. Consented children are identified by name; all others are identified as "Extra boy" or "Extra girl;" this process ensures that the behaviors of all the children in the classroom are observed. The *COP* measures:

- How much and to whom the children talk and listen
- The learning settings in which children are found
    - Whole Group (with and without teacher)
    - Small Group (with and without teacher)
    - Centers
    - Make Believe Play
    - Transitions
    - Other: classroom activities not captured by above, such as book reading at the beginning of nap.
- The different types of learning foci of the activities in which children are engaged
    - Specific Learning Focus: math, literacy, science, social studies
    - Other: art, music, fine motor, drama, etc.
    - No Learning Focus
- The level of involvement of the children

As with the *TOP*, *COP* codes are only collected when the children are in the room and learning interactions could take place. The *COP* is not coded during naptime -- stopping when the lights go off -- or during meal times. If meals happen in the room, one data collection sweep is allowed.

For more information see: Farran, D. et al. (2006), *Child Observation in Preschools* (2008 revision). Tools of the Mind Adaptation, (2010). Peabody Research Institute, Nashville, TN.

The Tools of the Mind Fidelity

The Tools of the Mind Fidelity (TAP) (Vorhaus & Meador 2010) captures the specific Tools of the Mind curriculum activities that occur within a prekindergarten classroom observation period along with information about the specific implementation steps that occur and mediators that are used. In addition, the curriculum developers furnished a list of behaviors that "should not" happen during

each activity that are also captured by observers. The Tools of the Mind Fidelity Measure provides an in-depth look at the degree of curriculum implementation across the year within experimental classrooms. Although this instrument was used in both Tools and comparison classrooms, relatively few Tools activities were ever coded in comparison rooms. TAP was used with the Narrative Record to document the amount of time activities occurred in the classroom.

For more information see: Vorhaus, E. & Meador, D. (2010). Tools of the Mind Curriculum Implementation Fidelity Checklist. Peabody Research Institute, Nashville, TN. (https://my.vanderbilt.edu/toolsofthemindevaluation/)

An extensive report of the fidelity of implementation using this instrument can be found in Meador, D., Nesbitt, K., & Farran, D. (September, 2015). *Experimental Evaluation of the Tools of the Mind Pre-K Curriculum: Fidelity of Implementation Technical Report.* (Working Report) Peabody Research Institute, Vanderbilt University. Available: https://my.vanderbilt.edu/toolsofthemindevaluation/

The Post Observation Rating Scale (PRS)

The *PRS* is completed immediately after a classroom is observed and is a 5-point Likert-type researcher-developed scale for rating classroom-level characteristics. This instrument was developed following extensive discussions with the *Tools of the Mind* curriculum developers, during which they identified classroom attributes that were most likely to be different between *Tools* classrooms and other early childhood classrooms. The *PRS* includes items regarding general classroom characteristics as well as teacher practices, classroom activities, and children's social and academic behaviors. Both observers completed the *PRS* independently following the visit and then combined their ratings into a single consensus set of ratings used for analysis.

PRS Subscale Descriptions

Subscale	Number of Items	Description
General	11	Items related to the general classroom atmosphere.
Center Time	4	Items characterizing children's play during centers.
Classroom Management	7	Items describing teacher- and child-level factors in classroom management.
Teacher Responsiveness	3	Items related to teachers' interactions with children.
Community	6	Items describing peer interactions.
Academic-Learning Related	5	Items characterizing children's behaviors and engagement during academic activities.

For more information see: Yun, C., Farran, D.C., Lipsey, M., Vorhaus, E., & Meador, D. (2010). *Prekindergarten Classroom Dynamics Rating Scale.* Peabody Research Institute, Nashville, TN. https://my.vanderbilt.edu/toolsofthemindevaluation/

Environmental Scan and Checklist

The Environmental Scan and Checklist (Vorhaus, Meador, & Farran, 2010) is an observational tool designed to gauge a classroom's environment, themes, and materials. It is derived from a list of early childhood materials the *Tools of the Mind* developers indicate should be available in the classroom. The checklist focuses on the play centers and materials accessible to children.

Characteristics of the Children and Teachers

In this section, we summarize the demographic characteristics of the participating children and describe the background and experience of the teachers and assistant teachers in the study.

In cohort 1, 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 the upper panel of 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. There was a small but statistically significant difference between conditions on ethnicity at baseline. 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. In cohort 2, 266 children (147 Tools, 119 comparison) were age-eligible for pre-k and consented to participate in the study in the fall of 2011. Demographics for cohort 2 children are presented in the lower panel of Table 1. The cohort 2 sample was balanced on gender and ethnically diverse. Cohort 2 included a larger proportion of ELL students than cohort 1. There were no significant differences by condition on any of the demographic variables shown in Table 1 in cohort 2.

Average ages for the Tools and comparison groups in both cohorts are shown in Table 2. In both cohorts, there were significant differences in age between conditions at baseline. However, the magnitude of the differences was minor and amounted to about 2-3 weeks between the groups. 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 from some schools.

Classroom demographics are shown in Table 3 and reflect the diversity of students in prekindergarten classrooms in Tennessee and North Carolina. On average, classes included 15-17 students across conditions and cohorts. As seen in the student-level demographics, cohort 2 classrooms tended to contain larger proportions of ELL students than cohort 1 classrooms.

Eighty teachers participated in the study, with 42 assigned to the Tools condition and 38 assigned to the business-as-usual comparison condition. Information about years of teaching experience and education and licensure for lead teachers is shown in Tables 4 and 5. 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; experience and education for the assistant teachers is presented in Tables 6 and 7.

Cohort 1 comparison classrooms used a variety of curricula, with the modal one being Creative Curriculum. In cohort 2, the district had recently adopted Opening the World of Learning and Building Blocks as the standard pre-k curricula. Training on OWL happened for half the teachers while the Tools assigned teachers were being trained.

Table 1: Demographics for Tools of the Mind and Comparison Children

	Tools Condition		Comparison Condition	
	N	%	*n*	%
Cohort 1				
Male	261	52.4	218	57.5
White	192	38.6	157	41.4
Black	145	29.1	86	22.7
Hispanic	118	23.7	95	25.1
Asian	32	6.4	21	5.5
Multi-racial	4	0.8	16	4.2
Other Minority	7	1.4	4	1.1
ELL (=yes)	140	28.1	117	30.9
IEP (=yes)	68	13.7	58	15.3
FRPL (=yes)	329[a]	86.1	293[a]	88.0
Cohort 2				
Male	80	54.4	64	53.8
White	56	38.1	62	52.1
Black	45	30.6	24	20.2
Hispanic	39	26.5	25	21.0
Asian	2	1.4	4	3.4
Multi-racial	4	2.7	2	1.7
Other Minority	1	0.7	2	1.7
ELL (=yes)	60	40.8	62	52.1
IEP (=yes)	14	9.5	6	5.0

Note. ELL = English language learner; IEP = individualized education plan; FRPL = free or reduced price lunch. FRPL data were not available for Cohort 2 children. In Cohort 1, there were no significant differences between the two conditions on gender, ELL status, IEP status, or proportion of student on free or reduced price lunch. However there was a significant difference on ethnicity ($p < .05$). In Cohort 2, there were no significant differences between the two conditions on gender, ethnicity, ELL status or IEP status.

[a]Missing for 116 Tools children and 46 comparison children; percentages reflect percent of non-missing cases.

Table 2: Average Age of Tools of the Mind and Comparison Children

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Age (months) at pretest[a]	494	54.1 (3.6)	372	54.6 (3.7)
Age (months) at posttest	467	61.5 (3.5)	349	62.0 (3.7)
Cohort 2				
Age (months) at pretest	145	54.6 (3.6)	119	55.5 (3.5)
Age (months) at posttest[b]	141	62.0 (3.4)	114	62.9 (3.5)

[a]In Cohort 1, there was a significant difference between the two conditions on age at pretest ($p < .05$).
[b]In Cohort 2, there was a significant difference between the two conditions on age at posttest ($p < .05$).

Table 3: Classroom Demographics

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Fall Class Size	32	17.3 (2.0)	28	17.9 (1.6)
Spring Class Size	32	17.3 (2.0)	28	17.7 (1.7)
Fall ELL Status (% of class)	32	28.8 (26.4)	28	27.9 (23.0)
Spring ELL Status (% of class)	32	29.4 (26.0)	28	29.9 (22.6)
Fall IEP Status (% of class)	32	9.9 (8.6)	28	12.2 (9.3)
Spring IEP Status (% of class)	32	11.1 (9.7)	28	13.4 (9.6)
Cohort 2				
Fall Class Size	10	15.0 (0.0)	10	15.0 (0.0)
Spring Class Size	10	14.9 (0.3)	10	15.0 (0.0)
Fall ELL Status (% of class)	10	38.7 (30.9)	10	46.7 (17.5)
Spring ELL Status (% of class)	10	41.3 (27.0)	10	49.3 (17.3)
Fall IEP Status (% of class)	10	4.0 (4.7)	10	5.3 (6.1)
Spring IEP Status (% of class)	10	10.7 (10.0)	10	7.3 (4.9)

Table 4: Years of Experience for Lead Teachers

	Tools Condition			Comparison Condition		
	n	*M (SD)*	Range	*n*	*M (SD)*	Range
Cohort 1						
Years Teaching	32	12.0 (8.0)	2-30	28	12.1 (8.4)	1-34
Years Teaching Pre-K	32	7.7 (5.2)	2-22	28	6.6 (5.1)	1-17
Cohort 2						
Years Teaching	10	11.9 (11.9)	1-34	10	17.0 (7.5)	7-31
Years Teaching Pre-K	10	7.0 (6.2)	1-16	10	10.7 (5.9)	2-20

Table 5: Education and Licensure for Lead Teachers

	Tools Condition		Comparison Condition	
	n	%	*n*	%
Cohort 1				
Education Level				
Bachelor's Degree	12	37.4	17	60.7
Some Graduate Coursework	11	34.4	5	17.9
Master's Degree	9	28.1	6	21.4
Licensure Area				
Early Childhood (Birth-K)	19	59.4	18	64.3
Elementary Ed (Pre-K-3,4,8)	10	31.3	9	32.1
Early Childhood and Special Ed.	3	9.4	1	3.6
Cohort 2				
Education Level				
Bachelor's Degree	8	80.0	6	60.0
Some Graduate Coursework	1	10.0	4	40.0
Master's Degree	1	10.0	0	00.0
Licensure Area				
Early Childhood (Birth-K)	7	70.0	7	70.0
Elementary Ed (Pre-K-3,4,8)	2	20.0	2	20.0
Early Childhood and Special Ed.	1	10.0	1	10.0

Table 6: Years of Experience for Assistant Teachers

	Tools Condition			Comparison Condition		
	n	*M (SD)*	Range	*n*	*M (SD)*	Range
Cohort 1						
Years Teaching Pre-K	32	4.7 (3.1)	0.1-17.0	28	4.2 (3.1)	0.3-12.0
Years Working with Lead Teacher	32	3.5 (3.5)	0.1-17.0	28	2.5 (1.7)	0.3-8.0
Cohort 2						
Years Teaching Pre-K	10	5.2 (3.2)	0.1-10.0	10	7.7 (4.5)	1.0-17.0
Years Working with Lead Teacher	10	2.1 (1.8)	0.1-5.0	10	4.8 (3.5)	0.1-11.0

Table 7: Education for Assistant Teachers

	Tools Condition		Comparison Condition	
	n	%	*n*	%
Cohort 1[a]				
High School Diploma or GED	5	13.9	8	25.8
CDA	8	22.2	6	19.4
Montessori Training	0	0.0	1	3.2
Some College	3	8.3	2	6.5
Associate's Degree	5	13.9	5	16.1
Bachelor's Degree	13	36.1	6	19.4
Master's Degree	2	5.6	1	3.2
Other	0	0.0	2	6.5
Cohort 2				
High School Diploma or GED	0	0.0	2	20.0
CDA	2	20.0	1	10.0
Associate's Degree	6	60.0	6	60.0
Bachelor's Degree	2	20.0	1	10.0

Note: GED=General Education Development; CDA=Child Development Associate.

