Journal of School Psychology 50 (2012) 379–401

Contents lists available at SciVerse ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs☆

Angeline S. Lillard

Department of Psychology, University of Virginia, Charlottesville VA 22904-4400, USA

ARTICLE INFO

Article history:
Received 21 December 2010
Received in revised form 9 January 2012
Accepted 12 January 2012

Action Editor: Robin Hojnoski

Keywords:
Early childhood education
Fidelity of implementation
Academic achievement
Social development
School readiness

ABSTRACT

Research on the outcomes of Montessori education is scarce and results are inconsistent. One possible reason for the inconsistency is variations in Montessori implementation fidelity. To test whether outcomes vary according to implementation fidelity, we examined preschool children enrolled in high fidelity classic Montessori programs, lower fidelity Montessori programs that supplemented the program with conventional school activities, and, for comparison, conventional programs. Children were tested at the start and end of the school year on a range of social and academic skills. Although they performed no better in the fall, children in Classic Montessori programs, as compared with children in Supplemented Montessori and Conventional programs, showed significantly greater school-year gains on outcome measures of executive function, reading, math, vocabulary, and social problem-solving, suggesting that high fidelity Montessori implementation is associated with better outcomes than lower fidelity Montessori programs or conventional programs.

1. Introduction

Montessori education has been in existence for over 100 years, and today is used in over 4000 schools in the United States alone (Cossentino, 2005), including about 400 public and charter schools (Shapiro, 2007). However, studies of Montessori education's impact on development are rare (Walsh & Petty, 2007). The few studies that exist present a mixed picture, with some showing better outcomes than other programs (Besancon & Lubart, 2008; Dohrmann, Nishida, Gartner, Lipsky, & Grimm, 2007; Lillard & Else-Quest, 2006; Miller & Bizzell, 1984; Rodriguez, Irby, Brown, Lara-Alecio, & Galloway, 2005), and

☆ This research was supported by grants from the Brady Education Foundation and the National Science Foundation. I thank Alyssa Hart, Nicole Godine, Linda Kim, Matt Lerner, Jeff Spies, and Morgan Woodyard for their help with running the study and statistical analyses, and the parents, children, and schools who participated.

E-mail address: lillard@virginia.edu.

0022-4405/$ – see front matter © 2012 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsp.2012.01.001

others showing similar or even worse outcomes (Cox & Rowlands, 2000; Krafft & Berk, 1998; Lopata, Wallace, & Finn, 2005). This article examines whether different outcomes are associated with different implementations of Montessori, consistent with other domains in which fidelity of implementation has been related to different program outcomes (Blakely et al., 1987; O'Donnell, 2008).

1.1. Montessori education and its outcomes

Montessori education began in a slum in Rome in 1907 when Maria Montessori, one of the first woman physicians in Italy (Kramer, 1976; Povell, 2009), extended her sensorimotor work with children with mental disabilities to children who were typically developing and from low-income backgrounds. A distinctive feature of Montessori education is dozens of specific materials[1] (Lillard, 2011a) and activities that confer learning through action (Piaget, 1970, pp. 147–8). These materials were developed by Dr. Montessori and her collaborators over 45 years (with some limited modifications thereafter) and are integral to the program that also involves placing children in 3-year groupings, lessons that are almost exclusively in small groups or one-on-one (Hojnoski et al., 2008), a highly organized physical space, and free choice among constructive learning activities (Humphryes, 1998). The materials used in these activities provide corrective feedback, and are grouped into curricular areas of Sensorial, Language, Mathematics and Geometry, Geography, Culture, Music, and Art, as well as Practical Life (Montessori, 1967, 1989a, 1989b). Theoretically, Montessori education embodies many features known to enhance learning and development (Glenberg, Jaworski, Rischal, & Levin, 2007; Lillard, 2005, 2011b), but only some studies of Montessori education show that it confers better developmental outcomes.

