Early Childhood Research Quarterly 23 (2008) 299–313

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Educational effects of the Tools of the Mind curriculum: A randomized trial

W. Steven Barnett[a,*], Kwanghee Jung[a], Donald J. Yarosz[a], Jessica Thomas[a], Amy Hornbeck[a], Robert Stechuk[b], Susan Burns[b]

[a] National Institute for Early Education Research, Rutgers-The State University of New Jersey, United States
[b] George Mason University, United States

ARTICLE INFO

Article history:
Received 24 February 2006
Received in revised form 21 March 2008
Accepted 23 March 2008

Keywords:
Curriculum
Preschool education
Early childhood education
Learning
Social development
Self-regulation

ABSTRACT

The effectiveness of the *Tools of the Mind* (*Tools*) curriculum in improving the education of 3- and 4-year-old children was evaluated by means of a randomized trial. The *Tools* curriculum, based on the work of Vygotsky, focuses on the development of self-regulation at the same time as teaching literacy and mathematics skills in a way that is socially mediated by peers and teachers and with a focus on play. The control group experienced an established district-created model described as a "balanced literacy curriculum with themes." Teachers and students were randomly assigned to either treatment or control classrooms. Children (88 *Tools* and 122 control) were compared on social behavior, language, and literacy growth. The *Tools* curriculum was found to improve classroom quality and children's executive function as indicated by lower scores on a problem behavior scale. There were indications that Tools also improved children's language development, but these effects were smaller and did not reach conventional levels of statistical significance in multi-level models or after adjustments for multiple comparisons. Our findings indicate that a developmentally appropriate curriculum with a strong emphasis on play can enhance learning and development so as to improve both the social and academic success of young children. Moreover, it is suggested that to the extent child care commonly increases behavior problems this outcome may be reversed through the use of more appropriate curricula that actually enhance self-regulation.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Most children in the United States now attend public or private preschool programs for 1 or 2 years prior to kindergarten (Barnett & Yarosz, 2004). Public investments in preschool education have largely focused on improving the school success of low-income and other children at high risk of school failure. The federal and state governments now invest considerable sums in such preschool programs, motivated by research demonstrating that preschool programs can contribute to impressive short- and long-term gains in cognitive, language, and social-emotional development (Bowman, Donovan, & Burns, 2001). However, many questions remain about the most effective approaches to educating young children, including curricular questions about the appropriate balance of teaching methods and children's experiences, and the knowledge and skills that young children should be expected to learn (Zigler & Bishop-Josef, 2006).

* Corresponding author at: National Institute for Early Education Research, Rutgers-The State University of New Jersey, 120 Albany Street, Suite 500, New Brunswick, NJ 08901, United States. Tel.: +1 732 932 4350.
E-mail addresses: sbarnett@nieer.org, wbarnet@rci.rutgers.edu (W.S. Barnett).

0885-2006/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2008.03.001

This study investigates the effects on learning and development of the Tools of the Mind (*Tools*) curriculum, with particular emphasis on self-regulation and emergent literacy. *Tools* is based on a socio-cultural perspective pioneered by Vygotsky (1978) that construes child development to be interactive and constructivist in its orientation (Bodrova & Leong, 2001, 2007). Its design is consistent with the principle that success in preschool is best promoted when children experience environments in which they have an opportunity to be active participants in learning and they are challenged and supported in that learning (Krafft & Berk, 1998). While child-centered, *Tools* also emphasizes the teacher's role in guiding and supporting the child's learning. As we explain below, it does not fit neatly into frameworks that classify curricula as teacher-directed or child initiated, child-centered or content-centered, and academic-focused or socialization-focused.

Our research design employs random assignment of both teachers and students to either treatment (*Tools*) or control (a "generic" district-developed curriculum) conditions. Previous quasi-experimental studies of *Tools* led us to expect that this curriculum would change children's classroom experiences in ways that would improve children's learning and development particularly with respect to emergent literacy and self-regulation (Bodrova & Leong, 2001). Self-regulation has been defined in a variety of ways, but self-regulatory capacity encompasses the ability to control one's physical, emotional, and cognitive functioning (Bronson, 2000). Progress in the development of self-regulation is regarded as "one of the really central and significant cognitive developmental hallmarks of the early childhood period" (Flavell, 1977, p. 64). This experiment was conducted in a low-income urban school district with a high proportion of children from low-income and non-English-speaking families. Such children are especially vulnerable to reading failure and other adverse outcomes if they do not receive high quality education in preschool and later in primary school (Bowman et al., 2001; Snow, Burns, & Griffin, 1998).

We have three primary research hypotheses. First, the *Tools* curriculum will lead to higher quality educational experiences for children, and these should be related to specific *Tools*' curricular emphases including the teacher's use of scaffolding. Second, *Tools* will produce larger gains in children's self-regulation and these gains should be evident in teacher reports of behavior problems (Blair, 2002a,b). Third, *Tools* will produce increases in children's cognitive and language development, especially emergent literacy skills, primarily because of elements of the curriculum directly related to literacy and only secondarily because of its emphasis on self-regulation (Diamond, 2007; Duncan et al., 2006). The comparative advantage for *Tools* over the control curriculum is expected to be greater for self-regulation as the control curriculum also emphasizes literacy development.

2. Literature review

The history of research on preschool curriculum is at once rich and limited. Debates about the appropriate extent of play, teacher directedness, individualization and other differences in curricular approaches span the entire history of early childhood education in the United States (Nourot, 2004). Among the most hotly debated questions are how much classroom time should be devoted to play, and how and to what extent play should be guided (Singer, Golinkoff, & Hirsh-Pasek, 2006). The nature and content of education to support to development of young children's literacy as a means to better prepare them for learning to read is similarly contentious. In a recent debate, the only point of agreement between Whitehurst (2001) and Elkind (2001) was that there is a lack of rigorous research on the relative effectiveness of preschool curricula that would help to resolve these issues. Few studies of curricula have employed random assignment, and non-experimental studies commonly confound curriculum differences with other preschool program characteristics and characteristics of the children attending the programs (Bowman et al., 2001). The U.S. Department of Education funded seven randomized trials to study preschool curriculum in 2002 and another six in 2003. However, few studies using randomized trials to evaluate the effects of curricula have actually been published in recent years (e.g., Assel, Landry, Swank, & Gunnewig, 2007; Domitrovich, Cortes, & Greenberg, 2007).

Among the more noteworthy studies on curriculum are randomized, small-scaled trials, some dating to the 1960s and 1970s, comparing the effects of well-specified alternative models on children's learning and development with long-term follow-up. These studies compared: Direct Instruction, High/Scope's open framework model, and a traditional unit-based approach (Schweinhart & Weikart, 1997); Montessori, Direct Instruction, DARCEE, and traditional nursery school (Miller & Bizzell, 1983; Miller & Dyer, 1975); Montessori, Direct Instruction, traditional nursery school, and the Community Integrated Program (Karnes, Shwedel, & Williams, 1983); and, Direct Instruction and Mediated Learning (Mills, Cole, Jenkins, & Dale, 2002).

Taken together these studies yielded several conclusions. First, differences in curricular emphases tend to be reflected in immediate differences in children's learning. The Direct Instruction models produced larger gains on achievement tests of subject matter specific content, for example. Second, the differences in cognitive outcomes did not persist more than a few years after leaving the program. Third, there was some evidence that curriculum effects varied with child characteristics, specifically gender and level of ability at program entry. However, these last findings emerge from post hoc analyses rather than tests of hypotheses specified prior to analysis. As these findings did not appear to replicate across studies, these apparent interactions may simply be due to random variation. Finally, differences in curricula also produce differences in social and emotional development. These differences in social and emotional outcomes may be more persistent than differences in cognitive outcomes.

Long-term follow-up of High/Scope's curriculum comparison study has received much attention because it found that the Direct Instruction curriculum produced far worse outcomes for social and emotional development over a long period of time (Schweinhart, Weikart, & Larner, 1986). The children in the Direct Instruction model had less pro-social behavior and

more anti-social behavior. In follow-up at age 23, the Direct Instruction model apparently failed to produce the decreases in crime and delinquency that had been found earlier in the Perry Preschool study, in contrast to the curriculum models that allowed more child-initiated activity (Schweinhart & Weikart, 1997). These results have been hotly contested (Bereiter, 1986; Gersten, 1986; Schweinhart et al., 1986).

Follow-up through age 15 of another randomized trial found no differences in anti-social behavior and delinquency between Direct Instruction and Mediated Learning, which is a more cognitively oriented model with more child initiation (Mills et al., 2002). This study would appear to contradict the conclusion that direct instruction negatively affects social and emotional development. However, comparison of results suggests that neither of the models in this study improved social behavior to the same extent as the High/Scope model, which specifically focuses on social problem solving and planning. Mills et al. (2002) suggest that gender differences across groups might actually account for this finding in the High/Scope study, but analyses demonstrate that this is not the case (Schweinhart & Weikart, 1997).