[a]In Cohort 1, seven classrooms had two assistants and education information is reported on all assistants. However, data on years of experience in Table 6 was only collected for the primary assistant in the classroom, or the one with which the lead teacher had worked for the longest amount of time.

Table 8: Curricula Used by Comparison Condition Teachers

	Comparison Condition *n*
Cohort 1	
Creative Curriculum	15
Literacy First	4
Houghton Mifflin	3
Scott Foresman	5
CSEFEL	6
Other[a]	10
Cohort 2	
Building Blocks	10
OWL	10
Other[b]	3

Note: Many comparison condition teachers (14 in Cohort 1 and all 10 Cohort 2 teachers) reported using more than one curriculum.

[a]The "Other" category for Cohort 1 teachers includes the following curricula: Letter People, DLM Early Childhood Express, Incredible Years, Early Years, P.A.S.T., Conscious Discipline, Handwriting Without Tears, Frog Street Press Letter Books, and Talking about Touching.

[b]The "Other" category for Cohort 2 teachers includes the following curricula: Incredible Years and Work Sampling System.

Child Outcomes

The following section describes the results of the study on children's achievement and self-regulation outcomes, separately for the two cohorts. 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 in cohort 1, 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 1st 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.

Tables 9-11 report means, standard deviations, and sample sizes for each outcome variable at each data collection point by condition for cohort 1. Tables 12-14 report the results of the impact analysis on each outcome variable and provide standardized mean difference effect sizes for each contrast. Effect sizes were computed using the adjusted means from the regression models and the unadjusted standard deviations. Across the eight academic achievement outcomes measured in pre-k and kindergarten, there was one statistically significant effect at the end of pre-k (oral comprehension) and two statistically significant effects at the end of kindergarten (letter word and quantitative concepts). There was also a significant effect at the end of kindergarten on the composite achievement measure (computed from the sum of the z-scores for the seven outcomes). All of the significant effects favored the comparison condition. Across the eight outcomes collected at the end of 1st grade, there was one statistically significant effect (spelling), which also favored the comparison condition.

The impact estimates and effect sizes for the six self-regulation outcomes and the composite self-regulation score for cohort 1 are shown in Table 13. There were no statistically significant differences between conditions on self-regulation at the end of pre-

k. At the end of kindergarten, there was a statistically significant difference between conditions on the backward digit span assessment. At the end of 1st grade, there were significant condition differences on the copy design task and on the composite self-regulation score. As with the academic achievement outcomes, the differences on these outcomes favored the comparison condition.

Results of the impact analyses on teacher and assessor assessments for cohort 1 are shown in Table 14. There were no significant condition differences on any of the assessments at any time point.

In addition to the main effects analyses reported in Tables 12-14, we tested several condition by subgroup interactions for cohort 1 in order to determine if the Tools curriculum was more effective for particular subgroups of students. The statistically significant interactions are shown in bold and italics in the table. Across the table as a whole, we identify no consistent pattern of subgroup effects for any outcome, subgroup, or time point. These rather idiosyncratic findings do not result in any defensible conclusions about whether Tools is more or less effective for certain subgroups of children.

Parallel results for the children in cohort 2 are shown in Tables 16-22. Note that we elected to report the impact results for the two cohorts separately because changes in the amount and focus of coaching were implemented for cohort 2 based on the cohort 1 findings. The coaching changes included more focused attention on the implementation of Make Believe Play and were expected to result in larger impacts on achievement and self-regulation than those observed for cohort 1.

Of the consented children in cohort 2, 263 had pretest scores on one or more direct assessments of achievement or self-regulation; we collected teacher reports of behavior on 266 children. The consented children who did not receive either a pretest or a teacher report in the fall of 2011 had either withdrawn from preschool prior to the assessment period or refused to complete one or more of the assessments. In the spring of 2012, 253 children had at least one direct assessment of achievement or self-regulation and teacher reports were received on 254 children.

We obtained follow-up assessments on 810 children and teacher reports on 249 children in the spring of 2013 (when most children were completing kindergarten). In the spring of 2014 (at the end of most children's 1st grade year), we obtained assessments and teacher reports on 240 children.

At baseline, there was a significant difference between the conditions on the spelling assessment. There were no other condition differences on the measures at baseline for cohort 2. Means, standard deviations and sample sizes for all outcomes for cohort 2 are presented in Tables 16-18.

The cohort 2 impact estimates and effect sizes for the academic achievement outcomes are shown in Table 19. There was one statistically significant condition difference—on picture

vocabulary at the end of kindergarten. This difference favored the Tools condition. The impact estimates for the self-regulation outcomes are shown in Table 20. There were no significant differences between the Tools and comparison conditions observed for any of the self-regulation outcomes.

The results for cohort 2 teacher and assessor ratings are shown in Table 21. At the end of kindergarten and 1st grade, a statistically significant condition difference was observed on the Adaptive Language Inventory. This difference favored the comparison group.

Finally, as above, we tested several condition by subgroup interactions for the cohort 2 sample. The results of these analyses are shown in Table 22. Again, no consistent pattern of significant interaction effects is observed across subgroups, outcomes, or time points.

Cohort 1

Table 9: Cohort 1 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 of 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)

Note: There were no significant baseline condition differences at the beginning of Pre-K.

Table 10: Cohort 1 Means, Standard Deviations and Sample Sizes for the Self-Regulation Direct Assessment by Condition at Each Assessment

	Fall Pre-K		Spring Pre-K		Spring of K		Spring of 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. There were no significant baseline condition differences at the beginning of Pre-K.

Table 11: Cohort 1 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	
	n	Unadjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted 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. There were no significant baseline condition differences at the beginning of Pre-K.

Table 12: Cohort 1 Impact of Tools of the Mind on Academic Achievement at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*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, computed using the adjusted means and unadjusted standard deviations.
* $p < .05$.

Table 13: Cohort 1 Impact of Tools of the Mind on Self-Regulation at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*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, computed using the adjusted means and unadjusted standard deviations.
* $p < .05$.

Table 14: Cohort 1 Impact of Tools of the Mind on Teacher and Assessor Ratings at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*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, computed using the adjusted means and unadjusted standard deviations. CFBRS=Cooper-Farran Behavior Rating Scales.
* $p < .05$.

Table 15: Cohort 1 Subgroup Analyses: *p*-values from Tests of Interactions

	End of Pre-K					End of Kindergarten					End of 1st Grade				
		Condition by					Condition by					Condition by			
Variable	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. DCCS=Dimensional Change Card Sort; HTKS=Head-Toes-Knees-Shoulders.

Cohort 2

Table 16: Cohort 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 of 1st Grade	
	n	Mean *(SD)*	*n*	Mean *(SD)*	*n*	Mean *(SD)*	*n*	Mean *(SD)*
Tools Condition								
Letter Word	144	88.2 (13.5)	139	96.7 (11.1)	138	110.5 (11.6)	133	108.3 (11.4)
Spelling[a]	144	79.6 (10.6)	139	87.0 (15.0)	138	102.0 (13.0)	133	99.5 (14.7)
Academic Knowledge	144	80.9 (21.4)	139	88.0 (16.4)	138	95.1 (11.7)	133	93.4 (10.7)
Oral Comprehension	145	88.0 (13.7)	140	92.3 (14.7)	138	96.7 (12.8)	133	97.4 (13.7)
Picture Vocabulary	145	87.5 (22.7)	140	93.2 (13.2)	138	94.5 (10.6)	133	92.5 (10.7)
Applied Problems	145	90.5 (14.0)	140	98.2 (10.6)	138	101.6 (12.1)	133	99.4 (11.5)
Quantitative Concepts	145	84.7 (12.1)	140	90.6 (11.8)	138	96.5 (10.9)	133	93.2 (10.8)
Passage Comprehension							133	95.7 (12.1)
Comparison Condition								
Letter Word	119	85.7 (12.0)	114	95.5 (10.9)	111	108.9 (10.9)	107	105.8 (10.6)
Spelling[a]	119	75.1 (11.7)	114	84.5 (13.7)	111	99.1 (12.2)	107	95.4 (13.8)
Academic Knowledge	119	75.1 (18.6)	114	84.8 (15.3)	111	91.3 (11.0)	107	88.9 (11.3)
Oral Comprehension	119	83.6 (11.3)	114	89.0 (14.1)	111	92.4 (13.6)	107	94.1 (12.3)
Picture Vocabulary	119	84.1 (22.1)	114	88.5 (17.2)	111	90.1 (10.8)	107	89.0 (10.4)
Applied Problems	119	88.6 (13.3)	114	96.7 (12.0)	111	99.4 (12.5)	107	99.1 (11.0)
Quantitative Concepts	119	81.5 (10.8)	114	92.1 (12.5)	111	94.7 (9.2)	107	91.5 (9.2)
Passage Comprehension							107	93.6 (9.6)

[a]There was a significant baseline condition difference at the beginning of Pre-K on Spelling (p = .049).

Table 17: Cohort 2 Means, Standard Deviations and Sample Sizes for the Self-Regulation Direct Assessment by Condition at Each Assessment

	Fall Pre-K		Spring Pre-K		Spring of K		Spring of 1st Grade	
	n	Unadjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*
Tools Condition								
Forward Span	144	2.6 (1.0)	139	3.3 (1.1)	138	4.2 (1.1)	133	4.6 (1.0)
Backward Span	144	1.0 (1.1)	139	1.5 (1.3)	138	3.2 (1.4)	133	3.8 (1.2)
DCCS	144	1.2 (0.6)	139	1.7 (0.6)	138	2.2 (0.5)	133	2.7 (1.0)
Copy Design	145	0.9 (1.4)	140	4.4 (2.6)	138	8.0 (2.7)	133	8.2 (2.7)
HTKS	144	9.0 (13.0)	140	21.8 (17.0)	138	38.4 (12.6)	133	44.9 (9.3)
Peg Tapping	144	4.1 (6.2)	140	9.1 (5.7)	138	13.8 (3.4)	133	14.9 (2.1)
Comparison Condition								
Forward Span	119	2.5 (1.2)	114	3.1 (1.2)	112	4.2 (1.1)	107	4.6 (0.9)
Backward Span	119	1.1 (1.1)	114	1.4 (1.3)	112	3.0 (1.2)	107	3.7 (1.2)
DCCS	119	1.2 (0.5)	114	1.6 (0.5)	112	2.1 (0.5)	107	2.4 (0.9)
Copy Design	119	0.7 (1.2)	114	4.0 (2.7)	111	7.6 (2.5)	107	8.1 (2.9)
HTKS	119	7.5 (11.3)	114	19.5 (16.9)	111	37.5 (13.4)	107	45.1 (11.0)
Peg Tapping	119	3.4 (5.3)	114	9.0 (6.2)	111	13.5 (3.9)	107	14.8 (2.7)

Note. DCCS=Dimensional Change Card Sort; HTKS=Head-Toes-Knees-Shoulders. There were no significant baseline condition differences at the beginning of Pre-K.