Three published studies of Montessori education have used some form of random assignment to minimize possible selection effects. The most recent of these found better outcomes for low-income children in a public Montessori school than those in business-as-usual control schools on a variety of measures, both at ages 5 and 12 (Lillard & Else-Quest, 2006). However, the other two randomized controlled studies, both involving Head Start, found that immediate outcomes for Montessori programs were similar to those of other types of preschool programs (Karnes, Shewedel, & Williams, 1983; Miller & Dyer, 1975), although effects did appear later ("sleeper effects"). In one study (Karnes et al., 1983) delayed social and cognitive benefits appeared by high school for all children who had previously attended Montessori programs, and in the other (Miller & Bizzell, 1984) later benefits emerged only for boys.

Other studies have examined outcomes of Montessori education using samples that were matched at either the individual or school level. One recent study using four demographically-matched traditional public and magnet city schools found equal math and worse language arts outcomes for 8th graders enrolled in public Montessori schools as compared to children in the matched schools; 4th graders in Montessori and other types of programs performed equally (Lopata et al., 2005). These findings contrast with a different study finding that children who had attended public Montessori programs from ages 3 to 11 performed better at ages 15 to 18 in math and science than demographically-matched classmates who had been in other programs through the 5th grade (Dohrmann et al., 2007). Other studies have shown better socio-emotional outcomes for children attending Montessori middle schools than controls in demographically-matched conventional middle schools (Rathunde & Csikszentmihalyi, 2005a, 2005b), whereas another study documented less private speech—a developmentally-important activity—by children attending a Montessori school than by controls attending a conventional preschool who were matched by family socioeconomic status (Krafft & Berk, 1998). In sum, the existing research presents a mixed picture regarding Montessori education outcomes. One possible explanation for this mixed picture concerns implementation of the Montessori program in the different studies.

1.2. Implementation fidelity

Fidelity of implementation refers to how well a program is implemented relative to the original or the ideal. The concept has long been of concern in the health literature, but its importance in education literature is only recently being recognized (O'Donnell, 2008). Higher implementation fidelity is generally

[1] Hereafter, *Montessori materials* refers both to Montessori materials (e.g., the Pink Tower) and activities (e.g., Table Washing and the Silence.)

associated with better program outcomes, and lower fidelity, in which programs are altered from the original, is often associated with poorer outcomes (Carroll et al., 2007; Durlak & DuPre, 2008; Odom, 2009; O'Donnell, 2008). However, there is a flip side to implementation fidelity: longevity. Programs last to the extent that they "reinvent" (O'Donnell, 2008, p. 44) themselves by adapting to the cultures and circumstances into which they disperse (Goodman & Steckler, 1989). By compromising fidelity in the adaptation process, programs can lose some of their effectiveness, but they can gain in endurance. Indeed, research on scaling interventions up frequently shows that positive effects diminish with dissemination (Durlak & DuPre, 2008), which common sense would suggest entails adaptation. A purpose of this research is to examine whether Montessori education might also reflect this tradeoff between high fidelity and endurance.

1.3. Types of Montessori implementation

The existing literature rarely discriminates Montessori program implementations, but two basic styles can be observed in the United States: a classic approach that adheres tightly to Dr. Montessori's original program as outlined in her books (Montessori, 1967, 1989a, 1989b) and a supplemented one in which conventional school activities and materials are added to the core program.[2] The classic implementation is strict and traditional, and often identified with the Association Montessori Internationale (AMI), the organization Dr. Montessori founded to carry on her work. It includes 3-hour work periods in the morning and, for children over age 4, in the afternoon as well; a 3-year age grouping; a single trained teacher; and a specific set of Montessori materials. However, few Montessori programs in the United States have a strict and traditional program. Indeed, as Povell (2009) discussed at length, adaptation to the local culture was an important aim of Dr. Nancy Rambusch, the founder of the American Montessori Society (AMS), America's largest Montessori society that is credited with reviving Montessori education in 1960s America (O'Donnell, 2007). As stated on the AMS website, "From the beginning, Dr. Rambusch and AMS advanced a modified version of Montessori education, more attuned than the original to mid-20th century American culture" (American Montessori Society, 2011). Adapting Montessori education to the local culture was integral to its re-rooting in America in the 1960s (it had been popular in the 1910s but diminished thereafter). Although this American adaptation had several aspects, such as making a college degree a prerequisite for teacher training, one aspect—the focus here—was adding to the standard set of Montessori materials. Rambusch is said to have added "easel painting, clay modeling, and educational toys" (O'Donnell, 2007, p. 87), and today, the AMS School Accreditation standards state that, "Classroom environments in quality Montessori schools are equipped with a full range of Montessori *and curriculum support* materials" (emphasis added; American Montessori Society, 2009, p. 6, Standard 3.71). Regardless of organizational affiliation, many Montessori classrooms worldwide adapt the basic program. Besides supplementary materials, other common Montessori adaptations are shorter work periods, special classes and additional teachers for extracurriculars, grades, and homework. Without such flexible adaptation, Montessori might not have taken root and thrived in America as it has since 1960. At issue is whether adaptation might be associated with different levels of effectiveness.

