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

Fidelity of Implementation Technical Report

Deanna Meador, MA
Kimberly Nesbitt, PhD
Dale Farran, PhD

9/1/15

PEABODY research institute

VANDERBILT PEABODY COLLEGE

Peabody Research Institute

Vanderbilt University

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

Recommended Citation: 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 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 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

Table of Contents

PRI Project Staff

Principal Investigators

Dale C. Farran, PhD
Mark W. Lipsey, PhD
Sandra 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

Post-Doctoral Fellow

Mary Wagner Fuhs, PhD

Asha Spivak

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

Introduction

The Experimental Evaluation of the *Tools of the Mind* Pre-K Curriculum study was designed to compare the effectiveness of the *Tools of the Mind* (*Tools*) curriculum to the curricula the school system is currently using in enhancing children's self-regulation skills and their academic preparation for kindergarten. In addition, we compared the effectiveness of each into kindergarten and first grade. The research plan in 2010-11 (Cohort 1) and 2011-12 (Cohort 2) involved assessing the consented children in both *Tools* and Comparison classrooms at the beginning and end of the school year using a variety of self-regulation tasks and child achievement measures. Each classroom was also observed three times during the year using multiple measures designed to capture the differences between classrooms using *Tools* and those using other curricula.

This research study is being funded by a grant from the US Department of Education Institute of Education Sciences. Drs. Dale Farran, Mark Lipsey and Sandra Wilson of the Peabody Research Institute at Vanderbilt University are conducting this five-year study. 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.

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

A full description of the research design, the participants, assessments, immediate and follow up outcomes is provided in the final report of the project available online:

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. https://my.vanderbilt.edu/toolsofthemindevaluation/

Development of the Fidelity of Implementation Instrument

Tools of the Mind is a complex and dynamic curriculum with a specific sequence of preschool activities designed to develop children's ability to learn, as well as direct content knowledge, across the school year, with this sequence varying to some extent upon the strengths of the students within the classroom. At the time this project began, there were 62 activities but no developed measure of fidelity of implementation for the curriculum in this version. Project staff and the curriculum developers and trainers participated in a series of meetings to discuss (a) the important aspects of the curriculum that set it apart from other early childhood curricula, (b) how these characteristics could be measured or quantified, and (c) once measured, what implementation with fidelity would look like.

This report focuses on vertical fidelity, or measures of fidelity designed to differentiate among classrooms enacting *Tools*. We collected other data designed to measure horizontal fidelity, or the aspects of the curriculum thought to differentiate classrooms using *Tools* from those using another curriculum. Details on those data can be found in the overall technical report referenced above.

We started formulating our observation scheme by thinking about ways to measure the aspects of the curriculum the developers and trainers provided about behaviors/materials that were unique to *Tools* and should be present in every classroom enacting *Tools*. The use of make-believe play to build self-regulatory skills in children is the central focus of the curriculum. Thus, one of the main aspects the developers identified as being both critical and unique to *Tools of the Mind* was the presence of a defined make-believe play theme that cuts across all centers to encourage purposeful interactions and high level, authentic dramatic play complete with defined roles and role speech. Part of the make-believe play block should include having children plan what they are going to play through the use of a scaffold writing process and visual tool called a sound map. Individualization based on a child's zone of proximal development (ZPD) would be evident through the use and subsequent withdrawal of physical mediators and presence of differentiated levels of scaffolding. Children were expected to be working in pairs and using private speech to guide their actions. Rote copying and worksheet activities, assigned seating, and the use of external behavior reinforcements are all discouraged and thus would not be observed in classrooms utilizing the *Tools* curriculum.

We modified a classroom observation measure already in existence, the Narrative Record to serve as a framework for the fidelity instrument. The Narrative Record, described in more detail later in the report, captures not only behaviors we were interested in measuring across all classrooms in the study (experimental and control), but also behaviors we only expected to observe among Tools classrooms. We developed a fidelity measure to accompany the Narrative Record (to be completed by the same observer) and this measure was initially designed with the idea that an observer could assign a point value to a Tools activity based on a combination of whether or not the essential element (what defines the activity and makes it Tools) was present and how many of the important, but non-essential elements were present.

In trying to determine an essential element for each activity, it became apparent, during discussions with the developers and meetings after piloting the system in classrooms using Tools, that for most of the activities every one of the steps was considered essential. This led to the development of a fidelity measure that relied on a list of all the curriculum activities and their corresponding steps, mediators, and "should nots" (actions the developers said should not be happening if the activity were being done properly).

Documenting implementation of a curriculum this complex would not have been possible without the assistance of digital recording. Tablet computers were used for all observations. The fidelity system was created within FileMaker, a database. Run times of the fidelity system were installed on the tablets of each observer, enabling an observer to access the system but not to be able to make changes in the program. With the Narrative Record as the base, observers first made a decision about the major type of Tools activity occurring. Any time a Tools activity took place, the observer marked what activity was happening and then was given access to a list of steps, mediators, and "should nots" for that activity. The observer marked each step that was completed, marked any mediators that were observed, as well as whether anything occurred that would negate or violate the purpose of the activity (list from the developers of items/behaviors that we should not see during an activity). When the activity concluded, the observer returned to the Narrative Record home page.

A screen shot of one of the activity pages is attached at the end of this report. Similar screen shots for all of the activities are available at https://my.vanderbilt.edu/toolsofthemindevaluation/.

This approach removed the requirement for an observer to make any judgment about the quality of the implementation and allowed us to combine the concrete, behavioral data after the fact into a measure of implementation fidelity. This approach meant however that both the complexity of the curriculum and the implementation system required observers familiar with the curriculum. All observers participated in the Tools training plus receiving more extensive training from research staff before each round of observations.

Outlined in Tables 1-5 below are the activities grouped by major type, along with the number of specific steps in the prescribed sequence to be carried out at each observational time point (steps varied for activities at different time points during the year). The tables indicate which activities are indicated in the curriculum as appropriate to implement at the beginning of the school year when observation 1 was conducted (e.g., Mystery Question, Mystery Shape), and which are not appropriate to implement until the end of the year (e.g., Mystery Pattern, Mystery Letter).