Table 18: Cohort 2 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	
	n	Unadjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*	*n*	Adjusted Mean *(SD)*
Tools Condition								
CFBRS Interpersonal Skills Scale	147	5.2 (0.8)	142	5.5 (0.8)	138	5.6 (1.2)	133	5.7 (0.9)
CFBRS Work-Related Skills Scale	147	4.4 (0.9)	142	5.0 (0.9)	138	4.9 (1.2)	133	4.8 (1.2)
Adaptive Language Inventory	147	2.8 (0.6)	142	3.2 (0.7)	138	3.1 (0.6)	133	2.9 (0.6)
Assessor Ratings of Attention	144	2.3 (0.7)	141	2.5 (0.6)	138	2.8 (0.4)	133	2.9 (0.2)
Comparison Condition								
CFBRS Interpersonal Skills Scale	119	5.2 (0.9)	112	5.4 (0.9)	112	5.5 (1.3)	108	5.8 (1.0)
CFBRS Work-Related Skills Scale	119	4.3 (0.9)	112	4.8 (1.0)	112	5.0 (1.4)	108	5.0 (1.2)
Adaptive Language Inventory	119	2.7 (0.6)	112	3.2 (0.7)	112	3.3 (0.8)	108	3.2 (0.8)
Assessor Ratings of Attention	119	2.4 (0.7)	114	2.6 (0.7)	112	2.8 (0.3)	107	2.9 (0.3)

Note. CFBRS=Cooper-Farran Behavior Rating Scales. There were no significant baseline condition differences at the beginning of Pre-K.

Table 19: Cohort 2 Impact of Tools of the Mind on Academic Achievement at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*b*	*SE*	*d*	*b*	*SE*	*d*	*b*	*SE*	*d*
Letter Word	-1.27	3.99	-0.06	-2.63	3.00	-0.11	1.71	3.11	0.07
Spelling	1.03	3.05	0.04	0.90	2.77	0.05	2.70	2.85	0.12
Academic Knowledge	-0.72	2.04	-0.04	1.37	1.65	0.11	2.03	1.43	0.18
Oral Comprehension	-0.74	1.66	-0.04	1.15	1.99	0.08	0.40	1.54	0.03
Picture Vocabulary	2.73	1.66	0.16	2.95*	1.05	0.26	2.06	1.04	0.19
Applied Problems	0.87	2.83	0.04	1.53	2.32	0.09	-1.29	1.82	-0.09
Quantitative Concepts	-4.39	2.35	-0.32	-0.78	1.38	-0.07	0.46	1.41	0.04
Composite Achievement	-0.32	0.59	-0.06	0.17	0.47	0.03	0.33	0.51	0.06
Passage Comprehension							-0.91	2.13	-0.05

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

Table 20: Cohort 2 Impact of Tools of the Mind on Self-Regulation at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*b*	*SE*	*d*	*b*	*SE*	*d*	*b*	*SE*	*d*
Forward Span	0.16	0.18	0.15	0.00	0.14	0.00	-0.02	0.11	-0.02
Backward Span	0.18	0.17	0.14	0.23	0.21	0.18	0.07	0.19	0.06
DCCS	0.08	0.08	0.14	0.14	0.09	0.27	0.26	0.15	0.27
Copy Design	0.42	0.38	0.16	0.40	0.32	0.15	0.10	0.34	0.04
HTKS	2.30	2.15	0.14	0.88	1.70	0.07	-0.18	1.32	-0.02
Peg Tapping	0.11	0.81	0.02	0.24	0.55	0.07	0.12	0.32	0.05
Composite Self-Regulation	0.52	0.42	0.14	0.68	0.47	0.17	0.27	0.43	0.07

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

Table 21: Cohort 2 Impact of Tools of the Mind on Teacher and Assessor Ratings at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K			End of Kindergarten			End of First Grade		
Variable	*b*	*SE*	*d*	*b*	*SE*	*d*	*b*	*SE*	*d*
CFBRS Interpersonal Skills Scale	0.15	0.14	0.18	0.09	0.18	0.07	-0.11	0.17	-0.12
CFBRS Work-Related Skills Scale	0.23	0.20	0.24	-0.01	0.16	-0.01	-0.19	0.18	-0.17
Adaptive Language Inventory	0.02	0.14	0.03	-0.21*	0.08	-0.28	-0.28*	0.08	-0.40
Self-Regulation Assessor Ratings	-0.15	0.11	-0.23	-0.02	0.06	-0.05	0.03	0.03	0.10

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

Table 22: Cohort 2 Subgroup Analyses: *p*-values from Tests of Interactions

	End of Pre-K				End of Kindergarten				End of 1st Grade			
		Condition by				Condition by				Condition by		
Variable	Pretest	Gender	ELL	Age	Pretest	Gender	ELL	Age	Pretest	Gender	ELL	Age
Letter Word	0.814	0.730	0.514	0.064	0.615	0.239	0.362	0.836	0.820	0.237	0.537	0.461
Spelling	0.833	0.839	0.998	0.527	0.422	0.999	0.900	0.727	0.803	0.260	0.840	0.111
Academic Knowledge	***0.039***	0.286	***0.007***	0.999	0.379	0.682	0.136	0.973	0.514	0.282	0.513	0.406
Oral Comprehension	0.560	0.494	0.209	0.939	0.822	0.795	0.756	0.195	0.666	0.773	0.992	0.673
Picture Vocabulary	***0.001***	0.722	0.654	0.064	0.985	0.903	0.934	0.943	0.987	0.984	0.776	0.589
Applied Problems	0.555	0.969	0.786	0.123	0.197	0.860	0.382	0.278	0.173	0.270	0.644	***0.003***
Quantitative Concepts	0.309	0.276	0.444	0.092	0.411	0.213	0.563	0.757	0.751	0.955	0.656	0.416
Passage Comprehension									0.254	0.378	0.329	0.976
Forward Span	0.270	0.270	0.465	0.817	0.956	0.320	0.838	0.452	0.400	0.993	0.063	0.876
Backward Span	0.349	0.281	0.342	0.220	0.616	0.599	0.854	0.328	0.636	0.244	0.276	0.351
DCCS	0.813	0.425	0.636	0.673	***0.039***	0.330	0.154	0.291	0.368	0.329	0.307	0.548
Copy Design	0.229	0.671	0.344	0.684	0.828	0.980	0.835	0.130	0.815	0.313	0.802	***0.022***
HTKS	0.856	0.247	0.640	0.118	0.607	0.118	0.238	0.108	0.901	0.523	0.427	0.515
Peg Tapping	0.076	0.792	0.650	0.111	0.308	0.838	0.234	0.410	0.869	0.097	0.167	0.844
Interpersonal Skills	***0.031***	0.344	0.853	0.957	0.328	0.270	0.309	0.793	0.097	0.921	0.527	0.066
Work-related Skills	0.743	0.641	0.911	0.505	0.604	0.610	0.078	0.401	0.254	0.233	0.269	0.783
Adaptive Language	0.642	0.557	0.687	0.477	0.759	0.230	0.153	0.101	0.940	0.751	0.406	0.160
Assessor Ratings	0.173	0.544	0.467	0.806	0.977	0.479	0.368	0.427	0.056	0.517	0.111	0.745

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

Narrative Record

The following section summarizes the results on some of the key indicators derived from the Narrative Record. As described above, observers used the Narrative Record to record the amount and timing of a variety of activities throughout the school day. Table 23 shows the average proportion of the school day observed in the basic activities that make up a preschool day for each of the two cohorts. There are few differences between the conditions, despite the fact that one group was implementing a very different curriculum.

It is notable that nearly 50% of the school day was spent on non-instructional activities such as napping and eating. In addition, the proportion of the day spent in non-instructional transitions between activities averaged 11-13% across conditions and cohorts. It is also notable that the proportion of time spent in centers was larger in cohort 2 classrooms (23% in cohort 2 vs. 12% in cohort 1 Tools classrooms and 21% vs. 16% in comparison classrooms). Cohort 2 came exclusively from one school system while cohort 1 included five school systems; differences in how the day was structured could reflect a school system effect.

In both cohorts, children in Tools classrooms spent more time in small group activities. Putting the children into smaller sized groups is an integral component of the Tools curriculum. Teachers work with half the children while the Assistant works with the other half daily during Play Planning episodes and throughout the week for other activities. The classroom is always split in half and the Assistant is expected to play a role equal to the teacher. These "small groups" range in size from 8 to 10 children, depending on the size of the class.

The Narrative Record data may also be broken down by content area. The proportions of the school day spent in the various content areas by condition and cohort are presented in Table 24. Many preschool activities involve a mix of content areas; mixed content occurs during center times as well as whole group sessions where content is covered sequentially for less than a minute each. We found relatively large proportions of the school day spent in mixed content in both conditions and cohorts. As expected for preschool classrooms, language- and literacy-related content is the next largest content area after mixed content. Notable in Table 24 are the very small proportions of time spent on math and science activities.

Table 25 shows the proportion of daily episodes in which particular teacher behaviors were observed. These behaviors were specifically added because the Tools developers believed teachers would have different amounts of them when compared to business as usual early childhood classrooms. Tools classrooms were similar to comparison classroom on most behaviors with the exception of Paired Activities and Teacher Encourages Private Speech. Tools teachers enacted many more instances of those behaviors.