The Planned Variation Head Start study, in which curriculum comparisons were made on a large scale (6000 children at 37 sites) is another important source of evidence (Datta, McHale, & Mitchell, 1976). Although it did not employ random assignment, this study did introduce new curriculum models with training to increase fidelity of implementation and controlled conditions across models in that all models were implemented within the relatively homogeneous Head Start program. Again, curriculum models were found to produce results consistent with their curricular emphases. Direct Instruction was found to produce the largest achievement gains, and High/Scope's cognitively oriented model produced the largest gains on IQ tests, though none of these relative advantages were sustained as children moved into elementary school.

A similar and even larger study of 20 curriculum models in elementary school, Project Follow Through, was conducted at about the same time as the Planned Variation study. It also found that the Direct Instruction approach produced the largest achievement test gains. However, for these older children the gains appeared to be somewhat more persistent. Moreover, it appeared that the Direct Instruction model also produced more positive effects on some aspects of social and emotional development compared to other models. These results and their interpretation have been subject to considerable dispute (House, Glass, McLean, & Walker, 1978; Mac Iver & Kemper, 2002; St. Pierre, Anderson, Proper, & Stebbins, 1978; Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977). While not directly applicable to preschool age children, it has helped to fuel the debates about preschool curriculum.

Given this history it is hardly surprising that much of the research on preschool curriculum that followed compared more child-centered to more didactic practices. Most of these studies have investigated the association between natural variation in measured differences or beliefs and practices and children's achievement and social-emotional development (Hirsh-Pasek, Hyson, & Rescorla, 1990; Marcon, 1999, 2002; Stipek, Daniels, Galluzzo, & Milburn, 1992; Stipek, Daniels, Galluzzo, Millburn, & Salmon, 1998; Stipek, Feiler, Daniels, & Millburn, 1995). Some of these studies have suggested that less didactic methods in preschool result in larger long-term achievement gains and others have added to concerns that more didactic practices with a greater emphasis on academic content are less optimal for children's social and emotional development. Concerns regarding social and emotional development have been intensified by studies finding an association between participation in early child care and children's behavior problems (Belsky et al., 2007; Vandell, 2004).

3. Description of the Tools of the Mind curriculum

The Tools of the Mind (*Tools*) curriculum was developed by Bodrova and Leong (1996), based on the theories and practical insights on cognitive development of Luria (1966) and Vygotsky (1978), including the promotion of self-regulation through a comprehensive system of activities. Basic principles of the curriculum include: (1) children construct their own knowledge; (2) development cannot be separated from its social context; (3) learning can lead development; and (4) language plays a central role in mental development (Bodrova & Leong, 2007). Guided by these principles, *Tools* has two primary emphases. First, the curriculum focuses on broad foundational skills, including children's abilities to regulate their own social and cognitive behaviors, to attend and to remember on purpose, the use of symbolic representation, and early math skills (Leong & Hensen, 2003). Second, there is at the same time an emphasis on specific literacy prerequisites for reading and writing (e.g., oral language, phonemic awareness, knowledge of letters, and familiarity with the conventions of print) and on specific mathematics pre-requisites (e.g., counting meaningfully, one-to-one correspondence, patterns, numeral recognition, etc.) (Bodrova & Leong, 2007). The activities promoting these literacy and math pre-requisites have a self-regulatory component built into them.

As the name suggests, *Tools* guides teachers' daily practices to support children's acquisition and development of various psychological "tools." Psychological tools are cultural-based, symbolic artifacts, such as symbols, texts, or graphic organizers, that, when internalized, help individuals to master their own psychological functions, including perception, memory, and attention (Kozulin, 2003). The *Tools* curriculum incorporates 40 Vygotsky-inspired activities designed to promote mature dramatic play, encourage the use of self-regulatory private speech, and teach the use of external aids to facilitate attention and memory (Luria, 1965, 1979; Vygotsky, 1962, 1978, 1997). As self-regulation is considered learned behavior, each specific learning activity is designed to teach self-regulation. Play is viewed as the primary source of self-regulation as well as leading children to higher levels of development. Thus, *Tools* teachers do not simply "let" children play, but use a play planning process as well as specific interactions to actively support children's development of "mature" play in which pretend scenarios are complex, planned, sustained, and involve multiple roles. Other defining characteristics of *Tools* are Scaffolded Writing

(Bodrova & Leong, 1995), directions for oral language use between teachers and children, and movement activities that incorporate self-regulation and symbol use.

Dramatic play is a leading mechanism for the development of self-regulation, and *Tools* is similar to other constructivist and play-based curricula in room arrangement, materials, and the balance between small group, large group, and play activities. However, *Tools* differs from other curricular models in that the teacher actively supports the development of mature intentional dramatic play, while ensuring that each child is active in the selection of activities and the creation of the play scenarios (Bodrova & Leong, 2007). In *Tools,* the teacher helps children to write play plans, teaches children to plan their dramatic play together, and helps children to think about next steps during their play with the intention of fostering the development of self-regulation. Dramatic play contributes to the development of children's self-regulation because it is an imaginary situation governed by social rules. While pretending to be a store clerk, the child must seek to behave in ways that meet the social rules for that role, curbing immediate impulses in order to think about how to represent how a clerk would act. Dramatic play leads to the internalization of rules and expectations and places demands and constraints on the child's behavior. In thinking about the role while talking and acting the child is relying on private speech, while play plans provide an external support and children remind each other of the rules and roles thereby helping each other regulate their behavior. In addition, the extensive use of symbols (e.g., pretending a jar lid is the cat's ball) develops abstract thought more generally.

It may be that the strongest difference between the *Tools* curriculum and others is the extent to which it directly addresses the idea that learning traditional academic content can be inefficient or difficult if children lack underlying cognitive skills such as self-regulation, but at the same time, it also can provide a context in which self-regulation can be practiced if it is organized with specific structural features (use of private speech, mediation, and specific teacher/peer interactions). All activities within the curriculum are designed to promote the development of such underlying skills along with more academic subject matter. The sheer number of activities designed to foster self-regulation, as well as the details of those activities (e.g., the development of children's play plans in concert with the teacher through Scaffolded Writing, turn-taking in Buddy Reading facilitated by concrete external aids, and the Pattern Movement game designed to increase the use of private speech and practice its use in self-regulation) distinguishes *Tools*. Finally, even though play is viewed as the leading activity for developing underlying foundational skills, and the curriculum, *Tools* makes more use of direct instruction than many other constructivist curricula.

In *Tools*, the teacher's role is not just to teach skills or facts, but to help children use tools and to learn to develop tools to facilitate learning. One of the tools expressly taught is the use of external mediators to regulate behavior. In "Buddy Reading" children are provided with books and take turns reading to each other (telling the story, turning pages, pointing to pictures, etc.). One child is given a picture of a mouth and the other member of the pair is given a picture of an ear. The child with the mouth picks a book and reads to the other. The child with the ear listens and waits for his or her turn. Then they switch pictures and roles. The pictures help the children to remember their roles and regulate their behavior. Several months into the year, the pictures are no longer necessary as the children move beyond the need for an external reminder.

The "Freeze Game" is another activity specifically designed to develop self-regulation and other abilities. During the Freeze Game, children practice self-regulation and symbolic representation. During circle time, the teacher plays music to which she and the children dance. While dancing the teacher holds a picture of stick figure representing a specific pose that children will take when the music stops. The children must control their own behavior by taking specific actions at specific times. This requires inhibition, not getting into the pose before the music stops, as well as holding the particular pose when the music does stop. By regulating their behavior, dancing and not dancing, posing and not posing, on purpose children developing their ability to regulate emotions and cognition. They are also learning to interpret symbols as they translate an iconic representation of a body position into an actual body position.

4. Description of the control curriculum

The control curriculum was developed by the local school district teachers and administration during the 3 years prior to the study. The curriculum developed by the school district was based on the idea that literacy should be taught to young children in a balanced way (i.e., through a combination of reading, writing, and listening activities) and in the context of thematic units, such as 'family' or 'transportation.' This focus on literacy is a characteristic it shared with *Tools*. In structured observations (discussed below) of the control group, frequently observed activities were art projects that correlated with the "letter of the week," free play for 30–60 min a day, large group movement and music, and other large group activities such as story-time. This curriculum represents common practices in many preschool programs that construct their own curriculum or adopt eclectic practices (Clifford et al., 2005). Although the control curriculum covered much of the same "academic" content and topics as *Tools*, it differed in educational philosophy and it did not weave activities promoting self-regulation throughout the day. In control classrooms, there was greater emphasis on teacher-imposed control and less on children regulating each other and themselves.

5. Method

5.1. Research site

The school selected for this research study is located in an urban school district in New Jersey and is part of the state-financed full-day "Abbott" preschool education program. Abbott programs are open to all children at ages three and four in

31 of New Jersey's high poverty school districts. Median income for families living in the city was $34,935 according to data from the 2000 U.S. census. Data reported by the district to the state Department of Education indicate that over 80% of the district's public school students qualified for free or reduced price lunch, and 70% came from homes where English is not the primary language.