In addition, as shown in the tables, activities varied in terms of their difficulty to implement. For example, implementing Make Believe Play Centers requires extensive preparation. Teachers must organize their classrooms around a central theme (e.g., restaurant, health clinic, grocery store), which includes removing props related to previous themes, incorporating new props that facilitate play around the new theme, and reorganizing centers around new scenarios related to the theme. On the other hand, Weather Graphing only requires development of a weather graph at the onset of the school year with implementation requiring teachers to have children update the graph each day.

Table 1. Tools of the Mind Large Group Activities by Observation

Large Group Activities	Level of Difficulty[1]	Observation 1		Observation 2		Observation 3	
		Tools Manual(s)[2]	Total steps[3]	Tools Manual(s)	Total steps	Tools Manual(s)	Total steps
Mystery Question	E	1	5				
Mystery Shape	E	2	4	2, 3	6		
Mystery Word	E			3	3	3, 4	7
Mystery Numeral	E			3	3	3, 4	6
Mystery Pattern	E					4	6
Mystery Letter	E					4	4
Mystery Rhyme	E					4	4
Timeline Calendar	E	1, 2	5	2, 3	7	3, 4	8
Weather Graphing	E	1, 2	6	2, 3	3	3, 4	3
Message of the Day	M	1, 2	6	2, 3	7	3, 4	8
Message of the Day Write Along	D					4	7
Share the News	E	1, 2	6	2, 3	5	3, 4	5
Share and Tell	E	1, 2	5	2, 3	5	3, 4	5
Tally	E					4	4
Write Along a Familiar Song/Finger Play	D					4	5
Make a Rhyme	M					4	5
Take Away Sounds	M					4	7
Class Schedules	E	1, 2	3	2, 3	3	3	3

Note. [1]Level of difficulty is the ease at which teachers can implement an activity without preparation or practice, E = Easy, M = Medium, and D = Difficult. [2]The *Tools of the Mind* curriculum is broken into four manuals that are implemented over the course of the school year, values indicate which manuals were being utilized at a given observation. Blank cells indicate that an activity was not to be implemented at that given observation. [3]Activities varied in the steps required to execute with fidelity, values indicate the required number of steps for each activity.

Table 2. Tools of the Mind Make Believe Play Activities by Observation

Make Believe Play Activities	Level of Difficulty[1]	Observation 1		Observation 2		Observation 3	
		Tools Manual(s)[2]	Total steps[3]	Tools Manual(s)	Total steps	Tools Manual(s)	Total steps
Make Believe Play Planning	D	1, 2	9	2, 3	10	3, 4	11
Make Believe Play Practice	D	1, 2	4	2, 3	4	3, 4	8
Make Believe Play	D	1, 2	5	2, 3	7	3, 4	11
Make Believe Play Cleanup	E	1, 2	3	2, 3	3	3, 4	3

Note. [1]Level of difficulty is the ease at which teachers can implement an activity without preparation or practice, E = Easy, M = Medium, and D = Difficult. [2]The *Tools of the Mind* curriculum is broken into four manuals that are implemented over the course of the school year, values indicate which manuals were being utilized at a given observation. Blank cells indicate that an activity was not to be implemented at that given observation. [3]Activities varied in the steps required to execute with fidelity, values indicate the required number of steps for each activity.

Table 3. Tools of the Mind Literacy and Story Lab Activities by Observation

Literacy and Story Lab Activities	Level of Difficulty[1]	Observation 1		Observation 2		Observation 3	
		Tools Manual(s)[2]	Total steps[3]	Tools Manual(s)	Total steps	Tools Manual(s)	Total steps
Graphics Practice	M	1, 2	9	2, 3	8	3, 4	11
Buddy Reading	M	1, 2	6	2, 3	9	3, 4	9
Elkonin Boxes 1: Jumping the Sounds	D					4	5
Elkonin Boxes 2: Token Game	D					4	4
I have who has Letters	E			3	8	3, 4	8
Story Lab: Active Listening	E	1, 2	6	2,3	6		
Story Lab: Connections	E	1, 2	5	2,3	5		
Story Lab: Vocabulary	D	1, 2	6	2,3	6		
Story Lab: Learning Facts	D	2	5	2,3	6		
Story Lab: Visualization	M	2	7	2,3	7		
Story Lab: Grammar	D			3	10		
Story Lab: Extensions	D			3	9		
Story Lab: Predictions and Inferences	D						

Note. [1]Level of difficulty is the ease at which teachers can implement an activity without preparation or practice, E = Easy, M = Medium, and D = Difficult. [2]The *Tools of the Mind* curriculum is broken into four manuals that are implemented over the course of the school year, values indicate which manuals were being utilized at a given observation. Blank cells indicate that an activity was not to be implemented at that given observation. [3]Activities varied in the steps required, values indicate the required number of steps for each activity at each time point.

Table 4. Tools of the Mind Math and Science Activities by Observation

Math and Science Activities	Level of Difficulty[1]	Observation 1		Observation 2		Observation 3	
		Tools Manual(s)[2]	Total steps[3]	Tools Manual(s)	Total steps	Tools Manual(s)	Total steps
Remember and Replicate	M	1, 2	8	2, 3	8	3	8
Puzzles and Manipulatives	E	1, 2	3				
Math Memory	M	2	8	2, 3	10	3, 4	12
Science Eyes	D	2	6	2, 3	8	3, 4	10
Numeral Game	M			3	5	3, 4	7
Venger Drawing	D			3	5	3, 4	6
Attribute Game	M			3	4	3, 4	6
Numberline Hopscotch	M			3	4	3	6
I have who has Colors	E			3	8	3	8
I have who has Numbers	E			3	8	3, 4	8
I have who has Shapes	E			3	8	3, 4	8
Making Collections	D	2	11	2, 3	11	3, 4	11
Patterns with Manipulatives	M					4	5

Note. [1]Level of difficulty is the ease at which teachers can implement an activity without preparation or practice, E = Easy, M = Medium, and D = Difficult. [2]The *Tools of the Mind* curriculum is broken into four manuals that are implemented over the course of the school year, values indicate which manuals were being utilized at a given observation. Blank cells indicate that an activity was not to be implemented at that given observation. [3]Activities varied in the steps required to execute with fidelity, values indicate the required number of steps for each activity.