Table 23: Proportion of Time Observed in Preschool Activities

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Whole Group	32	.18 (.06)	28	.17 (.07)
Centers	32	.12 (.04)	28	.16 (.06)
Small Group	32	.07 (.04)	28	.004 (.01)
Small Group Centers	32	.03 (.04)	28	.06 (.04)
Transition	32	.13 (.04)	28	.13 (.04)
Meal, Nap, Out, or Not Observed	32	.47 (.10)	28	.48 (.07)
Cohort 2				
Whole Group	10	.12 (.02)	10	.12 (.04)
Centers	10	.23 (.03)	10	.21 (.05)
Small Group	10	.06 (.02)	10	.01 (.01)
Small Group Centers	10	.05 (.04)	10	.04 (.04)
Transition	10	.11 (.02)	10	.13 (.05)
Meal, Nap, Out, or Not Observed	10	.43 (.04)	10	.49 (.08)

Table 24: Proportion of Time Observed in Preschool Content Areas

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Reading Readiness	32	.14 (.04)	28	.08 (.04)
Reading	32	.02 (.02)	28	.02 (.02)
Language Arts	32	.01 (.01)	28	.02 (.02)
Literacy	32	.11 (.04)	28	.04 (.03)
Math	32	.03 (.02)	28	.02 (.01)
Science	32	.01 (.01)	28	.02 (.02)
Social Studies	32	.02 (.01)	28	.01 (.01)
Morning Meeting	32	.005 (.005)	28	.01 (.01)
Art	32	.003 (.01)	28	.01 (.01)
Music & Movement	32	.03 (.01)	28	.02 (.01)
Mixed Content	32	.18 (.06)	28	.23 (.06)
Other	32	.004 (.01)	28	.003 (.01)
Meal, Nap, Out, Not Observed, Transition	32	.59 (.08)	28	.60 (.06)
Cohort 2				
Reading Readiness	10	.10 (.01)	10	.06 (.03)
Reading	10	.01 (.01)	10	.004 (.01)
Language Arts	10	.002 (.003)	10	.01 (.01)
Literacy	10	.08 (.01)	10	.04 (.02)
Math	10	.02 (.01)	10	.03 (.02)
Science	10	.01 (.01)	10	.01 (.01)
Social Studies	10	.01 (.01)	10	.003 (.01)
Morning Meeting	10	.001 (.002)	10	.003 (.003)
Art	10	.00 (.00)	10	.01 (.01)
Music & Movement	10	.03 (.01)	10	.02 (.01)
Mixed Content	10	.31 (.04)	10	.27 (.04)
Other	10	.001 (.003)	10	.00 (.00)
Meal, Nap, Out, Not Observed, Transition	10	.54 (.03)	10	.61 (.05)

Table 25: Proportion of Narrative Episodes in which Teacher Behaviors Observed

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Positive Behavior Reinforcers	32	.29 (.10)	28	.30 (.08)
Behavior Reminders	32	.39 (.15)	28	.42 (.12)
Choral Responses	32	.18 (.05)	28	.16 (.04)
Teacher Paired Activities	32	.06 (.03)	28	.01 (.01)
Individual Scaffolding	32	.11 (.04)	28	.09 (.05)
Teacher Encouraged Private Speech	32	.04 (.04)	28	.002 (.004)
Teacher Intentional Mistakes	32	.01 (.01)	28	.01 (.01)
Cohort 2				
Positive Behavior Reinforcers	10	.37 (.08)	10	.30 (.08)
Behavior Reminders	10	.35 (.14)	10	.34 (.07)
Choral Responses	10	.13 (.04)	10	.12 (.05)
Teacher Paired Activities	10	.08 (.03)	10	.004 (.007)
Individual Scaffolding	10	.18 (.04)	10	.11 (.03)
Teacher Encouraged Private Speech	10	.04 (.04)	10	.003 (.007)
Teacher Intentional Mistakes	10	.01 (.01)	10	.01 (.01)

Teacher Observation in Preschool

The *Teacher Observation in Preschool* (TOP) is a system for observing the teacher and assistant's behaviors in preschool classrooms across a daylong visit when the children are in the classroom. The *TOP* is based on a series of snapshots of the behaviors of both the teacher and assistant across a period of time when children and teachers are in the room. The TOP depicts classroom conditions from the teacher's perspective, a top-down approach, or instruction as delivered. It quantifies teacher behaviors such as instances of instruction, assessment, management, and behavior approval or disapproval. The TOP also records level and focus of instruction. Several of the key variables obtainable from the TOP are presented in this section. Results are presented for most variables as proportions of sweeps (or observation snapshots) for which a particular behavior was observed. These proportions should not be interpreted as a proportion of the school day but rather proportions of the observation (TOP is not coded during outdoor times, naps and only once during mealtimes.)

Table 26 reports the proportion of observational sweeps the teachers were observed providing instruction across a variety of content areas. As with the Narrative Record results shown above, teachers were observed not performing any instruction in over half of the sweeps, and language and literacy-related content was the most frequently observed content in the classrooms. There were condition differences in the types of literacy activities observed across cohorts, with comparison teachers engaging in more code-based instruction and Tools teachers doing more literacy.

The TOP also recorded a variety of teacher tasks, including whether they were engaging in instruction, management, or administrative activities. The results for teacher tasks are shown in Table 27. In both cohorts, Tools teachers were more often observed instructing than comparison teachers were. Other behaviors were remarkably similar.

Table 28 presents the average teacher tone observed in each type of task. Teacher Tone was rated on a 5-point scale with 5 being the most positive. Overall Tools teachers in cohort 1 appeared to display a more positive affect across tasks. There were no differences between Tools and comparison teachers in tone in cohort 2.

Verbal behaviors of teachers (talking, listening and to whom) are summarized in Table 29. Although the Tools developers predicted that teachers in Tools classrooms would talk less than comparison teachers, the observed rate of talking was the same or a little more for Tools teachers. The rate of listening to children was quite low in all cohorts, both conditions although cohort 2 Tools teachers were observed listening to children the most.

The average levels of instruction (rated on a 4 point scale) for cohorts and conditions are presented in Table 30. Across all groups, the levels averaged below a basic level. Tools teachers did not engage in more inferential/thinking interactions with children than comparison teachers.

Table 26: Teacher Instructional Learning Focus (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Literacy	32	.12 (.05)	28	.06 (.04)
Code-Based Skills	32	.04 (.04)	28	.08 (.05)
Reading	32	.06 (.03)	28	.05 (.04)
Math	32	.06 (.04)	28	.08 (.06)
Science	32	.03 (.03)	28	.05 (.03)
Social Studies	32	.03 (.02)	28	.03 (.03)
Drama	32	.04 (.04)	28	.005 (.01)
Other	32	.07 (.04)	28	.05 (.05)
None	32	.55 (.10)	28	.60 (.10)
Cohort 2				
Literacy	10	.12 (.05)	10	.08 (.04)
Code-Based Skills	10	.03 (.03)	10	.06 (.04)
Reading	10	.08 (.03)	10	.05 (.03)
Math	10	.08 (.04)	10	.09 (.04)
Science	10	.06 (.04)	10	.05 (.04)
Social Studies	10	.01 (.02)	10	.01 (.01)
Drama	10	.03 (.04)	10	.01 (.01)
Other	10	.06 (.03)	10	.04 (.05)
None	10	.53 (.08)	10	.62 (.07)

Table 27: Teacher Type of Task (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Instruct and Assess	32	.45 (.10)	28	.40 (.11)
Manage	32	.22 (.08)	28	.23 (.08)
Behavior Approving	32	.03 (.03)	28	.05 (.03)
Behavior Disapproving	32	.06 (.05)	28	.06 (.06)
Personal Care	32	.05 (.03)	28	.05 (.04)
Monitor	32	.10 (.08)	28	.12 (.07)
Administrative	32	.03 (.03)	28	.03 (.03)
Social	32	.03 (.04)	28	.04 (.04)
None[a]	32	.02 (.02)	28	.02 (.02)
Cohort 2				
Instruct and Assess	10	.47 (.08)	10	.38 (.07)
Manage	10	.27 (.10)	10	.37 (.07)
Behavior Approving	10	.05 (.03)	10	.05 (.04)
Behavior Disapproving	10	.04 (.03)	10	.02 (.03)
Personal Care	10	.02 (.02)	10	.04 (.03)
Monitor	10	.10 (.05)	10	.08 (.04)
Administrative	10	.02 (.01)	10	.02 (.02)
Social	10	.04 (.04)	10	.03 (.03)
None[a]	10	.00 (.00)	10	.01 (.02)

[a]None was coded when the teacher was engaged in a behavior unrelated to the classroom such as talking on a cell phone or eating separately from the children.

Table 28: Teacher Average Tone by Type Task

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Instruct	32	3.64 (0.19)	28	3.55 (0.22)
Manage	32	3.24 (0.17)	28	3.19 (0.19)
Behavior Approving	22	3.88 (0.24)	27	3.74 (0.36)
Behavior Disapproving	26	3.01 (0.44)	22	2.95 (0.32)
Personal Care	32	3.23 (0.30)	24	3.26 (0.31)
Monitor	29	3.29 (0.33)	27	3.21 (0.25)
Administrative	26	3.16 (0.33)	20	3.14 (0.32)
Social	21	3.71 (0.34)	21	3.66 (0.49)
None[a]	19	3.11 (0.28)	13	3.00 (0.00)
Cohort 2				
Instruct	10	3.71 (0.21)	10	3.70 (0.24)
Manage	10	3.26 (0.13)	10	3.25 (0.21)
Behavior Approving	9	3.79 (0.44)	9	3.94 (0.17)
Behavior Disapproving	8	3.18 (0.41)	6	3.00 (0.00)
Personal Care	6	3.08 (0.20)	8	3.19 (0.37)
Monitor	10	3.21 (0.21)	10	3.07 (0.11)
Administrative	9	3.15 (0.34)	6	3.22 (0.40)
Social	7	3.64 (0.35)	7	3.58 (0.36)
None[a]	0	-	4	3.00 (0.00)

[a]None was coded when the teacher was engaged in a behavior unrelated to the classroom such as talking on a cell phone or eating separately from the children.

Table 29: Teacher Verbal Behaviors (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Not Talking or Listening	32	.19 (.09)	28	.21 (.07)
Talking	32	.73 (.09)	28	.70 (.08)
Child	32	.34 (.09)	28	.36 (.07)
Small Group	32	.08 (.04)	28	.06 (.04)
Whole Group	32	.26 (.06)	28	.24 (.09)
Self	32	.01 (.01)	28	.01 (.01)
Parent or External Adult	32	.01 (.02)	28	.005 (.01)
Teacher	32	.02 (.02)	28	.02 (.03)
Listening	32	.08 (.06)	28	.10 (.05)
Child	32	.06 (.05)	28	.07 (.04)
Small Group	32	.003 (.01)	28	.001 (.005)
Whole Group	32	.004 (.01)	28	.004 (.01)
Parent or External Adult	32	.005 (.01)	28	.01 (.02)
Teacher	32	.005 (.01)	28	.01 (.02)
Cohort 2				
Not Talking or Listening	10	.17 (.08)	10	.20 (.09)
Talking	10	.70 (.07)	10	.70 (.12)
Child	10	.38 (.09)	10	.36 (.08)
Small Group	10	.07 (.03)	10	.08 (.05)
Whole Group	10	.21 (.03)	10	.23 (.07)
Self	10	.002 (.01)	10	.01 (.01)
Parent or External Adult	10	.01 (.02)	10	.01 (.01)
Teacher	10	.02 (.02)	10	.01 (.02)
Listening	10	.14 (.08)	10	.10 (.05)
Child	10	.13 (.07)	10	.08 (.04)
Small Group	10	.00 (.00)	10	.00 (.00)
Whole Group	10	.003 (.01)	10	.004 (.01)
Parent or External Adult	10	.002 (.01)	10	.00 (.00)
Teacher	10	.01 (.01)	10	.01 (.03)

Table 30: Average Level of Instruction

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Average Level of Instruction				
Across All Sweeps	32	1.85 (0.21)	28	1.75 (0.18)
Instructional Sweeps Only	32	1.89 (0.14)	28	1.89 (0.12)
Cohort 2				
Average Level of Instruction				
Across All Sweeps	10	1.88 (0.17)	10	1.73 (0.14)
Instructional Sweeps Only	10	1.89 (0.08)	10	1.93 (0.12)

Child Observation in Preschool

The *Child Observation in Preschool* (COP) is a system for observing the behaviors of individual children in preschool classrooms across a daylong visit. Consented as well as non-consented children were observed, the latter anonymously, to yield a picture of how children were functioning in the classroom. The COP is based on a series of snapshots of the behaviors across a period of time when children and teachers are in the room. The COP depicts classroom conditions from the child's perspective, a bottom-up approach, or instruction as received. It quantifies child behaviors along a variety of dimensions. Several of the key variables obtainable from the COP are presented in this section. Results are presented for most variables as proportions of sweeps (or observation snapshots) for which a particular behavior was observed. These proportions should not be interpreted as a proportion of the school day but rather proportions of the observation (COP is not coded during outdoor times, naps and only once during mealtimes.)