The studies reviewed earlier showing more positive outcomes all involved high fidelity or "classic" implementations of Montessori. Lillard and Else-Quest (2006), which showed that at 5 and 12 years of age children who attended Montessori programs scored higher on several social and cognitive outcome measures, and Dohrmann et al. (2007), which showed that in high school children who had previously attended Montessori programs excelled in science and math, both involved Montessori programs that would likely be considered classic implementations because the schools were associated with AMI. Four of the five schools used in the Rathunde and Csikszentmihalyi (2005a,b) studies examining social-emotional outcomes were associated with AMI; even the one that was not officially recognized by AMI, because it had not sought the status, was hand-picked for its classic implementation (Rathunde, personal communication, July, 10, 2009). In contrast, the Head Start studies (Karnes et al., 1983; Miller & Dyer, 1975) showing weaker and distal effects had low fidelity implementations. Both programs included only children of one age (4-year-olds) rather than the 3-year age grouping Dr. Montessori's books (e.g., 1967, p. 35,

[2] Supplemented programs might differ in other ways as well; in this study use of Montessori materials is the index of fidelity of implementation.

p. 302; 1989c, p. 68) describe. In one program, only one-half hour per day—not three hours—was spent working with the Montessori materials (Karnes et al., 1983). Schedules were not provided for the other Head Start study, but its Montessori programs' mean implementation score was 6.5 on a scale of 0 (*not at all*) to 10 (*best possible*), suggesting adaptation or a shift from the classic approach (Miller & Dyer, 1975). Krafft and Berk's (1998) article showing less private speech describes limited Montessori work time, supplementary materials, and "workstations" set up by the teachers at the Montessori school where they tested. In contrast, materials in classic Montessori classrooms are typically made available for children to choose from open low shelves (Montessori, 1966, p. 121). The Montessori elementary school used in the Lopata et al. (2005) study, which found worse language arts scores for children in 8th grade, issued evaluative grades to children and also had "specials" (both alterations to a classic Montessori program; see Montessori, 1967, pp. 59–60). Taken together, these reports leave open the possibility that fidelity of implementation might explain variations in outcomes, with higher fidelity programs possibly being associated with better outcomes.

Differences in implementation are multifarious, and there are no established measures of Montessori program fidelity. One difference that is simply and objectively measurable is the percentage of children in a classroom using classic Montessori materials at any given moment. This is essentially "a measure of the basic extent of use of the curricular materials," which is one of six accepted ways of measuring program fidelity (O'Donnell, 2008, p. 39). This unidimensional measure might index other more complex aspects of fidelity (see Carroll et al., 2007, for discussion of complexity), like organization of the environment or navigation of the curriculum. Children must virtually always use Montessori materials in classrooms that provide only those materials, and teachers who provide only the classic Montessori materials are likely to also be strict about how they implement other aspects of the program. The set of materials that are required for a true Montessori program is well-agreed upon by Montessori teacher trainers (Lillard, 2011a), and Dr. Montessori (1989c) advised against expanding on this set:

> The material of our schools today is based on the selection that the children have voluntary made themselves from the mass of things that was placed at their disposal … This selection brought us to the conception that there must be just that amount and no more. . . . The fundamental fact in the preparation of the environment is to have only one set of each type of material. (p. 64)

This selection of materials was considered complete: "The didactic materials, for instance, established through experimentation, is [sic] given complete to the teacher" (Montessori, 1997, p. 31). As O'Donnell (2007) noted regarding adding to this set of materials, "Modern embellishments to the curriculum did not please the . . . purists" (p. 87).