Table 5. Tools of the Mind Across the Day Activities by Observation

Across the Day Activities	Level of Difficulty[1]	Observation 1		Observation 2		Observation 3	
		Tools Manual(s)[2]	Total steps[3]	Tools Manual(s)	Total steps	Tools Manual(s)	Total steps
Attention Focusing Activities	E	1, 2	5	2, 3	5	3, 4	6
Freeze Game	E	1, 2	4	2, 3	5	3	5
Partner Freeze	E					4	7
Two Step Freeze	M					4	4
Freeze on Number	M			3	4	3, 4	5
Pattern Movement Game	M	2	7	2, 3	7	3	9
Complete and Continue	M			3	7	3, 4	7
Number Follow the Leader	M			3	4	3, 4	5
Pretend Transitions	E	1, 2	3	2, 3	3	3, 4	3
Community Building Activities	E	1, 2	3	2, 3	3		
I have who has Name Game	E	1, 2	6	2, 3	6	3, 4	6
Mousetrap	E					4	5
What are you doing Mr. Wolf?	E					4	5

Note. [1]Level of difficulty is the ease at which teachers can implement an activity without preparation or practice, E = Easy, M = Medium, and D = Difficult. [2]The *Tools of the Mind* curriculum is broken into four manuals that are implemented over the course of the school year, values indicate which manuals were being utilized at a given observation. Blank cells indicate that an activity was not to be implemented at that given observation. [3]Activities varied in the steps required to execute with fidelity, values indicate the required number of steps for each activity.

Procedures

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 for the entire length of the school day to note all instructional classroom activities. 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 Preschool (TOP)* and the *Child Observation in Preschool (COP).* The *TOP* and *COP* are not discussed in this report since the focus is on measures that gathered information designed to differentiate vertical fidelity (among *Tools* classrooms) and not horizontal fidelity (between *Tools* and control). These latter measures are described in detail in the final Technical report.

Each of the measures connected to the fidelity of implementation system will be described and descriptive data from its use will be provided for each separately.

Narrative Record

The *Narrative Record*[1] is an open-ended format for recording descriptive narrative data notes and also rating the activities occurring in the classroom. 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.

Because the Narrative Record was interlinked with the fidelity system but also served to record all the activities in the room including ones not curriculum specific, a variety of classroom summaries can be made available. Tables 6-9 focus only on those teachers who were supposed to be implementing the curriculum. Data for each cohort are presented separately.

Table 6 breaks out the hours and minutes devoted at each observational period overall to implementing the curriculum, to doing other activities not connected to the curriculum and to time spent non-instructionally (meals, routines, naps, etc.). It can be seen that cohort 2 spent more time implementing the curriculum than cohort 1. But for both cohorts, the time available to carry out this very complex curriculum is only an hour or a little more. Table 7 captures the same information but presented as proportions of the day.

The curriculum manuals do not provide a clear description of the organization of the day, and the information about how time should be spent varies across the different manuals teachers are given throughout the year. Research staff went through the curriculum manuals from each time point to determine (roughly) how much time teachers were expected to devote to each major domain of the curriculum (e.g., Literacy Activities, Make Believe Play). A comparison of time prescribed by the curriculum and time actually observed is provided in Tables 8 and 9. It will be apparent that teachers did not have time during the day to implement the activities for the amount suggested. Even if the teachers were to reduce their transitions and cease doing non-Tools activities, there still would not have been enough time available.

.

[1] For more information see: Farran, D.C. & Bilbrey, C. (2004). *Narrative Record Observation for Classrooms*. Nashville, TN: Peabody Research Institute, Vanderbilt University

Table 6. Descriptive Statistics Total Time (Hours: Minutes) in Tool and Non-Tools Activities by Observation

		Observation 1			Observation 2			Observation 3		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Cohort 1										
Tools Instructional Time	32	0:01	3:06	1:02 (0:40)	0:14	2:03	0:59 (0:30)	0:05	2:14	0:48 (0:30)
Non-Tools Instructional Time	32	0:19	2:58	1:40 (0:38)	0:31	2:52	1:50 (0:31)	0:34	2:31	1:50 (0:29)
Non-Instructional Time	32	1:21	5:15	3:37 (0:44)	1:28	4:54	3:30 (0:40)	2:24	5:39	3:41 (0:40)
Cohort 2										
Tools Instructional Time	10	0:15	1:56	1:24 (0:31)	0:20	2:05	1:32 (0:32)	0:20	2:05	1:32 (0:32)
Non-Tools Instructional Time	10	1:24	2:41	2:00 (0:26)	1:26	2:55	1:57 (0:27)	1:26	2:55	1:57 (0:27)
Non-Instructional Time	10	3:14	4:12	3:40 (0:20)	3:12	4:02	3:34 (0:17)	3:12	4:02	3:34 (0:17)

Note. Non-Tools instructional time includes free choice centers. Non-instructional time includes non-instructional transitions, meals, naps, and out of classroom time.

Table 7. Descriptive Statistics Proportion of School Day in Tool and Non-Tools Activities by Observation

		Observation 1			Observation 2			Observation 3		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Cohort 1										
Tools Instructional Time	32	0.05	0.54	0.27 (0.11)	0.08	0.41	0.29 (0.09)	0.09	0.42	0.29 (0.08)
Non-Tools Instructional Time	32	0.00	0.47	0.16 (0.10)	0.04	0.37	0.16 (0.08)	0.02	0.33	0.12 (0.08)
Non-Instructional Time	32	0.25	0.85	0.57 (0.10)	0.27	0.80	0.55 (0.09)	0.44	0.83	0.58 (0.09)
Cohort 2										
Tools Instructional Time	10	0.20	0.40	0.28 (0.07)	0.20	2:05	0.28 (0.07)	0.22	0.38	0.32 (0.06)
Non-Tools Instructional Time	10	0.04	0.27	0.20 (0.07)	0.05	2:55	0.22 (0.07)	0.05	0.29	0.17 (0.07)
Non-Instructional Time	10	0.46	0.58	0.52 (0.05)	0.45	4:02	0.51 (0.05)	0.42	0.60	0.51 (0.05)

Note. Non-Tools instructional time includes free choice centers. Non-instructional time includes non-instructional transitions, meals, naps, and out of classroom time.