The learning foci of the children are presented in Table 31. As one might expect, the divisions are similar to the instructional foci of the teachers (Table 26); these data were collected when children were in independent play (centers) and reflect also the child's attending during teacher led instruction. The bulk of the learning foci observed involved language and literacy activities. Children in Tools classrooms were observed engaging in three times as much dramatic play as children in the comparison classrooms, reflecting an emphasis in the curriculum.

Table 32 presents summaries of the types of social interactions observed. For most of the categories, children in Tools and comparison classrooms did not differ. About half of their interactions were parallel where children have similar materials but were not interacting to create something together. Although the Tools developers had predicted children would less often be observed unoccupied or in time out in Tools classrooms, the percentages are very similar across conditions. Of note is the reduction in time spent in routines (non-academic) for cohort 2 Tools teachers.

Tables 33 and 34 relate to scheduled activities during the day. Table 33 presents the proportion of the observation children were coded in the various scheduled activities and includes all classrooms even those in which a specific schedule type was not observed. Table 34 presents children's level of involvement (rated on a 5 point scale and aggregated across children) when they were observed in that activity. Children in cohort 1 Tools classrooms were generally rated as more involved across different types of scheduled activities, while children in cohort 2 comparison classrooms were rated as more involved. Cohort 2 comparison classrooms were implementing OWL for the first time.

The verbal behaviors of children are presented in Table 35. The rate of child talk (@25%) was very similar across conditions and across cohorts, despite the fact that Tools developers predicted that children in Tools classrooms would engage in more talk. The amount of listening was also similar though children in cohort 1 Tools classrooms listened more often than the comparison children especially to the teacher.

Table 31: Children Learning Focus (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Literacy	32	.11 (.04)	28	.06 (.03)
Code-Based Skills	32	.04 (.03)	28	.06 (.03)
Reading	32	.09 (.03)	28	.07 (.03)
Math	32	.06 (.03)	28	.06 (.03)
Science	32	.04 (.02)	28	.06 (.03)
Social Studies	32	.03 (.02)	28	.03 (.02)
Drama	32	.07 (.04)	28	.02 (.02)
Other[a]	32	.21 (.07)	28	.25 (.08)
None[b]	32	.36 (.06)	28	.37 (.07)
Cohort 2				
Literacy	10	.12 (.02)	10	.06 (.03)
Code-Based Skills	10	.02 (.01)	10	.05 (.04)
Reading	10	.09 (.03)	10	.07 (.03)
Math	10	.08 (.02)	10	.10 (.04)
Science	10	.05 (.03)	10	.06 (.03)
Social Studies	10	.01 (.01)	10	.01 (.01)
Drama	10	.06 (.02)	10	.02 (.01)
Other[a]	10	.29 (.04)	10	.32 (.10)
None[b]	10	.29 (.05)	10	.31 (.06)

[a]Other was coded when the learning focus was not one of the above named categories, most often toys, games, art, or music.

[b]None was coded when a learning opportunity existed but the child was not observed with a learning focus.

Table 32: Children Interaction Type (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Parallel	32	.47 (.05)	28	.48 (.05)
Associative	32	.10 (.04)	28	.07 (.03)
Cooperative	32	.02 (.02)	28	.01 (.01)
Alone	32	.04 (.02)	28	.07 (.03)
Onlooker	32	.01 (.004)	28	.005 (.003)
Social	32	.02 (.01)	28	.02 (.01)
Unoccupied	32	.05 (.02)	28	.04 (.02)
Time Out	32	.002 (.003)	28	.003 (.003)
Non-Academic[a]	32	.29 (.05)	28	.30 (.06)
Cohort 2				
Parallel	10	.50 (.04)	10	.51 (.07)
Associative	10	.10 (.03)	10	.07 (.01)
Cooperative	10	.03 (.01)	10	.02 (.02)
Alone	10	.09 (.02)	10	.09 (.03)
Onlooker	10	.003 (.003)	10	.003 (.003)
Social	10	.02 (.01)	10	.02 (.01)
Unoccupied	10	.04 (.02)	10	.03 (.02)
Time Out	10	.001 (.001)	10	.001 (.001)
Non-Academic[a]	10	.22 (.04)	10	.26 (.05)

[a]Non-Academic was coded during routine activities of the classroom such as hand washing, bathroom visits, mealtimes, and also if children were waiting for an activity to begin.

Table 33: Children Schedule (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Whole Group	32	.39 (.08)	28	.36 (.09)
Small Group	32	.13 (.07)	28	.01 (.02)
Centers	32	.18 (.07)	28	.25 (.12)
Small Group Centers	32	.04 (.07)	28	.11 (.09)
Transition	32	.19 (.05)	28	.21 (.07)
Meal Time	32	.07 (.04)	28	.06 (.04)
Other	32	.01 (.01)	28	.01 (.02)
Cohort 2				
Whole Group	10	.27 (.05)	10	.27 (.06)
Small Group	10	.13 (.03)	10	.02 (.02)
Centers	10	.37 (.08)	10	.43 (.10)
Small Group Centers	10	.06 (.06)	10	.06 (.05)
Transition	10	.15 (.04)	10	.20 (.06)
Meal Time	10	.01 (.02)	10	.02 (.02)
Other	10	.001 (.002)	10	.01 (.01)

Table 34: Average Involvement by Schedule

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Whole Group	32	2.69 (0.18)	28	2.57 (0.20)
Small Group	32	2.62 (0.29)	20	2.38 (0.98)
Centers	32	2.92 (0.17)	28	2.84 (0.17)
Small Group Centers	25	2.82 (0.63)	25	2.82 (0.44)
Transition	32	1.32 (0.13)	28	1.32 (0.18)
Meal Time	30	1.13 (0.23)	23	1.10 (0.15)
Other	8	2.15 (0.92)	14	1.59 (0.63)
Cohort 2				
Whole Group	10	2.77 (0.20)	10	2.84 (0.15)
Small Group	10	2.70 (0.28)	8	2.82 (0.18)
Centers	10	2.79 (0.11)	10	2.91 (0.18)
Small Group Centers	8	2.52 (0.65)	9	2.89 (0.25)
Transition	10	1.58 (0.29)	10	1.36 (0.16)
Meal Time	5	1.48 (0.85)	8	1.36 (0.33)
Other	3	1.00 (0.00)	5	1.63 (0.91)

Table 35: Children Talking and Listening (Proportion of Sweeps)

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Not Talking or Listening	32	.39 (.05)	28	.43 (.04)
Talking	32	.25 (.04)	28	.25 (.05)
Teacher	32	.04 (.01)	28	.05 (.01)
Child	32	.09 (.02)	28	.09 (.02)
Small Group	32	.01 (.01)	28	.01 (.01)
Whole Group	32	.04 (.03)	28	.05 (.02)
Self	32	.06 (.02)	28	.06 (.02)
Fuss/Cry	32	.01 (.005)	28	.01 (.003)
Listening	32	.36 (.07)	28	.31 (.06)
Teacher	32	.28 (.06)	28	.24 (.06)
Child	32	.08 (.02)	28	.07 (.02)
Small Group	32	.001 (.001)	28	.002 (.002)
Whole Group	32	.004 (.003)	28	.005 (.004)
Cohort 2				
Not Talking or Listening	10	.38 (.04)	10	.37 (.02)
Talking	10	.27 (.04)	10	.27 (.04)
Teacher	10	.06 (.01)	10	.05 (.01)
Child	10	.10 (.03)	10	.10 (.02)
Small Group	10	.02 (.01)	10	.01 (.01)
Whole Group	10	.03 (.02)	10	.04 (.03)
Self	10	.07 (.02)	10	.08 (.02)
Fuss/Cry	10	.005 (.01)	10	.004 (.002)
Listening	10	.34 (.04)	10	.36 (.04)
Teacher	10	.25 (.04)	10	.26 (.05)
Child	10	.08 (.02)	10	.09 (.02)
Small Group	10	.004 (.004)	10	.002 (.002)
Whole Group	10	.01 (.01)	10	.01 (.01)

Condition Comparisons

The Narrative Record, TOP and COP were chosen to provide details about the interactions within the classrooms that the Tools developers believed would distinguish classrooms implementing their curriculum from other early childhood classrooms. In October 2009, members of the research team met with Tools developers and trainers in a 4 day meeting to plan both the Fidelity of Implementation system (what we call vertical fidelity) and a plan to assess horizontal fidelity – how similar or different would the interactions be in the classroom when the Tools curriculum is being implemented. During this meeting, researchers also visited classrooms in Massachusetts run by experienced Tools teachers. From this meeting we developed a series of hypotheses about what would be different in Tools classrooms. All of these hypotheses involve interactions and could be tested with data generated by TOP and COP. The prior descriptive reports on data from COP and TOP present the data as simple probabilities. Most of the hypotheses the developers generated involve conditional probabilities (e.g., teachers will be more positive during instruction).

In this section of the report, we will present the hypotheses followed by tables of the means and standard deviations for the behaviors in question presented for each cohort separately. Analyses were conducted to determine the effect of treatment condition on each variable of interest. Multilevel models were conducted in SPSS with classrooms nested within schools and systems and were run separately for each variable. Analyses were also conducted separately for each cohort.

Hypothesis 1: The pattern of talk for children will be different in Tools classrooms from Comparison classrooms (COP).

1. There will be more instances of Child-to-Child talk in Tools classrooms.
2. Children who are talking to each other will be more likely to have a learning focus (all content areas) in Tools classrooms compared to Control classrooms.
3. Children in Tools classrooms will more often talk to themselves.
4. Children will more often be observed listening to other children in Tools classrooms.

As can be seen in Table 36, one of the hypotheses was confirmed. Children in Tools classrooms were more likely to be observed talking to another child about specific content. But children in Tools classrooms were not observed talking or listening any more frequently to other children nor were they more likely to talk to themselves.