In supplemented Montessori programs, however, at any given moment, children might be engaged with activities beyond the classic Montessori set of materials, such as crafts projects, commercial puzzles, games, and blocks, and worksheets. Working with such materials instead of the classic Montessori ones could in theory be beneficial, harmful, or neutral. This exploratory study sought preliminary empirical evidence on this issue.

Children's development across the school year was examined in two types of Montessori preschool programs: ones that use only classic materials and ones that supplement the classic materials with commercial ones. Development in conventional preschool programs that had no Montessori materials was also examined for comparison. The conventional programs were in highly regarded schools and were aligned with the basic principles of the National Association for the Education of Young Children and recognized by the state's Independent Schools association; in addition, parents of children in the Montessori programs most often chose them as the alternative private school their child would attend were a Montessori program not available. Each sample was composed of families that as a group were statistically the same in terms of income, education, and ethnicity. Importantly, with one exception the children's scores on the tasks at the beginning of the school year did not differ between the programs, mitigating the concern that unobserved family or child characteristics associated with school selection were the main drivers of results, since selection effects should have been observed in the fall as well. The exception (vocabulary score) favored the conventional programs, and this was used as a covariate in statistical analyses. The outcomes measured in this study fit under the overarching concept of "school readiness."

1.4. School readiness

Young children's school readiness, or preparedness to thrive in school settings, is of great concern (Graue, 2006; Pianta, Cox, & Snow, 2007). Children who arrive in first grade lacking some number and letter knowledge, age-appropriate language skills, and sufficient social skills and self-regulation disadvantage themselves and also their classmates (Blair et al., 2007; Raver, Garner, & Smith-Donald, 2007). School readiness is therefore a major goal of early childhood programs (Pianta et al., 2007). Here we tested for a multifaceted set of skills that help children thrive in early school settings: executive function, theory of mind, social problem solving, early reading, vocabulary, and early math (Blair & Diamond, 2008; Blair & Razza, 2007; Howes et al., 2008; Kendeou, van den Broek, White, & Lynch, 2009; Welsh, Parke, Widaman, & O'Neil, 2001). The value of reading, vocabulary, and math for school success is clear; the others are discussed in the sections that follow.

1.4.1. Executive function

Executive function (EF) is an umbrella term covering several component skills, such as working memory, inhibitory control, attention, planning, and flexibility (Bull, Espy, & Wiebe, 2008; Garon, Bryson, & Smith, 2008; Huizinga, Dolan, & van der Molen, 2006; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). It is sometimes viewed as redundant with self-regulation (Blair & Diamond, 2008; Bronson, 2001). Executive function skills are significantly related to children's success in school (Blair & Razza, 2007; Bull et al., 2008; Espy et al., 2004; McClelland, Acock, & Morrison, 2006; McClelland et al., 2007; Mischel, Shoda, & Rodriguez, 1989; Passolunghi, Vercelloni, & Schadee, 2007) as well as social skills (Hughes, 1998; Mischel et al., 1989; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006). Executive function is trainable (Diamond & Lee, 2011; Dowsett & Livesey, 2000; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Thorell, Lindqvist, Nutley, Bohlin, & Klingberg, 2009), diminished by some experiences (Lillard & Peterson, 2011), and helped by certain preschool programs (Diamond, Barnett, Thomas, & Munro, 2007; Lillard & Else-Quest, 2006).

1.4.2. Theory of mind

Theory of mind is a set of interrelated concepts through which we explain and predict others' actions (Wellman, 1990). Performance on theory of mind tasks has been shown to predict social competence in children (Baron-Cohen, 2000; Capage & Watson, 2001; Dunn & Hughes, 2001; Razza, 2005; Tager-Flusberg, 2003; Watson, Nixon, Wilson, & Capage, 1999), which in turn is related to success in school (Welsh et al., 2001) and life (Salovey & Mayer, 1990). Preschool performance on the central theory of mind task, the false belief task, also predicts math and reading performance in kindergarten (Blair & Razza, 2007).