The structural characteristics of preschool education in all classrooms are set by state regulations. Each classroom must have a teacher with a BA and certification in early childhood education who works together with an aide. Class size is limited to 15 students. Support includes mentor teachers to coach teaching staff. Guidance is provided by detailed state standards for children's learning in the preschool program. All classrooms operate for at least 6 h per day, 180 days per year with wrap-around care offered for an extended day and year to meet the child care needs of parents. The children's school year began right after Labor Day and continued through mid-June. In this district, 3- and 4-year olds were served in mixed age group classrooms.

5.2. Participants

District administrators offered the research team seven classrooms on one floor of the school for implementation of the *Tools* curriculum. Eleven classrooms on another floor were available for the control condition. Teachers and assistants were randomly assigned to the classrooms for fall of 2002. The district also provided the investigators with a list of children registered for preschool education ($n = 274$) at the research site, and these children were randomly assigned to classrooms before the start of school, as well.

Given the small number of teachers, random assignment by blocks was used to ensure that treatment was not confounded with key teacher characteristics. Specifically, teachers were stratified into four groups: eight teachers with a preschool-grade three license (P-3); four teachers with a K-8 license who were grandfathered (requires two full years of preschool teaching experience for preschool validation); four teachers with a N-8 license (an older certificate that preceded the P-3); and two teachers transferred from another school within the district (unlike the others, they did not choose to work at this school but were relocated there by the district). Two teachers from each of the first three groups and one teacher from the last were randomly chosen for the *Tools* classrooms.

Efforts were made to control treatment diffusion (Cook & Shadish, 1994) between the two groups of teachers by placing all the treatment classes on one floor of the building and all the control classes on another floor. This restriction of each curriculum to separate floors of the building resulted in unequal numbers of classrooms in the treatment and control groups. The smaller floor was used for *Tools* because it meant lower costs for training and technical assistance from outside the district. In addition, treatment group teachers signed documents in which they pledged not to divulge the practices they were learning to other teachers while the study was underway.

All classrooms looked very similar physically before the *Tools* curriculum intervention. For example, the school district ordered all classrooms exactly the same amount and type of furniture, toys, art supplies and books. Any incidental differences in the classroom materials were due to the teacher bringing in her own supplies from home.

There were a few materials specific to the *Tools* curriculum that were necessary for the curriculum implementation that the school district had not provided (in particular, wipe off boards and erasable markers) or had not provided in sufficient quantity (reams of white paper for daily planning). These materials were purchased for the treatment classrooms by the researchers and an equal amount of money was provided to control group teachers for purchase of similar amounts of educational materials at their discretion (total amount spent was $300 per class).

A total of 274 three- and four-year-old children were randomly assigned to either *Tools* curriculum classrooms ($n = 106$) or district curriculum classrooms ($n = 168$). At the open house in September two researchers (one native Spanish speaker, one native English speaker) were present to explain the study to parents and answer questions. As an incentive to join the study parents were offered $20 worth of children's books. Of the group of 274 randomly assigned children, 224 families (82%) consented to have their child participate in testing for the study. Lack of consent was higher for the control group ($n = 40$, 24%) than for the *Tools* group ($n = 10$, 9%). Teachers were asked to encourage families to return the permission letters, and *Tools* teachers may have been more enthusiastic about the study because they represented the "alternative" to the district's usual curriculum. In addition, three families requested that their children be transferred from *Tools* to the district curriculum and one family requested the reverse, though we do not know the reasons for the transfers. These four children (2% of the sample) did not participate in the study.

Among those who consented to the study, attrition was relatively minor. One child in each group moved out of the district prior to assessment. This left us with an initial sample of 218 children: 92 (42%) in *Tools* and 126 (58%) in the control group. Of these, four in each group were not tested in the Fall, due to the child's absence or discomfort with the testing situation. By Spring post-test, another six children in the *Tools* group and five children in the control group had moved. One child in each group was not tested due to absences so that 85 *Tools* (92%) and 120 control (95%) children were assessed in the Spring.

It was not possible to conduct extensive analyses of attrition, because most attrition in this study was due to lack of active consent from parents prior to any data collection. However, we do know gender, ethnicity, and home language for most of the original sample children. Thus, it was possible to test for differences between those whose parents agreed to participate and those whose parents declined or did not respond. Analysis of Variance revealed no statistically significant main effects of attrition or interactions between attrition and treatment (curriculum assignment).

5.3. Teacher training

All the teachers chosen to participate in the study freely agreed to do so and signed informed consent. For their participation, teachers were given a $100 honorarium plus $300 worth of educational supplies for their classrooms. The two groups of teachers received similar amounts of in-service training so that differences in this may be ruled out as an explanation for group differences. Teachers who were assigned to *Tools* received 4 full days of curriculum training before classes began in the 2002–2003 school year. Control group teachers attended workshops on the already established district curriculum given by the district for the same amount of time.

During the school year *Tools* teachers received 30 min classroom visits approximately once a week from a *Tools* trainer to address any difficulties they were having with the curriculum. If a teacher expressed a specific concern the *Tools* trainer would then schedule an appointment to come and spend more time in that teacher's classroom. In addition to classroom visits, *Tools* teachers received one half-day workshop and five 1-h lunchtime meetings to discuss aspects of the curriculum.

Control group teachers received similar support in implementing the pre-existing district curriculum from "master teachers." Master teachers are the district's specialized preschool professional development staff, and they oversee curriculum implementation and assist teachers with improving their teaching, including overcoming any problems they may encounter. During the year the master teachers made periodic visits to each control group teacher and provided a twice-monthly series of after-school or lunchtime workshops on various topics related to the district curriculum.

5.4. Fidelity of Tools implementation

Full implementation of the *Tools* curriculum requires specific materials and procedures. In addition to more general observation measures discussed later, highly specific fidelity measures were developed to assess the extent to which the study classrooms provided required materials and followed distinctive program procedures. For example, socio-dramatic play is considered to lead cognitive development during the preschool period. Thus, the curriculum calls for the use of toy figures (animal or human) in *each* of the learning centers in the classroom to support children's socio-dramatic play. To assess fidelity of implementation regarding materials, a 50-item "environmental features" observation instrument was developed. The scale measures the presence of materials for seven features: opening activity (circle time); play; equipment; classroom labeling; books; writing; and socio-dramatic play. A criterion of 80% was established as "full" implementation for each feature. Reliability scores of .85–.96 were obtained prior to collecting data from the study classrooms. At the beginning of the year, the *Tools* classrooms fully implemented three of the seven key environmental features. At years end, *Tools* classrooms fully implemented all environmental features except writing. Control classrooms did not fully implement any of the key environmental features at any time.

Assessment of fidelity regarding procedures focused on one particular activity that is common to many curricula, large group morning meeting (circle time). *Tools* sets aside a limited amount of time, 8–10 min, for large group morning meeting and is very specific regarding how it is to be implemented. *Tools* teachers are expected to pose questions to the *group* of children, rather than to specific individuals. Fidelity of implementation in this activity was examined using video tapes of large group meetings collected midyear. We found that on average large group sessions lasted 9 min in *Tools* classrooms and 19 min in control classrooms. This indicated fidelity to the *Tools* model and a clear difference between treatment and control conditions. In addition, *Tools* teachers are expected to pose questions to the entire group simultaneously during rather than to individual children during large group time. Thus, fidelity of teacher–child interaction was assessed by coding the video-taped teacher and child large group time interactions for instances in which teachers posed questions to the group as "group talk" and questions to individual children as "individual talk." Coders were naïve with respect to the study's focus and which classrooms were in the treatment or control conditions. Reliability scores ranged from .90 to .95. *Tools* teachers, but not control group teachers, were found to direct significantly more "group talk" than "individual talk."

5.5. Description of participating children and families

All children in the sample were age three or four at the time of pre-test administration, and they were 47% females and 53% males. The sample was composed of slightly more 4-year olds (54%) than 3-year olds reflecting overall participation rates in the program. Parent questionnaires were administered at study entry by telephone to obtain more information about the children and their family background including ethnicity, maternal employment, and income. No significant differences were found between *Tools* and control group children on these characteristics or on prior child care attendance, frequency of being read to in the home, and parent-reported knowledge of numbers, letters, and colors. Of those parents responding, 69% reported Spanish as their primary home language, while 30% reported that they use primarily English. The overwhelming majority (93%) described their children as Hispanic or Latino when asked about ethnicity. These statistics are consistent with the composition of the school district as a whole. Table 1 displays the descriptive characteristics of the sample, and compares the groups as initially assigned and as eventually tested.