Table 36: Patterns of Child Talk in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Talking To Child	32	.09 (.02)	28	.09 (.02)
Talk to Child with Content Focus	32	.04 (.02)	28	.02 (.01)
Talking To Self	32	.06 (.02)	28	.06 (.02)
Listening to Child	32	.08 (.02)	28	.07 (.02)
Cohort 2				
Talking To Child	10	.10 (.03)	10	.10 (.02)
Talk to Child with Content Focus	10	.04 (.01)	10	.02 (.01)*
Talking To Self	10	.07 (.02)	10	.08 (.02)
Listening to Child	10	.08 (.02)	10	.09 (.02)

* $p < .05$.

Hypothesis 2: The pattern of talk for teachers will be different in Tools classrooms from Control classrooms (TOP).

1. Teachers will talk less in Tools classrooms than in Control classrooms.
2. There will be a better balance between teacher and child talk in Tools classrooms than in Control classrooms.
3. Teachers will engage in less Management activities and a lower proportion of their talk will be during Management in Tools classrooms compared to Control classrooms.
4. Teachers will talk more with children during Center time in Tools classrooms than in Control classrooms.

Hypothesis 3 was partially confirmed; cohort 2 Tools teachers spent less time managing the classroom, but no other differences in the pattern of teacher talk were found between Tools and comparison classrooms

Table 37: Patterns of Teacher Talk in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Teacher Talking	32	.73 (.09)	28	.70 (.08)
Child Talking	32	.25 (.04)	28	.25 (.05)
Task Management	32	.22 (.08)	28	.23 (.08)
Talk during Task Management	32	.20 (.08)	28	.20 (.10)
Talk to Child during Task Management	32	.10 (.07)	28	.12 (.08)
Talk to Children during Centers	32	.58 (.21)	28	.55 (.18)
Instruction during Centers	32	.28 (.21)	28	.26 (.19)
Management during Centers	32	.23 (.19)	28	.25 (.18)
Cohort 2				
Teacher Talking	10	.70 (.07)	10	.70 (.12)
Child Talking	10	.27 (.04)	10	.27 (.04)
Task Management	10	.27(.10)	10	.37 (.07)*
Talk during Task Management	10	.27 (.08)	10	.36 (.09)
Talk to Child during Task Management	10	.20 (.09)	10	.23 (.10)
Talk to Children during Centers	10	.56 (.12)	10	.60 (.20)
Instruction during Centers	10	.31 (.15)	10	.22 (.14)
Management during Centers	10	.34 (.17)	10	.45 (.12)

* $p < .05$.

.

Hypothesis 3: Children's interactions will differ in Tools classrooms compared to Control classrooms (COP).

1. Children will more often be engaged in associative interactions in Tools classrooms than in Control classrooms.
2. Children will talk more during associative interactions in Tools classrooms than in Control classrooms.
3. Children will less often engage in parallel interactions in Tools classrooms compared to Control classrooms.

Data are presented in Table 38. Hypothesis 1 was confirmed for both cohorts; children in Tools classrooms spent more time in associative interactions. None of the other hypotheses was supported. Children did not talk differently during associative interactions in Tools classrooms nor were they less often observed in parallel interactions.

Table 38: Patterns of Children's Interactions in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Associative Interactions	32	.10 (.04)	28	.07 (.03)*
Talk during Associative Interaction	32	.49 (.08)	28	.53 (.11)
Talk to Child during Associative	32	.67 (.11)	28	.67 (.12)
Talk to Teacher during Associative	32	.21 (.11)	28	.21 (.11)
Talk to Small Group during Associative	32	.06 (.05)	28	.07 (.06)
Parallel Interactions	32	.47 (.05)	28	.48 (.05)
Cohort 2				
Associative Interactions	10	.10 (.03)	10	.07 (.01)*
Talk during Associative Interaction	10	.49 (.05)	10	.49 (.05)
Talk to Child during Associative	10	.59 (.11)	10	.59 (.14)
Talk to Teacher during Associative	10	.27 (.08)	10	.21 (.14)
Talk to Small Group during Associative	10	.05 (.05)	10	.07 (.07)
Parallel Interactions	10	.50 (.04)	10	.51 (.07)

* $p < .05$.

Hypothesis 4: Children's learning behaviors will differ in Tools classrooms compared to Control classrooms (COP).

1. Children will be rated as more highly involved in Tools classrooms compared to Control classrooms.
    a. When the learning focus is Drama.
    b. During Center time.
    c. During transitions.
2. Children will be less often seen as Disruptive in Tools classrooms than in Control classrooms.
3. Children in Tools classrooms will be less often observed to be Unoccupied than in Control classrooms.

As shown in Table 39, Tools children were more involved during Center time but only for cohort 1. No other of the proposed differences in learning behaviors was found.

Table 39: Children's Learning Behaviors in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Mean Involvement (All Sweeps)	32	2.38 (0.17)	28	2.33 (0.20)
Mean Involvement during Drama	32	3.28 (0.16)	28	3.38 (0.21)
Mean Involvement during Centers	32	2.92 (0.17)	28	2.84 (0.17)*
Mean Involvement during Transitions	32	1.32 (0.13)	28	1.32 (0.18)
Unoccupied	32	.05 (.02)	28	.04 (.02)
Disruptive	32	.01 (.01)	28	.01 (.01)
Cohort 2				
Mean Involvement (All Sweeps)	10	2.55 (0.14)	10	2.53 (0.17)
Mean Involvement during Drama	10	3.36 (0.16)	10	3.44 (0.18)
Mean Involvement during Centers	10	2.79 (0.11)	10	2.91 (0.18)
Mean Involvement during Transitions	10	1.58 (0.29)	10	1.36 (0.16)
Unoccupied	10	.04 (.02)	10	.03 (.02)
Disruptive	10	.01 (.004)	10	.003 (.003)

* $p < .05$.

Hypothesis 5: Teacher behaviors will be different in Tools classrooms compared to Control classrooms (TOP).

1. Teachers in Tools classrooms will have a warmer emotional tone compared to Control classrooms.
2. Teachers in Tools classrooms will engage in more Behavior Approving than in Control classrooms.
3. Teachers in Tools classrooms will engage in less Behavior Disapproving than in Control classrooms.
4. Teachers in Tools classrooms will engage in more instruction than in Control classrooms, and will have a higher level of instruction.
5. Teachers will prompt children to use private speech more often in Tools classrooms than in Control classrooms.
6. Teachers will prompt children to work in pairs more in Tools classrooms than in Control classrooms.
7. Teachers will make intentional mistakes for children to recognize more in Tools classrooms than in Control classrooms.

Summaries of the teacher behaviors are presented in Table 40. As can be seen, Tools teachers differed from teachers in comparison classrooms in their encouragement of private speech and in their use of paired activities. They did not differ in their emotional tone or in their use of behavior approving and disapproving.

Table 40: Teacher Behaviors in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Emotional Tone (rating 1-5 scale)	32	3.43 (.16)	28	3.38 (.20)
Behavior Approving	32	.03 (.03)	28	.05 (.03)
Behavior Disapproving	32	.06 (.05)	28	.06 (.06)
Instruction	32	.45 (.10)	28	.40 (.11)
Mean Level of Instruction (rating 1-4)	32	1.89 (.14)	28	1.89 (.12)
Teacher Encouraged Private Speech	32	.04 (.04)	28	.002 (.004)*
Teacher Paired Activities	32	.06 (.03)	28	.01 (.01)*
Teacher Intentional Mistakes	32	.01 (.01)	28	.01 (.01)
Cohort 2				
Emotional Tone (rating 1-5 scale)	10	3.49 (.20)	10	3.44 (.17)
Behavior Approving	10	.05 (.03)	10	.05 (.04)
Behavior Disapproving	10	.04 (.03)	10	.02 (.03)
Instruction	10	.47 (.08)	10	.38 (.07)
Mean Level of Instruction (rating 1-4)	10	1.89 (.08)	10	1.93 (.12)
Teacher Encouraged Private Speech	10	.04 (.04)	10	.003 (.01)*
Teacher Paired Activities	10	.08 (.03)	10	.004 (.01)*
Teacher Intentional Mistakes	10	.01 (.01)	10	.01 (.01)

* $p < .05$.

Hypothesis 6: The classroom day will be organized differently in Tools classrooms compared to Control classrooms (Narrative Record summaries)

1. There will be less intentional teaching in Tools classrooms compared to Control classrooms (i.e., teacher led instruction).
2. More time in Tools classrooms will be spent in Centers.
3. Less time in Tools classrooms will be spent in large group instruction. During large group instruction, children will be more involved in Tools classrooms compared to Control classrooms.
4. Less time in Tools classrooms will be spent in transitions.

As presented in Table 41, for both cohorts, the first hypothesis was confirmed in *reverse* – Tools teachers did more intentional teaching in both cohorts than comparison teachers. In cohort 1, the opposite effect for hypothesis 2 was also obtained – less time was spent in center based activities in Tools classrooms than comparison classrooms. No difference in the amount of center time was found for cohort 2. No differences were found between Tools and comparison classrooms in the amounts of time spent in whole group instruction or transitions. During whole group in cohort 1 only, children were rated as more involved.

Table 41: Organization of Classroom Day in Tools and Comparison Classrooms

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Proportion of Time in Teacher-led Instruction	32	.26 (.07)	28	.17 (.07)*
Proportion of Time in Centers	32	.12 (.04)	28	.16 (.06)*
Proportion of Time in Whole Group	32	.18 (.06)	28	.17 (.07)
Proportion of Time in Transition	32	.13 (.04)	28	.13 (.04)
Mean Involvement during Whole Group	32	2.69 (0.18)	28	2.57 (0.20)*
Cohort 2				
Proportion of Time in Teacher-led Instruction	10	.18 (.02)	10	.13 (.04)*
Proportion of Time in Centers	10	.23 (.03)	10	.21 (.05)
Proportion of Time in Whole Group	10	.12 (.02)	10	.12 (.04)
Proportion of Time in Transition	10	.11 (.02)	10	.13 (.05)
Mean Involvement during Whole Group	10	2.77 (0.20)	10	2.84 (0.15)

* $p < .05$.

Fidelity of Implementation

When this project began, no complete Fidelity of Implementation system existed for evaluating Tools classrooms. One developed some years before was partial and based on a former version of the curriculum. In partnership with the Tools developers, the PRI research staff created a complete Fidelity of Implementation system. The system was used in all classrooms though almost none of the comparison classrooms implemented any activities at all similar to Tools activities. Observers received extensive training on the curriculum, participating in the curriculum training sessions provided by the Tools trainers. The observers had to be able to recognize a Tools activity immediately.

The Narrative Record is the platform on which the Fidelity system was built. When a Tools activity began, observers chose the general category of activity (e.g., Tools Large Group; Tools Literacy block, etc.) and then were able to tab to the specific activity. On the specific activity page were listed each of the required steps to enact the activity, the mediators recommended as well as a list of "should nots" that the Tools developers had provided.