1.4.3. Social problem solving

Children with better social problem solving skills are both more socially competent (Pettit, Dodge, & Brown, 1988) and better-liked by their peers (Erwin, 1993; Rubin & Daniels-Beirness, 1983). Children with more social competence also do better in school (Ladd, Birch, & Buhs, 1999; Malecki & Elliott, 2002; Welsh et al., 2001). These three abilities and three academic abilities were examined in this study.

1.5. Study overview and hypotheses

The research conducted here was designed to address two main questions. First, do preschool children's school readiness skills change from fall to spring as a function of program type (Classic Montessori, Supplemented Montessori, or Conventional)? Second, within Montessori programs, does the percentage of children using Montessori materials in a classroom predict children's school readiness skills in the spring, after controlling for fall skill levels? We hypothesized that the gains of children in Classic Montessori programs would exceed those of children in Conventional programs for the outcomes tested, mainly because of an earlier study (Lillard & Else-Quest, 2006). In that study, low-income children in a public school implementing a classic Montessori program were compared with children who had lost the lottery for that Montessori program and were instead enrolled in other programs ranging from typical Nth Street Elementary to high caliber conventional language immersion schools (Lillard & Else-Quest, 2006).

Significant differences were seen for all but one of the outcomes measured here (i.e., vocabulary), although different tasks were used to measure some constructs. The Lillard and Else-Quest (2006) study was able to use randomization to control for possible selection effects, but it concerned a different population. Also, based on the existing research, we expected children in Classic Montessori programs would outperform children in Supplemented Montessori programs, although we know of no direct comparisons of children in these different types of Montessori program. In addition, because of past research, we hypothesized that Supplemented Montessori programs would not be associated with better performance than Conventional programs.

2. Method

2.1. Participants

One hundred seventy-two children ranging in age from 33 to 76 months at the fall test date participated. Table 1 shows demographic information by program type. According to the heads of schools, virtually all of the children attending Montessori programs entered their present classroom at age 3, and those attending the conventional programs entered the program at age 3 but changed classrooms in the fall after they turned 4 because the conventional programs divided classrooms by 3- to 4-year-olds and 4- to 6-year-olds.

Maternal education, reported for 98% of the sample, ranged from 4.45 to 4.73, with "4" indicating completing a 4-year college degree, and "5" indicating some graduate education. Over 90% of children were White, reflecting the composition of the communities in which the schools were located; 5% were African American, 3% were Asian, 1% were Hispanic, and 1% were of other races and ethnicities. All children were fluent in English.

2.2. Settings and teachers

2.2.1. Program type

Classrooms associated with three program types were targeted in this study. All three types of programs offered 3-hour sessions for 3- and 4-year-olds and 6- to 7-hour sessions for 5-year-olds. Classic and Supplemented Montessori classrooms shared many features that are discussed first, followed by the ways in which they diverged.

2.2.1.1. Montessori (including Classic Montessori). All 12 Montessori classrooms adhered to Montessori philosophy in important ways: they all employed a 3-year age grouping which aligned Dr. Montessori's developmental stages (ages 3 to 6), none used extrinsic reward programs (like gold stars), and the teachers gave mainly small group and individual lessons. All of the Montessori classrooms had four to five group tables seating four to six children, four to five tables that seated only one child, and a circle taped on

Table 1
Demographic features of children and programs.