5.6. Child assessment procedures

Children were assessed in the fall (October and November) and spring (late April through early June) of the 2002–2003 school year. Assessments were conducted one-to-one in a quiet section of the child's school environment and were scheduled

Table 1
Characteristics of students by treatment groups

	Total		Treatment		Control		Chi-square *p*
	n	Percent	*n*	Percent	*n*	Percent	
Gender							
Female	129	47.1	49	46.2	80	47.6	
Male	145	52.9	57	53.8	88	52.4	.370
Ethnicity							
Latino/Hispanic	249	92.6	98	92.5	151	92.6	.215
African-American	6	2.2	4	3.8	2	1.2	
Asian	10	3.7	4	3.5	6	3.7	
Multi-racial	4	1.5	0	0	4	2.5	
Primary home language							
Spanish	164	68.9	72	71.3	92	67.2	.489
English	71	29.8	27	26.7	44	32.1	
Another language	3	1.3	2	2.0	1	0.7	

to ensure that they did not disrupt children's school routines (i.e., avoided meals, nap time and outdoor play) seeking to make children as comfortable as possible during the process. Data collectors were graduate students in education and experienced researchers with advanced degrees. They were trained to conduct the child assessments and then shadow scored until they reached 100% agreement with the site coordinator. Throughout the data collection period the site coordinator monitored the testing to ensure administration remained reliable.

Children were tested in Spanish or English. The child's dominant language was ascertained from the classroom teacher who made judgments based on language proficiency tests administered by the schools at the beginning of the year, supplemented by parental report and their own experiences with the children. Assessors spoke to the child only in the language of each assessment to avoid code switching during testing sessions. Bilingual native Spanish speakers were employed to conduct assessments in Spanish.

Children were assessed with six different instruments, split between two testing sessions, to avoid extending the testing time per session beyond 30 min for any child. Children were first assessed with the Woodcock–Johnson Applied Math Problems and Letter–Word Identification Tests, Get Ready to Read, and the Wechsler Preschool Primary Scale of Intelligence Animal Pegs subtest. In the second testing session, we administered the Peabody Picture Vocabulary Test-III (PPVT-III), Expressive One-Word Picture Vocabulary Test (EOWPVT), and the Oral Language Proficiency Test (administered only to Spanish speakers). The Social Skills Rating System (SSRS) was completed by teachers at the end of the school year and did not require the child's participation. Each measure is described below.

5.7. Child measures

The PPVT-III was administered to all children regardless of their home language to obtain a measure of standard English vocabulary development. The PPVT-III (Dunn & Dunn, 1997) is a 204-item test of receptive vocabulary in standard English. The test is administered by having children point to one of four pictures shown when given a word to identify. The PPVT-III is often used as a quick indicator of general cognitive ability, and it correlates reasonably well with other measures of linguistic and cognitive development related to school success. The PPVT-III has a mean standard score of 100 and a standard deviation of 15. The PPVT-III was nationally standardized on a stratified normative sample of 2000 children and adolescents and has an internal consistency reported as Spearman–Brown split-half reliability coefficients ranging from .92 to .98. Test–retest reliability for a 1-month interval in four different age groups ranged from .91 to .93.

The Woodcock–Johnson Psycho-Educational Battery-Revised (WJ-R) and the Bateria Psico-Educativa Revisada de Woodcock–Muñoz (WM-R) are comprehensive sets of individually administered tests of cognitive abilities and achievement (Woodcock & Johnson, 1989; Woodcock & Munoz-Sandoval, 1996). We administered the Letter–Word Identification and Applied Problems subtests from these batteries. The 76-item Letter–Word Identification (Identification de Letras y Palabras) is a measure of reading decoding, which asks children to identify printed letters and words. The 60-item Applied Problems (Problemas Aplicados) subtest measures math skills.

The WJ-R and WM-R tests were calibrated and equated to U.S. norms through Rasch modeling and are particularly well-suited to the needs of assessment with bilingual populations (Woodcock & Munoz-Sandoval, 1996). The tests' standard scores have a mean of 100 and a standard deviation of 15. The English Form of the subtests was normed on a stratified random sample of 6359 English-speaking subjects in the United States. The Spanish Form was normed on 3911 primarily monolingual Spanish-speaking subjects from samples obtained both inside and outside the United States. Internal consistency reliabilities range from the high .70s to low .90s on both subtests for preschool-aged children. Correlations of the WJ-R and WM-R with other tests of cognitive ability and achievement are reported to range from .60 to .70.

Get Ready to Read (GRTR) was developed as a screening tool to assess a 4-year-old's progress in developing early literacy skills (Whitehurst & Lonigan, 2001). It was developed using a sample of 342 children from two different locations, ages 48–59

months. GRTR has reported internal consistency (alpha coefficient) of .78 and split-half reliability of .80. This tool has been found to correlate with other measures of language and letter knowledge (Whitehurst & Lonigan, 2001).

The Wechsler Preschool Primary scale of Intelligence Animal Pegs subtest (WPPSI) measures a child's nonverbal problem solving and visual-motor proficiency, particularly accuracy, concentration and speed of performance as they place pegs of correct colors in a series of holes under pictures of animals (Wechsler, 1989). The WPPSI performance IQ has adequate test–retest reliability (.87) and internal consistency reliabilities (.92). Studies have established adequate construct, concurrent, and predictive validity for diverse populations (Wechsler, 1989).

The EOWPVT-Revised is a norm-referenced assessment of expressive vocabulary in standard English designed for use from ages 2 through 18 (Brownell, 2000). The measure taps a child's ability to use words, requiring a child to access and retrieve words from memory, going beyond tests of receptive vocabulary. Normed on a randomly selected sample of 2327 children, the internal consistency of the EOWPVT is .98. To obtain a baseline of standard English expressive vocabulary, this assessment was administered to all children regardless of their home language.

The IDEA Oral Language Proficiency Test (OLPT) assesses the receptive and expressive language skills of Spanish-speaking children (Ballard & Tighe, 1999). Thus, it was administered to about 70% of the students in our sample to assess their Spanish language development. Scores reflect students' responses to items representative of common Spanish language speech patterns. The internal consistency reliability of the measure (Chronbach's alpha) is .97, and test–retest reliability is reported as .63 (Ballard & Tighe, 1999).

The teacher form of the Problem Behaviors Scale of the SSRS was completed by the child's teacher near the end of the school year (Gresham & Elliot, 1990). The SSRS was standardized on a nationally representative sample of 4170 children. The Problem Behavior Scale measures both externalizing and internalizing behaviors. Teacher forms were particularly reliable with an internal consistency coefficient and test–retest correlations ranging from .82 to .95.

5.8. Classroom assessment procedures

To assess the extent to which the *Tools* curriculum created the expected changes in children's experiences beyond those documented by the treatment fidelity measures, multiple instruments were used to examine various aspects of the classroom environment. These instruments are described individually in the next section of the paper. Our research team conducted training on the three observation instruments that were administered and scored by our staff. Observers typically had advanced degrees and experience teaching at the preschool level. Each observer was shadow scored and reached an 85% inter-rater reliability rate before qualifying to conduct observations for the study. Each classroom was observed once early in the second semester. The fourth instrument was coded and analyzed by the developers based on classroom video tapes.

Our team video-taped classrooms and another team that was completely blind to treatment or control status coded the video tapes using for the Classroom Assessment Scoring System (CLASS) (Pianta et al., 2005). Video was taken of an entire morning in each class. The data collector who shot the video was instructed to focus on the head teacher and her interactions with the class from breakfast until lunch. During free play, the data collector maintained her focus on the head teacher and her interactions with individual children by moving around the room as needed. In almost all cases the data collector was able to capture the teacher having opening group, playtime, large group and some sort of small group experience. All video was then shipped to the instrument's authors for coding and scoring.

5.9. Classroom measures

The Early Childhood Environmental Rating Scale-Revised (ECERS-R) provides a global look at classroom quality with 43 indicators ranging from safety, teacher–child interaction to parent involvement (Harms, Clifford, & Cryer, 1998). The ECERS-R has been used as a tool to measure preschool classroom quality in numerous studies within the United States and abroad (Harms et al., 1998). The ECERS-R scale uses a scoring system where one is considered inadequate quality, three is minimal quality, five is good quality and seven is excellent quality. In the standardization process of this measure, an extensive set of field tests of the ECERS-R was conducted in 45 classrooms. Substantial revisions were made to the first field-test draft of the scale, and then followed by a second test of 21 classrooms which focused on inter-rater reliability. The total scale internal consistency for ECERS-R is .92, and subscale internal consistencies range from .71 to .88.

The Supports for Early Literacy Assessment (SELA) was used to examine the quality of the literacy environment and instruction (Smith, Davidson, Weisenfeld, & Katsaros, 2001). The SELA was developed based on research and professional opinion regarding best practices with regard to the development of young children's literacy. Scores on the SELA range from one through five, with one considered very low quality, three fair or mediocre quality, and five ideal quality.