Table 8. Cohort 1: Classroom Schedule in Minutes as Designated by the Tools of the Mind Manual and Observed

		Observation 1		Observation 2		Observation 3	
	n	Manual 1 & 2	Observed *M* (*SD*)	Manual 3	Observed *M* (*SD*)	Manual 4	Observed *M* (*SD*)
Large Group	32	30	18.22 (13.64)	30	16.04 (6.86)	30	21.31 (12.32)
Free Choice Centers	32	60	25.70 (22.59)	35	29.92 (17.81)	35	19.01 (19.12)
Make Believe Play Planning	32	15	8.95 (5.69)	15	11.09 (6.18)	15	12.02 (5.58)
Make Believe Play Centers	32	25	14.01 (10.81)	45	20.54 (13.82)	45	18.51 (11.35)
Make Believe Play Practice	32	15	3.48 (7.43)	0	4.29 (6.72)	0	2.19 (3.57)
Literacy	32	30	19.03 (9.25)	30	20.77 (14.81)	30	18.97 (11.09)
Math/Science	32	10	0.87 (4.58)	10	4.47 (7.48)	10	3.15 (7.37)
Mixed Tools Instruction	32	--	18.61 (14.40)	--	17.43 (12.90)	--	19.67 (15.21)
Non-Tools Instruction	32	--	32.77 (26.38)	--	25.98 (23.16)	--	25.75 (17.26)
Tools Transitions	32	10	12.91 (5.03)	10	10.86 (4.81)	5	9.95 (4.53)
Non-Tools Transitions	32	--	39.48 (17.92)	--	33.89 (11.11)	--	36.37 (18.60)
Meal/Nap/Out	32	165[1]	165.98 (45.87)	185[1]	164.73 (36.35)	190[1]	173.10 (40.47)

Note. Values reported represent minutes of a 6 hour school day. [1]Reflects the proportion of the school day not allocated to other activities by the *Tools of the Mind* manuals.

Table 9. Cohort 2: Classroom Schedule in Minutes Designated by the Tools of the Mind Manuals and Observed

		Observation 1		Observation 2		Observation 3	
	n	Manual 1 & 2	Observed *M* (*SD*)	Manual 3	Observed *M* (*SD*)	Manual 4	Observed *M* (*SD*)
Large Group	10	30	14.75 (10.00)	30	13.32 (7.13)	30	11.60 (2.82)
Free Choice Centers	10	60	54.26 (26.74)	35	61.92 (25.95)	35	45.13 (23.68)
Make Believe Play Planning	10	15	6.48 (4.70)	15	11.38 (2.56)	15	11.33 (1.97)
Make Believe Play Centers	10	25	29.24 (18.19)	45	30.24 (19.69)	45	27.62 (15.88)
Make Believe Play Practice	10	15	2.27 (3.86)	0	0.41 (1.29)	0	0.91 (1.77)
Literacy	10	30	14.78 (7.31)	30	10.05 (6.26)	30	14.19 (6.65)
Math/Science	10	10	2.07 (4.39)	10	2.32 (7.35)	10	4.20 (6.85)
Mixed Tools Instruction	10	--	22.98 (8.91)	--	22.85 (12.38)	--	35.28 (16.00)
Non-Tools Instruction	10	--	16.40 (16.98)	--	16.13 (13.82)	--	16.51 (9.47)
Tools Transitions	10	10	9.63 (3.59)	10	9.16 (3.49)	5	8.30 (2.16)
Non-Tools Transitions	10	--	29.35 (10.92)	--	29.04 (8.63)	--	30.76 (7.73)
Meal/Nap/Out	10	165[1]	157.79 (16.79)	185[1]	153.18 (17.32)	190[1]	154.17 (20.31)

Note. Values reported represent minutes of a 6 hour school day. [1]Reflects the proportion of the school day not allocated to other activities by the *Tools of the Mind* manuals.

Fidelity of Implementation

Determining what scores to summarize from the measure to determine variations in implementation fidelity was not simple. On the one hand, we could simply count the number of activities we observed being implemented and the number of steps enacted. On the other hand, we were aware of the differences among the activities as described above. To account for the differences in difficulty and preparation required for an activity, we combined our behavioral data into a weighted fidelity score for each observation. The weighted score adjusted for the difficulty level of each activity and the time-appropriateness of the steps enacted from the curriculum manuals. Based on the implementation difficulty, each activities was either worth a maximum score of 10 (Easy), 20 (Medium), or 40 (Difficult). This categorization was determined through collaboration with the developers and trainers.

Scores were adjusted based on the proportion of required steps implemented by the teacher. For example, if a teacher implemented 4 of the 5 steps indicated in the manual (based on the time of year in which the observation took place) for Mystery Question, an Easy activity, a score of 8 would be given. If a teacher implemented 4 of the 9 steps indicated in the manual for Make Believe Play Planning, a difficult activity, a score of 17.8 would be given. Weighted fidelity scores were only calculated for activities and steps that were indicated in the manual as appropriate to implement at a given observation. A total weighted fidelity scores was estimated by summing across the scores for all appropriate activities observed at a given observation. Based on the curriculum manuals, if fully implemented, weighted fidelity scores could range from 380 to 460 at Observation 1, 370 to 530 at Observation 2, and 350 to 570 at Observation 3. These scores correspond to the manuals' indication that 22 activities should be implemented over the course of a school day.

The Tools final technical report provides information about how well the simple activity count variable and the weighted fidelity score predicted child gain in achievement and self-regulation, both immediately and through first grade.