Variable	Program type		
	Classic Montessori	Supplemented Montessori	Conventional
Number of classrooms	3	9	6 (3 at each age)
N (girls)	36 (22)	95 (47)	41 (26)
Average age of children in months (*SD*s)	55.6 (10.27)	54.01 (10.84)	59.17 (8.43)
Number of children at ages 3, 4, and 5+	9, 15, 12	28, 40, 27	2, 21, 18
Maternal education	4.58 (0.55)	4.61 (0.66)	4.73 (0.51)
Percent White	85	92	95
Average years teacher has taught	2.0 (range 1–3)	8.3 (range 1–30)	7.7 (range 5–12)
Average proportion (and *SD*) of children engaged in Montessori work across snapshots	.98 (.03)	.48 (.17)	NA

the floor for circle time. Low shelves were placed around the room and had materials organized into the Montessori curriculum areas. The Montessori classrooms all had 2 to 3 hours of dedicated work time in the mornings, during which children freely chose work off the shelves, and some outdoor time midday. Snacks in all the Montessori classrooms were available to children to take individually when they wanted. In all the Montessori classrooms, there was some whole class circle time, during which teachers read, conversed with students, and gave whole-class lessons in Grace and Courtesy (a part of the Montessori curriculum). Older children in all the Montessori classrooms stayed for lunch and another 2- to 3-hour work period in the afternoon. Classic Montessori classrooms as well as Supplemented Montessori classrooms shared these features.

2.2.1.2. Supplemented Montessori implementations. Classrooms targeted as implementing Supplemented Montessori differed from those targeted as implementing Classic Montessori in three key ways. First, in addition to offering many of the classic Montessori materials, they also offered conventional preschool materials, like craft projects, beads, puzzles, workbooks, commercial games like Memory, and commercial materials like LEGOs, on classroom shelves. Second, once or twice a week, the morning work period was interrupted for specials, such as 30-minute music or Spanish lessons. Third, a second adult was involved with the children as the main teacher. In contrast, in the Classic Montessori classrooms, a second adult observed the children but rarely interacted with them.

2.2.1.3. Conventional. In the conventional classrooms, there were areas set up for pretend play (e.g., supplying dress-ups and a toy kitchens) and reading (supplying books and pillows), and shelves of typical preschool materials, like blocks, LEGOs, beads, and puzzles. Each classroom had five or six tables seating 4 to 6 children and a circle on the floor for circle time. Art supplies were also present, including paints, crayons, markers, and playdough. Children in both programs also had workbooks for math and language. In both conventional schools the day was organized in a mix of lessons, specials, outside time, circle time, snack and lunch time, and free play time in the classroom, reflecting the array of activities typically seen in American preschools (Hamre & Pianta, 2007). In lessons, which took a total of about 60 minutes each morning, children sat with the teacher (sometimes as a whole class and other times in small groups) while they discussed, for example, days of the week and different letters and their sounds and practiced counting and elementary math. The daily schedule of lessons was structured in one school program (e.g., 30 minutes of reading daily, 20 minutes of math daily, and 20 minutes of Science on Wednesdays and History on Thursday) and more variable in the other. Specials involved children going out of the class for art, music, and Spanish. Outside time was usually free play outside, but occasionally teachers organized games like tag. Snack time involved children sitting at group tables and being served a snack. During free playtime in the classroom, about 90 minutes in the morning at each school, children were able to choose their own activities. Like most conventional preschools (Mashburn et al., 2008), these two did not subscribe to any particular named curriculum, like High/Scope (Epstein, 2006), but had put together their own. None of the Conventional classrooms had any Montessori materials. In both conventional programs, ages were stratified, with 3- to 4-year-olds in one classroom and 4- to 6-year-olds in another.

2.2.2. Teachers

All of the teachers were White women. Teachers at the Classic Montessori school were in their 1st to 3rd years of teaching, teachers at the Supplemented Montessori schools were in their 1st to 30th years, and those at the Conventional schools were in their 5th to 12th years. Prior research has shown that teacher experience has no effects or at most modest effects on student outcomes (Greenwald, Hedges, & Laine, 1996; McDonald Connor, Son, Hindman, & Morrison, 2005; NECCRN, 2002), so teacher experience was not expected to be important, but analyses were conducted to verify this in the current study. All the Montessori teachers had been certified by a Montessori organization and had college degrees; all but one of the conventional teachers had a college degree. Teachers at the Classic Montessori had all completed AMI training courses; 6 of the teachers at the Supplemented Montessori had AMS training and 3 had AMI training. Every teacher had an assistant. Teacher education as indicated by college degrees has not been associated with child outcomes (Early et al., 2007), and thus, was not included in analyses for this study.