The Preschool Classroom Implementation (PCI) rating scale (Frede, 1989) measures the frequency of use of scaffolding techniques by teachers in their interactions with children. Items are scored according to teachers' typical behavior throughout the day, with justification for each rating including examples noted by the assessor. Scores on the scale range from one (indicating that the technique was not observed) to five (indicating that the technique was used consistently throughout the day). Examples of items include: "Staff extend children's activities and problem solving by playing alongside children for a few minutes to model new possibilities;" and, "Adults make specific comments that extend children's thinking and focus on cognitive concepts such as classification, seriation, time, and space." The scale employed here is a shortened version of a more comprehensive instrument. The full instrument has been found to predict the effects of preschool education on

children's learning into first grade (Frede, Austin, & Lindauer, 1993; Frede & Barnett, 1992). Analyses of the subscale used in this study found that it correlated with the ECERS-R at .6 ($p < .001$). Inter-rater reliability exceeded 80% and Cronbach's alpha was .89.

The CLASS measures emotional climate, classroom management, and instruction (Pianta et al., 2005). Unlike the ECERS-R, the CLASS does not assess physical or structural features of the classroom. Among the dimensions scored are the sensitivity of teacher behavior to the children, behavior management, effectiveness of behavior management, productive management of time and activities, and the quality of instruction and feedback, and the extent to which activities stimulate conceptual development and engagement (LaParo, Pianta, & Stuhlman, 2004). Each item within the instrument is rated on a seven-point Likert-type scale. The coding and analyses for the CLASS instrument was done by the instrument developers who were blind to treatment or control status of the classrooms in the video tapes.

6. Results

We present results in two parts. The first reports findings from analyses of ratings of observed classroom activities and environment, including the estimated effects of curriculum. The second reports estimated effects of the *Tools* curriculum model on children's learning and development. In the statistical analyses, *Tools* is coded 1 and the control condition is coded 0.

6.1. Analyses of classroom observation data

Table 2 presents the results of comparisons of the *Tools* and control classrooms on three observation instruments: the ECERS-R, SELA, and PCI. The *Tools* classrooms scored substantially higher than control classrooms on total scores for all three measures: 3.9 versus 3.1 on the ECERS-R, 3.0 versus 2.0 on the SELA, and 2.5 versus 1.6 on the PCI. All of these differences are highly statistically significant ($p < .01$). These differences correspond to effect sizes of about 2. Having found differences on the total scores, we then conducted analyses on the ECERS-R subscales to look for more specific areas of differences. The *Tools* classrooms scored significantly higher on the three ECERS-R subscales most closely related to curriculum: Language and Reasoning ($p = .01$), Activities ($p = .004$), and Interactions ($p = .08$). *Tools* classrooms reached or exceeded an average score of five on both the Language–Reasoning and Interactions subscales.

As there is no summary score for the CLASS, we conducted a factor analysis on the CLASS subscales. This yielded only one factor that loaded on multiple items, which were positive classroom climate, teacher sensitivity, behavior management techniques, and productivity. As the statistical power to detect even large differences is quite limited (data were collected from only 16 classrooms on this measure), we note differences that are significant at the .10 level, but not at the conventional .05 level (two-tailed). A *t*-test suggests a possible difference between the curricula on the CLASS factor ($p = .096$). Comparisons on individual items found the *Tools* classrooms to score significantly higher than the control classrooms on the productivity item ($p = .04$), with indications of a possible difference on the teacher sensitivity item ($p = .07$). Productivity measures teacher management of instructional time and routines, and teacher sensitivity measures responsiveness of the teacher to children's academic and emotional needs and the degree to which the teacher provides a secure base for children to volunteer answers and responses (Table 3).

6.2. Analyses of child assessment data

Two-sample *t*-tests presented in Table 4 reveal that there were no significant differences at the beginning of the study year on any of the assessments of children's learning and development. This is consistent with the lack of any family background differences between *Tools* and control groups noted earlier. This provides reassurance that randomization worked to ensure comparability of groups despite attrition.

Table 2
Comparisons of treatment (Tools) and control Classrooms on ECERS-R, SELA, PCI

	Treatment, $n = 7$		Control, $n = 11$		p^a
	M	S.D.	*M*	S.D.	
Overall ECERS score	3.9	0.3	3.1	0.5	.003
Space and furnishings	3.4	0.5	3.0	0.7	.248
Personal care subscale	3.6	1.2	2.7	1.1	.132
Language–Reasoning	5.0	0.9	3.4	1.3	.010
Activities	3.7	0.7	2.7	0.7	.004
Interactions	5.7	0.9	4.5	1.7	.081
Program	2.0	0.7	2.2	0.3	.427
Parents and staff	3.7	0.3	3.5	0.5	.232
Overall SELA score	3.0	0.5	2.0	0.5	.001
Overall PCI score	2.5	0.6	1.6	0.4	.002

ECERS, the Early Childhood Environmental Rating Scale-Revised; SELA, Supports for Early Literacy Assessment; PCI, Preschool Classroom Implementation.
[a] Two-tailed *t*-test.

Table 3
Comparison of treatment (Tools) and control classrooms on the CLASS

	Treatment, $n = 7$		Control, $n = 9$		p^a
	M	S.D.	*M*	S.D.	
Positive classroom climate	5.6	1.1	4.8	1.2	.201
Negative classroom climate	1.6	0.5	1.9	1.3	.547
Teacher sensitivity to children	5.4	1.1	4.3	1.1	.074
Over controlling environment	1.7	0.5	1.6	0.5	.547
Behavior management techniques	5.4	1.0	4.6	1.7	.240
Productivity	6.1	0.7	5.1	1.1	.042
Concept development	2.6	1.1	2.3	0.5	.580
Learning formats/engagement of children	5.0	1.3	5.4	0.7	.436
Quality of teacher feedback	2.7	1.1	2.1	0.9	.256

[a] Two-tailed *t*-test.

Table 4
Comparisons of pre-test scores by treatment group (Tools versus control)

	Treatment (Tools)			Control			p^a
	n	*M*	S.D.	*n*	*M*	S.D.	
PPVT-III	88	63.8	20.8	122	67.0	15.8	.211
EOWPVT-R	88	63.6	13.4	120	64.0	12.4	.813
WJ-R AP	88	75.7	16.4	122	76.9	14.9	.565
WJ-R LW	88	89.0	11.6	122	88.7	10.1	.832
WIPPSI (raw score)	87	23.0	15.4	122	20.7	14.6	.276
OLPT (raw score)	59	7.3	5.4	82	7.1	5.6	.813

PPVT-III, Peabody Picture Vocabulary Test; EOWPVT-R, Expressive One-Word Picture Vocabulary Test-Revised; WJ-R AP, Woodcock–Johnson Applied Problems; WJ-R LW, Woodcock–Johnson Letter–Word Identification; WIPPSI, Wechsler Preschool Primary Scale of Intelligence Animal Pegs subtest; OLPT, IDEA Oral Language Proficiency Test.

[a] Two-tailed *t*-test.

Two sets of analyses were conducted on child outcomes. Each has its strengths and limitations. Effects on child outcomes were estimated through regression analysis with post-test as the dependent variable and pre-test, curriculum, and the child's primary language as independent variables. Standard scores were employed in the analyses when available. All standardized measures employed have a mean of 100 and a standard deviation of 15. Standard scores are not available for GRTR, the WPPSI Animal Pegs subtest, and the OLPT. Hence, raw scores were used in those regressions, and age was added as an independent variable (no significant interactions were found between treatment and age). The SSRS and GRTR were administered only at post-test. Tables 5 and 6 report the results of regression analyses conducted to estimate curriculum effects on children's learning and development. Statistically significant effects of curriculum were found on the SSRS and on the PPVT-III and OLPT. The effect on behavior problems as measured by the SSRS is about half a standard deviation (es = .47, Glass' delta). This indicates that behavior problems (as rated by the teacher) were substantially less common for children in the *Tools* classrooms than for those in the control classrooms. In further analyses, reductions in both externalizing and internalizing problems were found to contribute to this effect. The two effects on language development are of moderate size. Estimated effect sizes are .22 for the PPVT-III and .35 for the OLPT. Those estimated effects indicate that *Tools* was more effective than the control curriculum in promoting both English and Spanish language development.

Given the multi-level nature of the data, we also estimated hierarchical linear models with treatment at the classroom level. These analyses are more conservative as they recognize that children are clustered within classrooms. Both methods

Table 5
Regressions estimating effects of tools on learning and development (standard score measures)

	SSRS		PPVT-III		EOWPVT-R		WJ-R AP		WJ-R LW	
	B	S.E.	*B*	S.E.	*B*	S.E.	*B*	S.E.	*B*	S.E.
Constant (intercept)	120.9*	7.43	31.59*	3.56	11.31*	2.81	35.41*	5.40	35.97*	6.30
Tools versus control	−7.31*	1.92	3.72*	1.79	1.39	1.06	2.80	2.10	−1.26	1.45
Pre-test score			0.60*	0.05	0.86*	0.04	0.590*	0.07	0.63*	0.07
Primary language (English = 1, other = 0)			6.51*	1.99						
Age	−0.38*	0.13								
Model R^2	0.109		0.474		0.670		0.267		0.277	
n	198		209		204		213		213	
Effect size	−.47		.22		.11		.15		−.10	

PPVT-III, Peabody Picture Vocabulary Test; EOWPVT-R, Expressive One-Word Picture Vocabulary Test-Revised; WJ-R AP, Woodcock–Johnson Applied Problems; WJ-R LW, Woodcock–Johnson Letter–Word Identification; SSRS, Problem Behaviors Scale of the Social Skills Rating System.