Presented in Tables 10-11 below is information about the fidelity scores achieved at each time point and separately for each cohort. The Tools developers changed the emphasis of training and coaching for cohort 1 once they reviewed the results for cohort 2, emphasizing more implementation of make believe play activities. As can be seen, there was much less variability in implementation among the teachers in cohort 2. Table 12 provides evidence of the relations between observational periods – how consistent did teachers implement the curriculum. It is evident that there was fall off in implementation at the last observational period conducted in the late spring. Table 13 presents evidence of the relationship among parts of the observational system.

Table 10. Cohort 1: Descriptive Statistics Fidelity of Implementation by Observation

		Observation 1			Observation 2			Observation 3		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Weighted Fidelity[1]	32	28	259	154.05 (55.14)	23	290	180.36 (58.42)	48	293	164.42 (53.42)
Time Appropriate Activities[1]	32	4	19	12.47 (3.28)	3	20	13.81 (3.58)	5	19	13.75 (3.29)
All Activities[2]	32	5	22	13.97 (3.57)	4	22	14.91 (3.77)	6	20	14.84 (3.34)
Time Appropriate Steps[1]	32	12	70	45.25 (14.40)	8	84	55.75 (16.55)	15	86	58.31 (16.35)
All Steps[2]	32	16	78	53.66 (16.18)	11	95	61.81 (18.27)	15	91	62.00 (17.13)
Mediators[2]	32	12	46	30.75 (7.96)	12	48	32.31 (7.84)	7	44	32.47 (8.21)
Should Nots[2]	32	0	8	4.78 (2.57)	0	9	3.19 (3.06)	0	12	4.44 (2.91)

Note. [1]Values reported only include activities and steps that were indicated as appropriate to implement by the curriculum at the given observation point in the school year. [2] Values reported include all 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 11. Cohort 2: Descriptive Statistics Fidelity of Implementation by Observation

		Observation 1			Observation 2			Observation 3		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Weighted Fidelity[1]	10	156	204	182.72 (18.02)	146	210	170.69 (23.46)	126	228	172.01 (33.23)
Time Appropriate Activities[1]	10	10	16	14.10 (1.79)	10	15	12.80 (1.93)	10	17	14.10 (2.02)
All Activities[2]	10	14	18	15.70 (1.34)	11	18	15.40 (2.07)	12	17	15.30 (1.83)
Time Appropriate Steps[1]	10	48	68	55.20 (6.21)	43	60	52.10 (6.21)	42	79	57.80 (10.17)
All Steps[2]	10	55	81	66.10 (8.63)	44	65	56.80 (7.54)	44	79	60.10 (9.63)
Mediators[2]	10	28	46	35.50 (4.62)	24	38	33.60 (4.09)	26	45	35.70 (5.77)
Should Nots[2]	10	0	10	3.40 (3.17)	0	10	4.70 (3.09)	0	9	4.90 (2.42)

Note. [1]Values reported only include activities and steps that were indicated as appropriate to implement by the curriculum at the given observation point in the school year. [2] Values reported include all 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 12. Fidelity of Implementation Zero-Order Correlations Among Observations (Cohorts 1 and 2, *N* = 42)

	Observation 1 & 2	Observation 1 & 3	Observation 2 & 3
Weighted Fidelity	.637**	.678**	.433**
Time Appropriate Activities	.653**	.579**	.524**
All Activities	.685**	.619**	.566**
Time Appropriate Steps	.643**	.645**	.553**
All Steps	.626**	.619**	.600**
Mediators	.681**	.643**	.556**
Should Nots	.230	.212	.293

**$p < .01$. *$p < .05$.

Table 13. Zero-Order Correlations Among Weighted Fidelity and Implementation of Activities and Steps by Observation (Cohorts 1 and 2, *N* = 42)

	Weighted Fidelity		
	Observation 1	Observation 2	Observation 3
Time Appropriate Activities	.887**	.836**	.817**
All Activities	.878**	.818**	.793**
Time Appropriate Steps	.925**	.926**	.912**
All Steps	.897**	.914**	.896**