Table 42 presents the percentage of the day devoted to each of the major Tools Activity Blocks. The two cohorts enacted Tools for similar amounts of time for each activity block except for Make Believe Play. As mentioned, the Tools developers strongly emphasized enacting Make Believe Play in the second cohort. Classrooms in cohort 2 also spent twice as much time in Free Choice Centers as the cohort 1 classrooms. In part this was due to their spending less time in non Tools activities – the routines of the classroom (Table 32 showed a meaningful reduction in non-academic time.). But it is apparent that in a 6 ½ hour day (390 minutes), teachers have less than half of that time to devote to curriculum implementation. In cohort 1, 136 minutes of the day could be spent implementing the curriculum.

When this study was conducted, the curriculum contained 63 separate activities, not all of which were to be implemented at one time, some were prescribed for different points during the year, some were to be done twice a week, some every day. Each activity had a series of 9 to 13 steps. Table 43 summarizes the number of Tools activities observed at each observation, the total number of steps carried out, the number of mediators observed across the activities and the number of should-nots observed. Cohort 2 teachers carried out an average of 1 more Tools activity at each time point than cohort 1. Otherwise their implementation levels appeared to be the same.

We created two analytic variables to investigate the effects of curriculum implementation on immediate and longer-term outcomes. One variable was the amount of Tools instruction, constructed from the Tools time blocks (presented in Table 42). The other is a weighted fidelity score created by giving more weight to activities that required more effort and planning to implement and less weight to simpler activities. The weighted activities and their steps were combined to yield a single score. Tables 44 through 49 present the results of analyses using these two scores as predictors. Results for each cohort are presented separately. For each, the first set of results involve the achievement measures, the second the self-regulation assessments and the third

the teacher and assessor ratings. It will be evident from reviewing the tables that neither fidelity measure was systematically related to any outcomes for either cohort. More information can be downloaded online:
https://my.vanderbilt.edu/toolsofthemindevaluation/resources/fidelitymeasures/

Table 42: Proportion of Time in Tools Time Block Activities (across the entire day)

	Cohort 1		Cohort 2	
	n	*M (SD)*	*n*	*M (SD)*
Activity Blocks				
Large Group	32	0.05 (0.02)	10	0.04 (0.01)
Literacy	32	0.05 (0.02)	10	0.04 (0.01)
Math/Science	32	0.01 (0.02)	10	0.01 (0.01)
Activities through the Day	32	0.02 (0.01)	10	0.01 (0.01)
Make Believe Play (MBP)	32	0.05 (0.02)	10	0.08 (0.05)
MBP Planning	32	0.03 (0.01)	10	0.03 (0.01)
MBP Practice	32	0.01 (0.01)	10	0.003 (0.004)
MBP Cleanup	32	0.01 (0.01)	10	0.011 (0.004)
Pretend Transitions	32	0.003 (0.005)	10	0.002 (0.002)
Free Choice Centers	32	0.07 (0.04)	10	0.15 (0.06)
Mixed Tools Activities[a]	32	0.03 (0.03)	10	0.04 (0.02)
Tools and Non-Tools Activities[b]	32	0.02 (0.03)	10	0.04 (0.03)
Non-Tools Activities[c]	32	0.65 (0.07)	10	0.56 (0.03)

[a]Teachers combined two or more Tools activities into the same instructional episode.

[b]Teachers combined a Tools activity with one of their own.

[c]The classroom was engaged in a non Tools activity. Such activities include outside time, meals, and naps as well as activities that were special to the teacher but unrelated to the curriculum.

Table 43: Descriptive Statistics for *Tools of the Mind* Fidelity of Implementation by Observation

	Observation 1			Observation 2			Observation 3		
	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Cohort 1									
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)
Cohort 2									
Activities (#)	14	18	15.70 (1.34)	11	18	15.40 (2.07)	12	17	15.30 (1.83)
Steps (#)	55	81	66.10 (8.63)	44	65	56.80 (7.54)	44	79	60.10 (9.63)
Mediators (#)	28	46	35.50 (4.62)	24	38	33.60 (4.09)	26	45	35.70 (5.77)
Should Nots (#)	0	10	3.40 (3.17)	0	10	4.70 (3.09)	0	9	4.90 (2.42)

Note. Cohort 1 = 32 and Cohort 2 = 10 *Tools of the Mind* Classrooms. Values reported represent all *Tools of the Mind* activities, steps mediators, and should nots observed irrespective of whether or not the feature was indicated as appropriate to implement by the curriculum at the given observation point in the school year.

Table 44: Cohort 1 Impact of Fidelity of Implementation on Academic Achievement at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
Letter Word	10.76 (19.73)	0.28 (1.05)	-0.85 (16.42)	1.02 (0.78)	-4.03 (13.15)	-0.24 (0.87)
Spelling	7.45 (21.78)	0.18 (1.10)	16.83 (18.94)	0.82 (0.95)	-24.82 (16.30)	-1.24 (0.88)
Academic Knowledge	-14.51 (8.57)	-0.58 (0.44)	-3.46 (8.86)	-0.14 (0.45)	-5.58 (8.28)	-0.41 (0.43)
Oral Comprehension	0.10 (7.53)	-0.67 (0.36) [†]	-11.26 (10.22)	-0.74 (0.50)	-7.43 (7.32)	-0.54 (0.40)
Picture Vocabulary	-1.92 (5.70)	-0.04 (0.29)	5.58 (5.91)	0.29 (0.29)	3.34 (6.59)	0.12 (0.32)
Applied Problems	19.45 (11.08)[†]	0.33 (0.58)	-1.06 (9.26)	-0.23 (0.46)	-11.38 (10.99)	-0.62 (0.56)
Quantitative Concepts	4.68 (9.76)	0.11 (0.50)	-2.73 (8.14)	-0.03 (0.41)	-2.40 (7.69)	-0.35 (0.39)
Composite Achievement	0.40 (2.42)	-0.04 (0.13)	-0.22 (3.33)	0.02 (0.17)	-2.85 (2.71)	-0.19 (0.15)
Passage Comprehension	--	--	--	--	-19.71 (12.90)	-1.04 (0.67)

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest.
$^{**}p < .01$. $^{*}p < .05$. $^{\dagger}p < .10$.

Table 45: Cohort 1 Impact of Fidelity of Implementation on Self-Regulation at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
Forward Span	0.17 (0.74)	0.01 (0.04)	-0.02 (0.86)	-0.04 (0.04)	-0.68 (0.94)	-0.04 (0.05)
Backward Span	0.26 (0.86)	0.04 (0.04)	-0.55 (1.04)	0.02 (0.05)	-0.40 (1.06)	-0.08 (0.05)
DCCS	0.11 (0.41)	0.02 (0.02)	-1.05 (0.50)*	-0.04 (0.03)	-0.19 (0.68)	0.01 (0.03)
Copy Design	4.50 (2.24)ŧ	0.09 (0.12)	1.34 (2.10)	0.06 (0.10)	2.42 (1.93)	0.19 (0.10)ŧ
HTKS	-5.13 (11.22)	-0.22 (0.57)	4.58 (9.43)	0.14 (0.47)	-10.11 (6.23)	-0.63 (0.33)ŧ
Peg Tapping	2.67 (3.47)	-0.16 (0.18)	1.43 (2.51)	0.08 (0.12)	-2.88 (2.00)	-0.10 (0.10)
Composite Self-Regulation	0.52 (2.23)	0.01 (0.11)	-0.29 (2.01)	-0.06 (0.10)	-3.02 (2.19)	-0.21 (0.11)ŧ

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. DCCS = Dimensional Change Card Sort. HTKS = Head Toes Knees Shoulders.

**$p < .01$. *$p < .05$. ŧ$p < .10$.

Table 46: Cohort 1 Impact of Fidelity of Implementation on Teacher and Assessor Ratings at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
CFBRS Interpersonal Skills Scale	0.24 (0.93)	0.06 (0.05)	0.51 (0.88)	0.02 (0.05)	-1.35 (0.79)[†]	-0.03 (0.04)
CFBRS Work-Related Skills Scale	-0.05 (1.18)	0.02 (0.06)	0.79 (0.88)	0.04 (0.04)	-0.78 (0.91)	0.01 (0.05)
Adaptive Language Inventory	-0.20 (0.77)	0.02 (0.04)	0.61 (0.63)	0.04 (0.03)	0.09 (0.69)	0.04 (0.04)
Self-Regulation Assessor Ratings	-0.43 (0.48)	-0.03 (0.03)	0.39 (0.58)	0.04 (0.03)	-0.07 (0.37)	0.01 (0.02)

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. CFBRS=Cooper-Farran Behavior Rating Scales.

$^{**}p < .01$. $^{*}p < .05$. $^{†}p < .10$.

Table 47: Cohort 2 Impact of Fidelity of Implementation on Academic Achievement at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
Letter Word	5.37 (66.64)	-5.35 (5.56)	-44.04 (41.36)	1.30 (4.29)	-38.44 (49.42)	1.69 (5.00)
Spelling	20.91 (51.98)	2.35 (4.44)	-44.18 (34.98)	-0.35 (3.93)	-1.45 (42.50)	-4.05 (4.36)
Academic Knowledge	-21.84 (23.60)	3.84 (1.98)[†]	17.19 (23.62)	2.47 (2.25)	-0.78 (24.16)	2.89 (2.16)
Oral Comprehension	-1.41 (27.22)	-0.66 (2.30)	-21.72 (19.39)	3.14 (2.02)	-0.03 (23.60)	-0.51 (2.37)
Picture Vocabulary	4.29 (14.83)	1.04 (1.26)	3.00 (13.45)	-0.41 (1.40)	3.61 (16.19)	-0.72 (1.65)
Applied Problems	5.18 (32.80)	2.03 (2.83)	-50.19 (28.87)[†]	5.04 (2.95)[†]	-18.10 (29.67)	1.81 (3.04)
Quantitative Concepts	-37.18 (30.10)	1.62 (2.85)	-11.95 (20.54)	1.83 (2.13)	-14.36 (22.73)	2.28 (2.31)
Composite Achievement	-2.09 (6.21)	0.29 (0.57)	-10.31 (6.54)	0.75 (0.68)	-3.35 (7.76)	0.27 (0.79)
Passage Comprehension	--	--	--	--	-50.91 (38.18)	0.49 (3.68)

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest.
$^{**}p < .01$. $^{*}p < .05$. $^{\dagger}p < .10$.

Table 48: Cohort 2 Impact of Fidelity of Implementation on Self-Regulation at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
Forward Span	1.66 (2.80)	-0.03 (0.24)	1.24 (2.03)	-0.31 (0.21)	-1.38 (2.33)	0.05 (0.24)
Backward Span	0.31 (2.70)	0.27 (0.23)	-0.59 (2.59)	0.15 (0.26)	1.10 (2.59)	0.01 (0.26)
DCCS	-2.17 (1.43)	0.26 (0.11)*	1.46 (1.08)	-0.01 (0.12)	0.19 (2.61)	-0.16 (0.25)
Copy Design	1.48 (4.59)	0.72 (0.39)[ł]	2.47 (4.84)	-0.01 (0.52)	-1.07 (5.23)	-0.31 (0.53)
HTKS	2.30 (29.39)	-0.03 (2.50)	13.02 (22.30)	-2.65 (2.32)	-9.35 (17.68)	-0.31 (1.80)
Peg Tapping	-9.69 (10.83)	-1.00 (0.92)	-7.49 (6.60)	0.06 (0.66)	1.09 (4.18)	-0.68 (0.42)
Composite Self-Regulation	0.50 (6.02)	0.55 (0.52)	3.47 (6.83)	-0.56 (0.71)	0.39 (6.72)	-0.63 (0.68)

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. DCCS = Dimensional Change Card Sort. HTKS = Head Toes Knees Shoulders.
**$p < .01$. *$p < .05$. [ł]$p < .10$.