* $p < .05$.

Table 6
Regressions estimating effects of tools on learning and development (raw score measures)

	GRTR		WIPPSI		OLPT	
	B	S.E.	*B*	S.E.	*B*	S.E.
Constant (intercept)	−3.61*	1.87	−14.37*	6.70	−14.32*	3.86
Tools versus control	0.19	0.49	1.02	1.63	2.27*	1.00
Pre-test score			0.53*	0.06	0.61*	0.10
Age	0.23*	0.03	0.64*	0.13	0.35*	0.07
Model R^2	0.182		0.481		0.433	
n	220		211		140	
Effect size	.05		.06		.35	

GRTR, Get Ready to Read; WIPPSI, Wechsler Preschool Primary Scale of Intelligence Animal Pegs subtest; OLPT, IDEA Oral Language Proficiency Test.
* $p < .05$.

Table 7
HLM analysis estimating effects of tools on learning and development (standard score measures)

Variable	SSRS (*n* = 198)		PPVT-III (*n* = 209)		EOWPVT-R (*n* = 204)		WJ-R AP (*n* = 213)		WJ-R LW (*n* = 213)	
	B	S.E.	*B*	S.E.	*B*	S.E.	*B*	S.E.	*B*	S.E.
Fixed effects[a]										
Class-level										
Intercept	96.66*	1.65	74.57*	1.10	66.93*	.52	81.70*	1.20	91.30*	.78
Tools versus control	−7.68*	3.33	3.66	2.12	1.42	1.05	2.56	2.45	−1.32	1.58
Student-level										
Primary language (English = 1, other = 0)			6.49*	1.96						
Age	−.33*	.12								
Pre-test score			.61*	.05	.86*	.04	.60*	.07	.63*	.07
Random effects[b]										
Between-class variance	33.51		6.12		.06		7.66		1.96	
Between-student variance	142.18		151.52		54.28		216.37		105.63	
Intra-class correlation	.270		.001		.0003		.005		.001	
Effect size	−.50		.22		.11		.14		−.11	

PPVT-III, Peabody Picture Vocabulary Test; EOWPVT-R, Expressive One-Word Picture Vocabulary Test-Revised; WJ-R AP, Woodcock–Johnson Applied Problems; WJ-R LW, Woodcock–Johnson Letter–Word Identification; SSRS, Problem Behaviors Scale of the Social Skills Rating System.
[a] Coefficient (S.E.).
[b] Variance.
* $p < .05$.

yield essentially the same estimated effects, but single-level regression analyses tend to underestimate standard errors and, thus, overestimate statistical significance. However, for all of the standard score language and literacy measures, the intra-class correlation (ICC) is .005 or less, raising questions about whether the HLM is necessary for these measures. HLM results are presented in Tables 7 and 8. Given the limited sample size at level two, it is noteworthy that the estimated effect of *Tools*

Table 8
HLM analysis estimating effects of tools on learning and development (raw score measures)

Variable	GRTR (*n* = 220)		WIPPSI (*n* = 211)		OLPT (*n* = 140)	
	B	S.E.	*B*	S.E.	*B*	S.E.
Fixed effects[a]						
Class-level						
Intercept	9.29*	.26	33.37*	.96	10.18*	.57
Tools versus control	.13	.52	.81	1.96	2.18+	1.17
Student-level						
Age	.23*	.03	.62*	.13	.34*	.07
Pre-test score			.53*	.06	.63*	.09
Random effects[b]						
Between-class variance	.18		5.80		1.90	
Between-student variance	12.21		125.60		30.23	
Intra-class correlation	.021		.063		.045	
Effect size	.03		.05		.34	

GRTR, Get Ready to Read; WIPPSI, Wechsler Preschool Primary Scale of Intelligence Animal Pegs subtest; OLPT; IDEA Oral Language Proficiency Test.
[a] Coefficient (S.E).
[b] Variance.
* $p < .05$.
+ $p < .10$.

on the SSRS remains statistically significant at the conventional .05 level while the estimated effect on the OLPT would be significant at the .10 level and the estimated effect on the PPVT narrowly misses ($p = .101$).

An issue of multiple comparisons arises because of the number of child outcome measures in our study. We have three hypotheses, one each relating to social development, cognitive development, and classroom processes. We have only one measure of social development to test the corresponding hypothesis. However, we have seven measures of cognitive development spanning language, literacy, and mathematics to test that relevant hypothesis. In addition, we have multiple measures of classroom processes relevant to our third hypothesis. Conducting multiple tests of significance across multiple outcome measures raises the risk of Type I, though there is disagreement about the need for adjustments to *p*-values and how such adjustments should be made (Cook & Farewell, 1996; Perneger, 1998). For the social development hypothesis there is only one test, so we make no adjustment. For the cognitive effects hypotheses, the estimated effects do not reach statistical significance at conventional levels in the HLM analyses even without an adjustment for multiple comparisons. The ordinary regression estimated effects on the PPVT and OLPT would not be significant at $p < .05$ adjusting for multiple comparisons. However, it should also be considered that six of the seven estimated effects of *Tools* on cognitive measures are positive ($p = .07$, two-tailed exact test). For the hypothesis regarding classroom practices, the estimated effects of *Tools* on total scores for the ECERS-R, SELA, and PCMI are statistically significant even after conservative Bonferroni corrections.

7. Discussion

The *Tools of the Mind* (*Tools*) curriculum was constructed on a foundation of Vygotsky's and Luria's theoretical work and practical studies of how children learn and how "tools" can be used to help children and teachers scaffold learning in the classroom. The *Tools* curriculum is a novel approach emphasizing intentional development of specific academic skills and self-regulation of behavior and emotions with play featured in a leading role in the curriculum. The *Tools* curriculum has been implemented in multiple sites, has well-developed training and curriculum materials, and has evidence of effectiveness from quasi-experimental pilot studies (Bodrova & Leong, 2001). This study was designed to conduct a more rigorous test of the educational effectiveness of *Tools* by a research team independent of the curriculum developers. A randomized trial was used to compare *Tools* to the standard practice in an urban school district, which was a district-developed curriculum.

Tools was found to improve children's classroom experiences, social development, and, with somewhat less confidence, cognitive development. Our findings are consistent with previous research findings discussed earlier indicating that differences in curricula can produce important differences in children's social development and behavior. The finding of substantial positive effects on social behavior is particularly important in view of concerns that typical child care programs increase problem behavior. Our results also demonstrate that self-regulation can be taught. The effect size for Tools is quite large compared to the estimated negative effect of child care, suggesting that this problem may be avoided by use of an appropriate curriculum. Moreover, our findings suggest that polarized debates – about academics versus play, child-initiation versus direct instruction, academic skills versus curiosity, and cognitive development versus socialization – pose false choices and are inadequately conceptualized.

Overall, classrooms in our study were similar in quality to state pre-kindergarten programs in many states (Clifford et al., 2005). However, *Tools* classroom environments were substantially better than the control classroom environments in several respects. Even though the research design equated the two groups on teacher and child characteristics, classroom structure and resources, and amount of in-service training, *Tools* classrooms attained higher overall levels of quality as assessed by the ECERS-R, SELA, and PCI. The advantages of *Tools* on the ECERS-R were particularly evident on the Language and Reasoning, Activities, and Interactions subscales. *Tools* classrooms also scored higher on the SELA, an assessment of the quality of the literacy environment and instruction, and on the PCI, which measures the frequency of use of scaffolding techniques by teachers. In addition, the *Tools* classrooms outperformed the control classrooms on teacher sensitivity and productivity as measured by the CLASS.

The observed differences between *Tools* and control classrooms are highly consistent with the design of the *Tools* curriculum. This provides an additional confirmation beyond the treatment fidelity measures that *Tools* was substantially implemented as designed and produced the changes in children's experiences that were predicted. The correspondence between observed and theoretical differences in curricula helps to rule out other potential explanations for differences in classroom processes and effects on children's learning and development. For example, one counter-explanation is that the effects were due to differences in the teachers. This is unlikely to begin with as teachers were randomly assigned blocking on their backgrounds, and teachers in both types of classrooms received equivalent amounts of training and professional development. Nevertheless, it is important that the specific differences in practice found by observation correspond to specific differences between the curricula. For example, several *Tools* techniques and activities are specifically designed to elicit language interactions between peers and there is an emphasis on teachers' scaffolding children's thinking through language interactions and the development of daily play plans. The implementation of these techniques would contribute to higher scores on the ECERS-R Language and Reasoning and Interactions, the SELA, the PCI, and the CLASS teacher sensitivity item. Similarly, *Tools* trains teachers to arrange play areas and play times so as to promote complex socio-dramatic play (i.e., long uninterrupted blocks of time, with materials for role play available in all center areas). These aspects of *Tools* would be expected to raise scores on the ECERS-R Activities subscale even though the materials in the classrooms are highly similar.