**$p < .01$. *$p < .05$.

Fidelity of Implementation Latent Profile Analysis

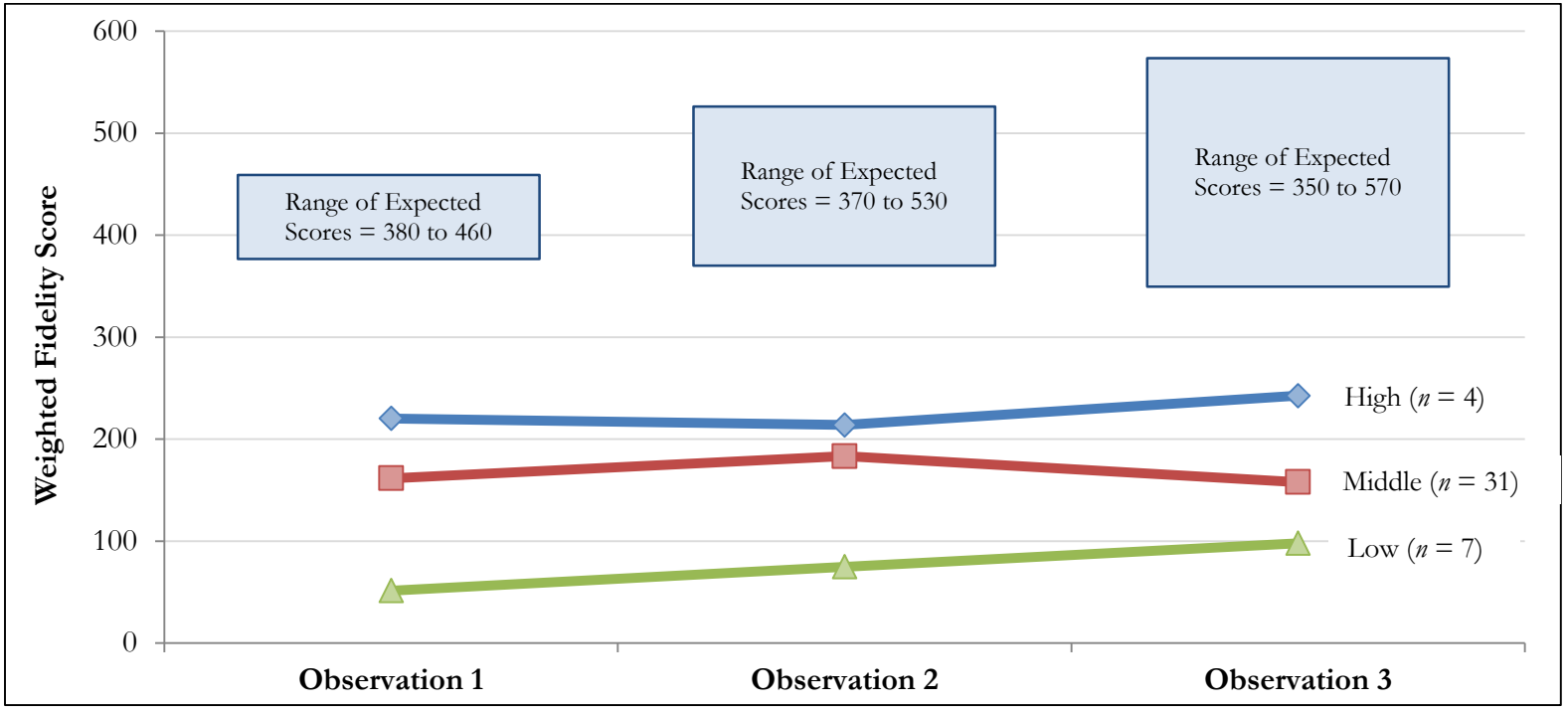
Latent profile analysis (LPA) was used to classify the 42 Tools of the Mind classrooms into subgroups based on their weighted fidelity of implementation scores at the three classroom observations. LPA was conducted starting with testing a one-profile model and then testing alternative models that successively increased numbers the number of profiles. Analyses included both cohorts. Models were compared using the Bayesian information criterion (BIC; Schwartz, 1978), the bootstrapped log ratio test (BLRT; McLachlan, 1987), and estimates of entropy. The three-profile model was found to be the best fitting with the smallest BIC (1335.64) and largest estimate of entropy (0.90). The BLRT indicated that the three-profile models had a significantly better fit than the two-profile model ($p < .001$) and was no different than the four-profile model ($p = .667$). Descriptive statistics for the three profiles are provided in Table 14.

A 3 x 3 repeated measures analysis of variance (RM-ANOVA) indicated that there was a marginal effect of observation and significant effect of profile membership, $F(2,78) = 2.62$, $p = .079$ and $F(2,39) = 64.90$, $p < .001$. Moreover, there was a significant interaction between observation and profile membership, $F(4,78) = 3.32$, $p = .014$. Exploration of pairwise comparisons indicated that at all three observations, classrooms in the *Low* profile had significantly lower weighted fidelity scores compared to classrooms in the *Middle* profile (ps $\leq .001$). At observation 1 and 3 classrooms in the *Middle* profile were significantly lower that classrooms in the *High* profile (ps $< .001$), yet the effect was marginal at observation 2 ($p = .068$). For classrooms in the *Low* profile, scores at observation 1 were significantly lower than scores at observation 3 ($p = 0.17$), which were both no different than scores at observation 2. For classrooms in the *Middle* profile, scores at observation 1 and 3 were significantly lower than scores at observation 2 (ps ≤ 0.08), which were no different from each other. Lastly, for classrooms in the *High* profile, there were no significant differences in scores across the three observations.

Table 14. Fidelity of Implementation Latent Profile Descriptives by Observation

	Cohort 1 *n*	Cohort 2 *n*	Observation 1 Min	Observation 1 Max	Observation 1 Mean (*SD*)	Observation 2 Min	Observation 2 Max	Observation 2 Mean (*SD*)	Observation 3 Min	Observation 3 Max	Observation 3 Mean (*SD*)
Low Profile	6	1	28	67	51.47 (17.36)	23	95	74.84 (34.47)	48	157	97.97 (47.28)
Middle Profile	22	9	104	214	161.59 (29.21)	122	290	183.30 (41.13)	106	222	157.80 (28.12)
High Profile	4	0	184	259	220.27 (26.89)	161	251	213.83 (28.10)	202	293	242.56 (33.66)

Figure 1. Fidelity of Implementation Latent Profiles



Environmental 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 indicated should be available in the classroom. The checklist focuses on the play centers, classroom themes, and materials accessible to children.

Table 15. Frequency of Make Believe Play Themes by Observation

	n	Observation 1	Observation 2	Observation 3
Cohort 1				
Community/Community Helpers	32	0	3	5
Grocery Store/Supermarket	32	1	10	5
Home Living/Family	32	5	0	0
Hospital/Doctors	32	0	8	5
Mall	32	0	1	2
Restaurant	32	17	3	5
School	32	0	1	2
Veterinarian/Pets	32	0	0	2
Other	32	0	1	3
None	32	9	5	3
Cohort 2				
Community/Community Helpers	10	0	2	0
Grocery Store/Supermarket	10	1	4	5
Home Living/Family	10	2	0	0
Hospital/Doctors	10	0	0	5
Mall	10	0	0	0
Restaurant	10	6	3	0
School	10	0	1	0
Veterinarian/Pets	10	0	0	0
Other	10	1	0	0
None	10	0	0	0

Table 16. Proportion of Classroom Materials Accessible to Children by Observation