Table 49: Cohort 2 Impact of Fidelity of Implementation on Teacher and Assessor Ratings at the End of Pre-K, Kindergarten, and 1st Grade

	End of Pre-K		End of Kindergarten		End of First Grade	
Variable	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]	Amount of Tools Instruction	Weighted Fidelity Score	Amount of Tools Instruction	Weighted Fidelity Score
CFBRS Interpersonal Skills Scale	-0.24 (2.46)	0.21 (0.19)	-5.03 (2.86)	-0.21 (0.30)	-1.89 (2.88)	-0.22 (0.27)
CFBRS Work-Related Skills Scale	0.68 (3.06)	0.25 (0.25)	0.24 (2.49)	-0.19 (0.24)	2.02 (3.00)	-0.30 (0.29)
Adaptive Language Inventory	-0.20 (2.46)	0.13 (0.22)	0.18 (1.09)	0.08 (0.11)	-0.12 (1.08)	0.08 (0.11)
Self-Regulation Assessor Ratings	1.09 (1.34)	0.06 (0.13)	0.73 (0.84)	0.06 (0.09)	0.48 (0.42)	-0.03 (0.04)

Note. [1]Proportion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument rescaled to dividing original estimate by 100. Coefficients in the table are unstandardized regression coefficients (standard errors) from multi-level regression models that account for nesting of teachers in schools and random assignment blocks. Covariates included in the models were pretest, gender, ELL and IEP status, age at pretest, and interval from pretest. CFBRS=Cooper-Farran Behavior Rating Scales.

$^{**}p < .01$. $^{*}p < .05$. $^{\dagger}p < .10$.

Post Observation Rating Scale

Both observers rated the 36 item PRS immediately after being in the classroom a full day (both Tools and comparison classrooms). As each observer had seen the classroom from a different perspective, they reached consensus on what the final ratings should be. Presented in Table 50 are the average ratings by subscale averaged across three observations for each cohort. In cohort 1 particularly the observers had a much more positive view of the classrooms as demonstrated by their higher PRS ratings. The ratings are more similar for Tools and comparison classrooms for cohort 2, whose ratings in general tend to be higher than those in cohort 1.

Table 50: Average Ratings on Post Observation Rating Scale Subscales and Overall

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
General	32	3.71 (0.36)	28	3.44 (0.37)
Center Time	32	2.76 (0.39)	28	2.31 (0.33)
Classroom Management	32	3.92 (0.44)	28	3.59 (0.60)
Teacher Responsiveness	32	3.41 (0.51)	28	3.06 (0.53)
Community	32	3.72 (0.46)	28	3.54 (0.40)
Academic-Learning Related	32	3.16 (0.50)	28	2.86 (0.57)
Overall Classroom Atmosphere	32	3.45 (0.38)	28	3.13 (0.37)
Cohort 2				
General	10	3.94 (0.31)	10	3.80 (0.26)
Center Time	10	2.78 (0.23)	10	2.39 (0.28)
Classroom Management	10	3.96 (0.40)	10	3.94 (0.36)
Teacher Responsiveness	10	3.56 (0.53)	10	3.53 (0.46)
Community	10	3.81 (0.42)	10	3.83 (0.23)
Academic-Learning Related	10	3.16 (0.48)	10	3.18 (0.29)
Overall Classroom Atmosphere	10	3.53 (0.33)	10	3.45 (0.23)

Note. The full Post Observation Rating Scale can be downloaded online (https://my.vanderbilt.edu/toolsofthemindevaluation/resources/classroomobservationmeasures/)

Environmental Checklist

The Environmental Checklist was created at the request of the Tools developers. They believed that classrooms implementing Tools would look different from other early childhood classrooms. The following tables present descriptive data on the number of centers present in each classroom for each of the three observations and a comparison of classrooms on the materials present, using a list provided by the Tools developers.

The classrooms differed in the number of centers, especially in the presence of Make Believe Play centers. Tools classrooms in cohort 2 had far more free play centers. Both cohorts and types of classrooms had proportionally the same early childhood materials.

Table 51: Cohort 1 Average Number of Centers by Observation

	Observation 1		Observation 2		Observation 3	
	M (SD)	Range	*M (SD)*	Range	*M (SD)*	Range
Tools Condition (*n* =32)						
Number of Free Play Centers	5.41 (2.56)	0-10	5.75 (2.50)	0-11	3.91 (2.94)	0-9
Number of Make Believe Play Centers	3.19 (2.40)	0-7	3.38 (2.20)	0-6	4.06 (2.08)	0-7
Comparison Condition (*n* =28)						
Number of Free Play Centers	8.36 (1.66)	5-12	8.11 (2.33)	0-11	8.39 (2.08)	5-12
Number of Make Believe Play Centers	0.04 (0.19)	0-1	0.11 (0.42)	0-2	0.04 (0.19)	0-1

Table 52: Cohort 2 Average Number of Centers by Observation

	Observation 1		Observation 2		Observation 3	
	M (SD)	Range	*M (SD)*	Range	*M (SD)*	Range
Tools Condition (*n* =10)						
Number of Free Play Centers	9.00 (1.63)	6-11	9.80 (1.23)	8-11	9.00 (2.40)	5-12
Number of Make Believe Play Centers	4.40 (1.71)	0-6	4.60 (0.52)	4-5	4.30 (1.57)	0-5
Comparison Condition (*n* =10)						
Number of Free Play Centers	8.80 (1.87)	5-12	10.00 (0.94)	9-11	9.40 (1.26)	7-11
Number of Make Believe Play Centers	0.00 (0.00)	0	0.00 (0.00)	0	0.00 (0.00)	0

Table 53: Proportion of Materials Present in Classrooms by Category

	Tools Condition		Comparison Condition	
	n	*M (SD)*	*n*	*M (SD)*
Cohort 1				
Arts and Music Materials (*n* = 12)	32	0.68 (0.17)	28	0.77 (0.16)
Furniture and Equipment (*n* = 3)	32	0.88 (0.13)	28	0.90 (0.13)
Manipulatives (*n* = 3)	32	0.60 (0.24)	28	0.67 (0.27)
Math and Science Materials (*n* = 5)	32	0.43 (0.22)	28	0.53 (0.24)
Make Believe Play Materials (*n* = 17)	32	0.57 (0.18)	28	0.58 (0.17)
Tools Materials (*n* = 12)	32	0.54 (0.16)	28	0.18 (0.11)
Cohort 2				
Arts and Music Materials (*n* = 12)	10	0.91 (0.06)	10	0.89 (0.09)
Furniture and Equipment (*n* = 3)	10	0.98 (0.05)	10	0.97 (0.07)
Manipulatives (*n* = 3)	10	0.74 (0.14)	10	0.76 (0.13)
Math and Science Materials (*n* = 5)	10	0.67 (0.10)	10	0.65 (0.16)
Make Believe Play Materials (*n* = 17)	10	0.72 (0.08)	10	0.76 (0.10)
Tools Materials (*n* = 12)	10	0.66 (0.15)	10	0.17 (0.10)

Note. The full Environmental Checklist can be downloaded online (https://my.vanderbilt.edu/toolsofthemindevaluation/resources/fidelitymeasures/)

Summary and Future Directions

Overall, we found no significant effects of the *Tools of the Mind* curriculum on literacy, language or mathematics achievement when compared to business as usual classrooms whose teachers used a variety of curricular approaches. Similarly, we found no effects on Self-Regulation. Gains in achievement and self-regulation were correlated, $r = .35$.

Following a report of these outcomes for cohort 1, the Tools developers and trainers focused their attention on cohort 2 emphasizing areas they felt had not been implemented sufficiently in cohort 1. However, despite that increased intervention, the outcomes for cohort 2 were the same, no difference in achievement, self regulation or teacher ratings at the end of the pre-k year.

The few significant interactions obtained in the analyses did not provide a consistent picture of the curriculum being more or less effective for subgroups of children.

Follow up assessments in kindergarten and 1st grade presented a mixture of results. One positive effect was found for cohort 2 children who had been in Tools classrooms; they scored higher on picture vocabulary. Significant negative effects were found on various measures for the Tools condition in cohort 1, the larger sample composed of children from a number of school systems.

According to observations of curriculum fidelity, there was variation among the teachers in the degree to which they implemented the curriculum. Virtually all of the *Tools* teachers implemented substantial portions of the curriculum. Mostly teachers implemented the activities at the appropriate times and chose a variety of easy, medium and difficult types of activities. Ambiguity about what constitutes full implementation makes it difficult to appraise accurately the level of implementation actually attained. Teachers in cohort 2 implemented more of the activities involved with make believe play than teachers in cohort 1 following an increased emphasis on this aspect of the curriculum.

Observational measures of fidelity were consistent with ratings of high implementation provided by *Tools* trainers, coaches, project classroom observers, and the teachers themselves lending validity to the fidelity of implementation system we developed.

Variations in fidelity of implementation measures across the full group of 42 *Tools* teachers were not associated with greater gains in achievement or self-regulation. Although teachers in Tools classrooms enacted many activities connected to the curriculum not observed in comparison classrooms, the curriculum itself was not associated with differential gains.

However, the more general classroom interactive behaviors between teachers and children were not changed by implementing the Tools curriculum. The Tools developers hypothesized that Tools classrooms would be distinct from other early childhood classrooms on a number of measures, including such important aspects as teacher and child talk and child involvement levels. While the classrooms differed on a few things, notably organizational patters dictated by the curriculum itself

(e.g., teachers urging private speech and pairing children for activities), most of their behaviors were similar. Across both conditions, teachers varied in all of these behaviors. Further analyses have demonstrated that variation in a number of these key classroom interactions are associated with gains in self regulation as well as achievement.

Across all classrooms, time on task and time spent in centers were related to achievement outcomes while the frequency of children's private speech was negatively related to both achievement and self-regulation outcomes.

Across all classrooms, frequency of behavior disapproving was negatively related to achievement and self-regulation outcomes, while the frequency of behavior approving was positively related to self-regulation outcomes.

Given the positive findings in favor of the comparison classrooms, it is important to examine 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. However, in the *Tools* version of pretend play 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, were 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.

In terms of future directions, the field of early childhood education may need to consider a new approach; reliance on a curriculum to affect child outcomes may be less important than changing interaction patterns in the classroom 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. Classroom procedures that find a way to actively engage young children in the learning process, particularly with materials that lend themselves to exploration, may be more effective. 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.