Finally, the fundamental emphasis on self-regulation would be expected to improve productivity as measured by the CLASS. To the extent that *Tools* is successful in enabling children to manage their own behavior, *Tools* teachers have much less need to spend time in managing children's behavior *per se* rather than facilitating their learning.

The findings regarding effects on children's learning and development are highly consistent with the classroom observation findings and the design of *Tools*. The consistency across differences in curriculum goals, observed classroom experiences, and children's learning strengthens confidence that the estimated effects on learning can be attributed to the *Tools* curriculum and are not explained by some other difference. The much lower level of problem behaviors (SSRS) among children in the *Tools* classroom is consistent with the emphasis of *Tools* on self-regulation as a fundamental skill to be developed and supported in the classroom and by results on the CLASS. In addition to focusing on activities that teach self-regulation, *Tools* provides other supports for children's self-regulation. For example, during partner reading, the child who is listening is given a picture of ears and the child who is "reading" has a picture of lips as continuous reminders of their roles. Stronger gains in language, literacy, and to some degree on mathematics would have been expected given the higher scores on Language and Reasoning, the SELA, and the PCI. The evidence of gains on the PPVT-III and OLPT likely reflect the effects of both staff–child language interactions and children's peer language interactions.

It was expected that gains would be smaller for language, literacy, and other cognitive measures than for social behavior. The most striking differences between the two curricula were with respect to activities promoting self-regulation. However, somewhat stronger findings with respect to cognitive development were expected, and it is worth considering why results were not stronger in this domain. One obvious reason is that the study was limited in its statistical power by the study because of the small number of classrooms. Effect sizes of .11–.33 were not statistically significant in this study after adjusting for clustering or multiple comparisons, but such effect sizes are as large as the estimated effects of preschool programs *per se* in some studies (e.g., U.S. Department of Health and Human Services, 2005). Another reason is that the curriculum was not fully implemented for the entire year as indicated by the change in the assessed fidelity over the course of the school year. Had this not been the first year of implementation by these teachers, results might have been stronger. In addition, as is often the case in preschool research, the outcome measures do not fully measure all of the goals of the curriculum. For example, *Tools* is designed to help children develop early literacy skills through writing, but early writing ability was not assessed by any of our measures. Finally, the control curriculum also focused on language and literacy.

Our study had significant strengths. Both teachers and children were randomly assigned to the treatments. Classroom resources, staffing structure, and other programmatic elements were identical between groups. In addition, the study ensured that the timing and amounts of teacher training and support and even experience with the curricula were similar. Although we cannot entirely rule out the possibility that *Tools* teachers were more energized by participation in the study, it seems unlikely that this would continue through the year. Yet, the fidelity measures show that practices improved over the course of the year, and the other measures of classroom practices reveal a pattern consistent with differences in the curricula rather than overall higher enthusiasm or effort by the *Tools* teachers. Multiple measures of both classroom experiences and children's learning and development are strengths, as well.

Our study also had significant limitations. It was conducted in 18 classrooms in a single school district. This limits the statistical power of the study, and there may have been more differences between *Tools* and the control curriculum than those found to be statistically significant. In the most conservative analyses, only the estimated effects on classroom experiences and self-regulation (social behavior) were statistically significant. In less conservative analyses, there was additional evidence for modest effects on language and literacy development. The setting and comparison to a locally developed curriculum limits generalization. While similar circumstances can be found elsewhere, it is unclear how Tools might have faired against other curricula in other places. These programs are well funded and have high standards for teacher qualifications, class size, and ratio. The population was almost entirely low-income and Hispanic. Most children came from Spanish-speaking homes, and their English language scores were quite low at entry to the preschool program. However, *Tools* has been implemented elsewhere in half- and full-day programs in head start, public schools, and private child care (Bodrova & Leong, 2001). There is no reason to believe that its advantages would be limited to particular types of programs, though one might hypothesize that the advantages would be greatest for populations for which the development of self-regulation is more problematic, whether because of family circumstances or the child's own special needs.

As noted earlier our evaluation took place during the first year that teachers had implemented Tools. Treatment fidelity measures indicated that the curriculum was not fully implemented during the early part of the year. By contrast, some of the control teachers had prior experience with the control curriculum. The *Tools* curriculum could be expected to produce larger effects when teachers have had more experience with the model. Finally, our sole measure of social and emotional development was teacher administered. It would be desirable to have objective measures, as well. Longitudinal studies would be useful to learn about the extent to which effects on self-regulation persist over time and might translate into long-term improvements in behavior and, perhaps, achievement.

In a subsequent study of the same programs with a smaller and somewhat different sample, *Tools* was found to produce positive effects on multiple objective measures of executive function, a construct that overlaps substantially with self-regulation (Diamond, Barnett, Thomas, & Munro, 2007). Important elements of executive function include inhibitory control, working memory, and cognitive flexibility. Each of those elements involves skills relevant to self-regulation, though this may be most obvious for its inhibition. Future studies might investigate these and other individual elements of executive function and self-regulation in greater detail while seeking to illuminate their effects on a range of outcomes including achievement test scores, school progress, internalizing and externalizing behavior problems, and mental health.

Future research is recommended to determine the extent to which our findings can be replicated, expanded, and generalized. Future studies should sample more classrooms (perhaps $n = 40$) to ensure that they are adequately powered and to offer a stronger basis for generalization. Such studies might look at stability and variation in curriculum effects across teachers who have a broad range of qualifications, preparation, and experience and across children with different socioeconomic backgrounds. Future studies also could expand the range of alternative curricula to which *Tools* is compared and might improve upon our approach to assessing treatment fidelity by having measures tailored to both alternatives rather than just *Tools*. No such instrument existed for the control curriculum in the present study.

In future studies, *Tools* should be compared to both more and less similar curricula. Other approaches to the education of young children that have been influenced by Vygotsky and have some similarities to Tools, though they also can differ in their interpretations of his theories and their implications for practice (Berk & Winsler, 1995). Although a detailed comparison is beyond the scope of this paper, the High/Scope and Mediated Learning curricula (though very different from each other) both can be said to reflect key principles articulated by Vygotsky in their practices (Dale, Jenkins, Mills, & Cole, 2005; Sylva, 1997). Reggio Emilia can be viewed as consistent with Vygotsky's principles in key respects (Berk & Winsler, 1995), and the widely used creative curriculum cites Vygotsky as an influence (Dodge, Colker, & Heroman, 2002). Curricula that are more similar to *Tools* would seem more likely to produce similar effects on self-regulation, and studies comparing them to *Tools* would be particularly interesting for that reason. It would also be interesting to compare *Tools* to Direct Instruction and other models with clearly opposed theoretical foundations. In light of previous research, such studies should be set-up to permit follow-up over time.

Finally, our focus was on the implementation of *Tools* and its effects on specific measures of children's learning and development. It should not be forgotten that the goals and objectives of early education are broader than our measures. Not only are the domains tapped here broader than our measures, but the educational goals people have for their children are broader yet. Research must be careful not to overly simplify the problem of improving early education. In addition, at least some future studies should also focus on the teacher's role in determining the ends of education and not just as a means of delivering a curriculum. Policy decisions about curriculum, as well as about teacher preparation, qualifications, and pay may profoundly influence this role.

References

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing*, *20*(5), 463–494.

Ballard, W. S., & Tighe, P. L. (1999). *IDEA Oral Language Proficiency Test*. Brea, CA: Ballard & Tighe Publishers.

Barnett, W. S. (1998). Long-term effects on cognitive development and school success. In W. S. Barnett & S. S. Boocock (Eds.), *Early child care and education for children in poverty: Promises, programs, and long-term results* (pp. 11–44). Albany, NY: SUNY Press.

Barnett, W. S., & Yarosz, D. J. (2004). *Who goes to preschool and why does it matter? Preschool policy matters, Vol. 8*. New Brunswick, NJ: NIEER.

Belsky, J., Vandell, D. L., Burchinal, M., Clark-Stewart, K. A., McCartney, K., Owen, M. T., et al. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701.

Bereiter, C. (1986). Does Direct Instruction cause delinquency? *Early Childhood Research Quarterly*, *1*(3), 289–292.

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington, DC: National Association for the Education of Young Children.

Blair, C. (2002a). Early intervention for low birth weight, preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology*, *14*, 311–332.

Blair, C. (2002b). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111–127.

Bodrova, E., & Leong, D. J. (1995). Scaffolding the writing process: The Vygotskian approach. *Colorado Reading Council Journal*, *6*, 27–29.

Bodrova, E., & Leong, D. J. (1996). *Tools of the Mind: The Vygotskian approach to early childhood education*. Upper Saddle River, NJ: Prentice-Hall.

Bodrova, E., & Leong, D. J. (2001). *Tools of the Mind: A case study of implementing the Vygotskian approach in American early childhood and primary classrooms. (UNESCO Innodata Monographs: Educational Innovations in Action No. 7)*. Geneva, Switzerland: International Bureau of Education, UNESCO.