		Observation 1			Observation 2			Observation 3		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Cohort 1										
Arts and Music Materials (12 items)	32	0.08	1.00	0.64 (0.21)	0.33	1.00	0.77 (0.18)	0.25	1.00	0.64 (0.21)
Furniture and Equipment (3 items)	32	0.33	1.00	0.85 (0.21)	0.67	1.00	0.93 (0.14)	0.33	1.00	0.86 (0.20)
Manipulatives (3 items)	32	0.24	0.94	0.55 (0.17)	0.18	0.94	0.57 (0.19)	0.18	0.88	0.57 (0.21)
Math and Science Materials (5 items)	32	0.33	1.00	0.65 (0.22)	0.00	1.00	0.55 (0.37)	0.00	1.00	0.60 (0.31)
Make Believe Play Materials (17 items)	32	0.00	1.00	0.43 (0.26)	0.00	1.00	0.45 (0.26)	0.00	1.00	0.41 (0.27)
Tools Materials (12 items)	32	0.17	0.83	0.52 (0.18)	0.08	0.83	0.52 (0.20)	0.17	0.92	0.56 (0.20)
Cohort 2										
Arts and Music Materials (12 items)	10	0.83	1.00	0.92 (0.07)	0.67	1.00	0.90 (0.11)	0.75	1.00	0.90 (0.10)
Furniture and Equipment (3 items)	10	1.00	1.00	1.00 (0.00)	0.67	1.00	0.97 (0.11)	0.67	1.00	0.97 (0.11)
Manipulatives (3 items)	10	0.53	0.82	0.66 (0.11)	0.65	0.88	0.77 (0.08)	0.59	0.88	0.74 (0.10)
Math and Science Materials (5 items)	10	0.33	1.00	0.70 (0.19)	0.33	1.00	0.73 (0.21)	0.67	1.00	0.80 (0.17)
Make Believe Play Materials (17 items)	10	0.40	0.80	0.62 (0.11)	0.20	1.00	0.68 (0.23)	0.60	1.00	0.70 (0.14)
Tools Materials (12 items)	10	0.42	0.92	0.63 (0.16)	0.17	1.00	0.77 (0.18)	0.25	0.92	0.69 (0.21)

Table 17. Cohort 1: Proportion of Specific Tools of the Mind Materials Accessible to Children by Observation

		Observation 1	Observation 2	Observation 3
	n	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Action prompt cards	32	0.41 (0.50)	0.47 (0.51)	0.44 (0.50)
Child-made props	32	0.56 (0.50)	0.31 (0.47)	0.50 (0.51)
Dry erase markers for entire class	32	0.69 (0.47)	0.75 (0.44)	0.91 (0.30)
ELL labels (for centers and materials)	32	0.22 (0.42)	0.22 (0.42)	0.25 (0.44)
Free play center labels	32	0.94 (0.25)	0.91 (0.30)	0.94 (0.25)
Individual wipe-off boards for class	32	0.75 (0.44)	0.78 (0.42)	0.94 (0.25)
Make-believe center labels	32	0.63 (0.49)	0.47 (0.51)	0.66 (0.48)
Math activities in non-math centers	32	0.28 (0.46)	0.09 (0.30)	0.22 (0.42)
Plastic tubs to sort buddy reading books	32	0.75 (0.44)	0.78 (0.42)	0.69 (0.47)
Role prompt cards	32	0.47 (0.51)	0.56 (0.50)	0.47 (0.51)
Vocabulary words posted	32	0.19 (0.40)	0.38 (0.49)	0.19 (0.40)
Writing materials in at least two non-writing centers	32	0.41 (0.50)	0.53 (0.51)	0.56 (0.50)

Table 18. Cohort 2: Proportion of Specific Tools Materials Accessible to Children by Observation

		Observation 1	Observation 2	Observation 3
	n	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Action prompt cards	10	0.50 (0.53)	0.60 (0.52)	0.60 (0.52)
Child-made props	10	0.50 (0.53)	0.70 (0.48)	0.90 (0.32)
Dry erase markers for entire class	10	0.90 (0.32)	0.90 (0.32)	0.90 (0.32)
ELL labels (for centers and materials)	10	0.40 (0.52)	0.40 (0.52)	0.50 (0.53)
Free play center labels	10	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Individual wipe-off boards for class	10	0.90 (0.32)	0.80 (0.42)	0.90 (0.32)
Make-believe center labels	10	0.70 (0.48)	0.70 (0.48)	0.70 (0.48)
Math activities in non-math centers	10	0.20 (0.42)	0.30 (0.48)	0.60 (0.52)
Plastic tubs to sort buddy reading books	10	1.00 (0.00)	0.80 (0.42)	0.80 (0.42)
Role prompt cards	10	0.30 (0.48)	0.50 (0.53)	0.40 (0.52)
Vocabulary words posted	10	0.40 (0.52)	0.70 (0.48)	0.50 (0.53)
Writing materials in at least two non-writing centers	10	0.70 (0.48)	0.70 (0.48)	0.50 (0.53)

Tools Trainer Ratings of Teacher Fidelity

The *Trainer's Rating Form* was designed by curriculum developer, Elena Bodrova. This information was collected at the end of the study year by the *Tools* trainer working with each school system. The *Tools* trainers observed in classrooms before/after each of the 4 trainings conducted throughout the school year. This rating form asked trainers to rate each teacher's level of understanding of the curriculum and the level of implementation.

Table 19. Tools Trainer Ratings of Teacher Fidelity

	n	Min	Max	Mean (*SD*)
Cohort 1				
Level of Implementation	32	1	5	3.25 (1.30)
Level of Understanding	32	1	5	3.25 (1.41)
Overall Score	32	1	5	3.25 (1.26)
Cohort 2				
Level of Implementation	10	1	5	3.90 (1.37)
Level of Understanding	10	1	5	2.80 (1.75)
Overall Score	10	1	5	3.35 (1.49)

Note. Items rated on a 1 to 5 scale with higher ratings indicating greater fidelity. Overall score is the average of Level of Implementation and Understanding.

Table 20. Zero-Order Correlations between Tools Trainer Ratings of Fidelity and Observed Fidelity of Implementation, Both Cohorts Combined

	Amount of Tools Instruction[1]	Weighted Fidelity Score[2]
Level of Implementation	.35**	.60**
Level of Understanding	.09	.31*
Overall Score	.23	.48**

Note. $N = 42$. [1]Proporion of school day in Tools Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument.
$^{**}p < .01$. $^{*}p < .05$.

Teacher Survey

The teacher survey was designed to gather both demographic information about the teacher and assistant (years teaching, certification, etc.) as well as *Tools* specific information, such as what aspects the teacher liked the most about the curriculum and how they would rate the training they received. This survey also asked teachers to self-report fidelity by asking how often they did each of the *Tools* activities over the past month (never, rarely, frequently, daily). If an activity reported doing an activity "never" or "rarely" during the past month, the teacher was prompted to select a reason (not enough time, not important, too difficult or complicated, did not like the activity, not a daily activity in the curriculum, not the right time of year). Each teacher completed a survey at the beginning and end of the study year.

Table 21 presents summaries of teacher ratings in different categories for each cohort separately. Teachers on average held a moderately positive view toward the curriculum but the range of opinions was quite large. In neither cohort did teachers tend to agree that the curriculum was easy to implement. Interestingly, the ratings from cohort 2 teachers were somewhat lower across the board than cohort 1 teachers. The cohort 2 teachers may have felt more pressure; the developers and the coach strongly pushed for greater implementation, which did occur (see Tables 10 and 11), but apparently not happily.

Table 22 presents the correlations between teachers' ratings of the curriculum and the degree of implementation captured by our fidelity system. The weighted fidelity score correlates more highly with most ratings and the teachers' overall evaluation of the curriculum is moderately related to what we saw them implement.

Table 21. Teacher Ratings of the Tools Curriculum in the Fall and Spring of the Implementation Year

		Fall			Spring		
	n	Min	Max	Mean (*SD*)	Min	Max	Mean (*SD*)
Cohort 1							
How easy have you found the *Tools of the Mind* curriculum to implement?	32	2	5	2.91 (0.86)	1	5	3.09 (1.06)
How would you evaluate the training you have been offered to implement the *Tools of the Mind* curriculum?	32	2	5	3.81 (0.82)	2	5	4.28 (0.85)
How would you evaluate the coaching you have received on the *Tools of the Mind* curriculum?	32	1	5	3.81 (1.06)	1	5	4.13 (0.98)
How would you evaluate the classroom visits you have received from the *Tools* trainer?	32	2	5	3.66 (0.83)	2	5	4.13 (0.83)
Overall, how would you rate the *Tools of the Mind* curriculum so far?	32	1	5	3.34 (0.97)	1	5	3.78 (1.10)
Cohort 2							
How easy have you found the *Tools of the Mind* curriculum to implement?	10	1	4	2.60 (0.97)	1	4	2.70 (0.95)
How would you evaluate the training you have been offered to implement the *Tools of the Mind* curriculum?	10	2	5	3.70 (0.95)	2	5	3.80 (0.92)
How would you evaluate the coaching you have received on the *Tools of the Mind* curriculum?	10	1	5	3.00 (1.49)	2	5	3.60 (1.07)
How would you evaluate the classroom visits you have received from the *Tools* trainer?	10	2	5	3.50 (0.97)	2	5	3.70 (0.95)
Overall, how would you rate the *Tools of the Mind* curriculum so far?	10	2	5	3.20 (0.92)	1	5	3.30 (1.25)

Note. Items rated on a 1 to 5 scale with higher ratings indicating a more favorable view.

Table 22. Concurrent Zero-Order Correlations between Teacher Survey Items and Observed Fidelity of Implementation (Cohorts 1 and 2, *N* = 42)

	Amount of Tools Instruction[1]		Weighted Fidelity Score[2]	
	Fall	Spring	Fall	Spring
Teacher Rating of Curriculum				
How easy have you found the *Tools of the Mind* curriculum to implement?	.15	.20	.35*	.33**
How would you evaluate the training you have been offered to implement the *Tools of the Mind* curriculum?	.25	.41**	.39*	.39*
How would you evaluate the coaching you have received on the *Tools of the Mind* curriculum?	-.01	.22	.08	.21
How would you evaluate the classroom visits you have received from the *Tools* trainer?	.27	.23	.28	.27
Overall, how would you rate the *Tools of the Mind* curriculum so far?	.49**	.36*	.52**	.32*
Teacher Reported Implementation				
All Activities	.34*	.39*	.35*	.53**

Note. [1]Proportion of school day in *Tools* Instruction from Narrative Record. [2]Weighted Fidelity Score from Fidelity of Implementation Instrument. Fall Survey correlated with Obs 1. Spring Survey correlated with Obs. 3.
**$p < .01$. *$p < .05$.

Conclusion and Future Directions

Determining whether teachers implemented a curriculum to which they were randomly assigned is not an easy task. Although teachers were not instrumental in choosing this curriculum, the process followed by the school systems was typical of administrative decisions to adopt a new curriculum. In fact, the training procedures were actually much more extensive than most school systems can afford. Training was provided and in-class coaching occurred for a full year before the teachers were expected to implement the curriculum fully. During the implementation year, training sessions and coaching continued.

The difficulties with a curriculum as complex as Tools are myriad. First, no predefined estimate of adequate implementation exists. Teachers had no goals to strive for, and trainers and coaches had no goals to pursue. As researchers, we had no way to know if the implementation had been all that could be achieved. Second, the curriculum continues to grow without reference to what is actually possible in a classroom. When we wrote the proposal, the curriculum consisted of 40 activities; by the time we were funded there were 60 activities. Three more were added during the year of training before implementation. As we have seen, the actual time teachers have available to implement a curriculum is substantially less than would be required to implement the Tools of the Mind curriculum as it existed at the time of this study.

The teachers completed many Tools activities; they did so at appropriate times and following the right number of steps. When the developers were provided the results from the first cohort, they changed the emphasis for the second cohort, choosing to stress higher levels of implementation especially of the make believe play activities. Cohort 2 teachers implemented more activities and there was much less variation among them in the number of activities they carried out. However, these changes were not related.to better outcomes for the children.

Curriculum developers often lose touch with real classrooms. It is understandable given the pressure of producing materials, training trainers, and selling their curricular approach to school systems. But without grounding in actual classrooms, developers can easily create unrealistic expectations for what teachers can actually do.

Curricula being sold to school systems need to be tested completely. Often only a few activities if that have been empirically verified. In order to determine if a curriculum can be implemented within the constraints of current classrooms, developers should implement the full curriculum experimentally.

INSERT PDF OF SAMPLE PAGE OF FIDELITY WHEN CONVERTED TO PDF