Bodrova, E., & Leong, D. J. (2007). *Tools of the Mind: The Vygotskian approach to early childhood education* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Bronson, M. B. (2000). *Self-regulation in early childhood: Nature and nurture*. New York, NY: Guilford Press.

Bowman, B., Donovan, S., & Burns, S. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test-third edition (EOWPVT)*. Novato, CA: Academic Therapy Publications.

Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*(3), 126–143.

Cook, R. J., & Farewell, V. T. (1996). Multiplicity considerations in the design and analysis of clinical trials. *Journal of the Royal Statistical Society, Series A*, *159*(Part 1), 93–110.

Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, *45*, 545–580.

Dale, P. S., Jenkins, J. R., Mills, P. E., & Cole, K. N. (2005). Follow-up of children form academic and cognitive preschool curricula at 12 and 16. *Exceptional Children*, *71*(3), 301–317.

Datta, L. E., McHale, C., & Mitchell, S. (1976). *The effects of the Head Start classroom experience on some aspects of child development: A summary report of national evaluations, 1966–1969*. Washington, D.C.: Office of Child Development.

Diamond, A. (2007). Interrelated and interdependent. *Developmental Science*, *10*(1), 152–158.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*, 1387–1388.

Dodge, D. T., Colker, L., & Heroman, C. (2002). *The Creative Curriculum for preschool*. Washington, DC: Teaching Strategies, Inc.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention*, *28*(2), 67–91.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2006). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Elkind, D. (2001). Much too early. *Education Next*, *1*(2), 9–15.

Flavell, J. H. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.

Frede, E. C. (1989). *Preschool Classroom Implementation Rating Scale*. Ewing, NJ: The College of New Jersey.

Frede, E. C. (1998). The role of program quality in producing early childhood program benefits. In W. S. Barnett & S. S. Boocock (Eds.), *Early care and education for children in poverty: Promises, programs, and long-term results* (pp. 77–98). Albany, NY: SUNY Press.
Frede, E. C., Austin, A. B., & Lindauer, S. K. (1993). The relationship of specific developmentally appropriate teaching practices to children's skills in first grade. *Advances in Early Education and Child Care*, *5*, 95–111.
Frede, E., & Barnett, W. S. (1992). Developmentally appropriate public school preschool: A study of implementation of the High/Scope curriculum and its effects on disadvantaged children's skills at first grade. *Early Childhood Research Quarterly*, *7*(4), 483–499.
Gersten, R. (1986). Response to consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly*, *1*(3), 293–301.
Gresham, F., & Elliot, S. (1990). *Social Skills Rating Scale*. Circle Pines, MN: American Guidance Service.
Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environmental Rating Scale-Revised*. New York: Teacher's College Press.
Hirsh-Pasek, K., Hyson, M. C., & Rescorla, L. (1990). Academic environments in preschool: Do they pressure or challenge young children? *Early Education and Development*, *1*(6), 401–423.
House, E. R., Glass, G. V., McLean, L. F., & Walker, D. F. (1978). No simple answer: Critique of the "Follow Through" evaluation. *Harvard Educational Review*, *28*(2), 128–160.
Karnes, M. B., Shwedel, A. M., & Williams, M. B. (1983). A comparison of five approaches for educating young children from low income homes. In Consortium for Longitudinal Studies (Ed.), *As the twig is bent. . . lasting effects of preschool programs* (pp. 133–170). Hillside, NJ: Erlbaum.
Kozulin, A. (2003). Psychological tools and Mediated Learning. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 15–38). Cambridge: Cambridge University Press.
Krafft, K. C., & Berk, L. E. (1998). Private speech in two preschools: Significance of open-ended activities and make-believe play for verbal self-regulation. *Early Childhood Research Quarterly*, *13*(4), 637–658.
LaParo, K., Pianta, R. C., & Stuhlman, M. (2004). Classroom Assessment Scoring System (CLASS): Findings from the pre-k year. *The Elementary School Journal*, *10*, 409–426.
Leong, D., & Hensen, R. (2003). *Tools of the Mind research training manual* (2nd ed.). Denver, CO: Center for Improving Early Learning, Metropolitan State College.
Luria, A. L. (1965). The directive function of speech in development and dissolution. In R. C. Anderson & D. P. Ausubel (Eds.), *Reading in the psychology of cognition* (pp. 350–363). New York: Holt, Rinehart and Winston.
Luria, A. R. (1966). *The higher cortical functions in man*. New York: Basic Books.
Luria, A. R. (1979). *The making of mind: a personal account of Soviet psychology*. Cambridge, MA: Harvard University Press.
Mac Iver, M., & Kemper, E., (2002). Guest editors' introduction: Research on Direct Instruction in reading. *Journal of Education for Students Placed at Risk*, *7*(2). Retrieved February 21, 2006 from http://www.louisville.edu/edu/jespar/vol_7_no_2_editors.htm.
Marcon, R. (1999). Differential impact of preschool models on development and learning of inner-city children: A three cohort study. *Developmental Psychology*, *35*(2), 358–375.
Marcon, R., (2002). Moving up the grades: Relationship between preschool model and later school success. *Early Childhood Research and Practice, 4*(1). Retrieved February 21, 2007, from http://ecrp.uiuc.edu/v4n1/marcon.html.
Miller, L. B., & Bizzell, R. P. (1983). The Louisville experiment: A comparison of four programs. In Consortium for Longitudinal Studies (Ed.), *As the twig is bent. . . lasting effects of preschool programs* (pp. 171–200). Hillsdale, NJ: Erlbaum.
Miller, L. B., & Dyer, J. L. (1975). Four preschool programs: Their dimensions and effects. *Monographs of the Society for Research in Child Development*, *40*(5–6, Serial No. 162), 727–741.
Mills, P., Cole, K., Jenkins, J., & Dale, P. (2002). Early exposure to Direct Instruction and subsequent juvenile delinquency: A prospective examination. *Exceptional Children*, *69*(1), 85–96.
Nourot, P. M. (2004). Historical perspectives on early childhood education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (pp. 3–31). Englewood, NJ: Prentice Hall.
Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, *321*, 1236–1238.
Pianta, R. C., La Paro, K., & Hamre, B. (2005). *The Classroom Assessment Scoring System pre-k manual*. Unpublished manuscript, University of Virginia.
Schweinhart, L. J., & Weikart, D. P. (1997). The High/Scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly*, *12*(2), 117–143.
Schweinhart, L. J., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age fifteen. *Early Childhood Research Quarterly*, *1*(1), 15–46.
Singer, D. G., Golinkoff, M. R., & Hirsh-Pasek, K. (2006). *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth*. New York: Oxford University Press.
Smith, S., Davidson, S., Weisenfeld, G., & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)*. New York: New York University.
Snow, C. E., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
St. Pierre, R. G., Anderson, R. B., Proper, E. C., & Stebbins, L. B. (1978). That Follow Through evaluation. *Phi Delta Kappan*, *59*(10), 729.
Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *A planned variation model: Vol. IV-A effects of Follow Through models*. Washington, DC: US Office of Education.
Stipek, D., Daniels, D., Galluzzo, D., & Milburn, S. (1992). Characterizing early childhood education programs for poor and middle-class children. *Early Childhood Research Quarterly*, *7*, 1–19.
Stipek, D., Daniels, D., Galluzzo, D., Millburn, S., & Salmon, J. M. (1998). Good beginnings: What difference does the program make in preparing young children for school. *Journal of Applied Developmental Psychology*, *19*(1), 41–66.
Stipek, D., Feiler, R., Daniels, D., & Millburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development*, *66*, 209–223.
Sylva, K. (1997). The quest for quality in curriculum. In L. J. Schweinhart & D. P. Weikart (Eds.), *Lasting differences: The High/Scope preschool curriculum comparison study through age 23* (pp. 89–93). Ypsilanti, MI: High/Scope Press.
U.S. Department of Health Human Services, Administration for Children and Families. (2005). *Head Start impact study: First year findings*. Washington, DC: Author.
Vandell, D. L. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly*, *50*, 387–414.
Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard Press.
Vygotsky, L. S. (1997). *The history of the development of higher mental functions* New York: Plenum Press.
Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-revised*. New York, NY: The Psychological Corporation.
Whitehurst, G. J. (2001). Much too late. *Education Next*, *1*(2), pp. 9, 16–21
Whitehurst, G., & Lonigan, C. (2001). *Get Ready to Read!*. Columbus, OH: Pearson Early Learning.
Woodcock, R., & Johnson, M. (1989). *Woodcock–Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.
Woodcock, R., & Munoz-Sandoval, A. F. (1996). *Bateria Woodcock–Munoz: Pruebas de Habilidad Cognitive-Revisada*. Chicago, IL: Riverside Publishing.
Zigler, E. F., & Bishop-Josef, S. F. (2006). The cognitive child vs. the whole child: Lessons from 40 years of Head Start. In D. G. Singer, M. R. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 15–35). New York: Oxford University Press.