2.3. Measures

2.3.1. Executive function (EF): The Head-Toes-Knees-Shoulders task

The Head-Toes-Knees-Shoulders (HTKS) task (Ponitz et al., 2008) asks that children carry out an opposing response, for example touching their toes when asked to touch their head. It is thought to involve several components of executive function: working memory to keep rules in mind, attention to tester instructions, planning to execute the action, and inhibitory control to avoid the prepotent response. The HTKS task predicts kindergarten achievement and is significantly related to parent and teacher reports of inhibitory control and behavior regulation in preschool populations (McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009; Ponitz et al., 2008). Six-month test–retest correlation coefficients are over .55 for kindergarteners (Ponitz et al., 2008). Scoring reliability for this measure has been demonstrated in prior studies (Ponitz et al., 2009, as cited in Ponitz et al., 2008).

To conduct this task, the experimenter taught the child a rule: "If I say to touch your head, I want you to touch your toes, and if I say to touch your toes, I want you to touch your head." After training, the child was given 10 such commands in a fixed order with no command repeated more than twice in succession. Children were scored 2 if they immediately followed the opposite instruction, and 1 if they did so after a quick touch of the wrong location. If children scored at least 10 (of a possible 20) on the first 10 commands, a second rule was added: "When I say to touch your knees, you touch your shoulders, and when I say to touch your shoulders, you touch your knees." After training on this command, children were given 10 additional trials involving all four commands, for a final score of 0 to 40.

In addition to HTKS, the choice modification of the classic delay of gratification task, in which children chose whether to have one desirable item (i.e., stickers and wind-up toys) immediately or have three later, was also used to assess inhibitory control (Mischel et al., 1989). Subsequent literature review of this task revealed that even when nine such trials are used, it is unrelated to other tests of inhibitory control (Hongwanishikul, Happeny, Lee, & Zelazo, 2005), so it is not discussed further (see also Schwarz, Schrager, & Lyons, 1983; Toner, Holstein, & Hetherington, 1977).

2.3.2. Theory of mind

Theory of mind was assessed in this study using the Theory of Mind Scale (Wellman & Liu, 2004), a set of tasks that 90% of American children come to pass in a particular order as indicated by Guttman scaling and Rasch analysis (Wellman, Fang, & Peterson, 2011), and which has been widely adopted and validated in a range of populations, including Chinese children (Wellman, Fang, Liu, Zhu, & Liu, 2006), German children (Aschersleben, Hofer, & Jovanovic, 2008), deaf children (Peterson & Wellman, 2009), and children with specific language impairment (Farrant, Fletcher, & Maybery, 2006). Because of the ages of children in this study, we used the 3 most difficult items on the 5-item scale: Perceptual Access (which data from 280 American and Australian children suggest is passed at age 4.45), False Belief (which the average American child passes at age 4.77), and Hidden Emotion (which the average American child passes at age 5.15; ages taken from Wellman et al., 2011). Others have presented these tasks in a variety of orders, although Hidden Emotion is typically last. In this study, to minimize testing time, the False Belief task was always presented first, Hidden Emotion was presented only if a child passed False Belief, and Perceptual Access was presented only if a child failed False Belief. Although data is not available for all three of these tasks across the ages studied here, the 4-week test–retest correlation coefficient for 5-year-olds on a set of similar false belief tasks was $r = .77$ (Hughes et al., 2000). Materials for this task were 4 dolls 3 to 4 inches high, a doll-sized side table with a drawer containing a small shell, a BandAid box containing a pencil, a small super-ball, and a placard showing a sad, neutral, and happy face.

First, children were given the False Belief task, in which they were shown the BandAid box and asked if they knew what was inside. Children almost always guessed the typical contents; if not, they were prompted ("What's usually in this box?") and then guessed correctly. The box was opened, revealing the content, and the experimenter commented, "Ha! Look at that! There's actually a pencil in there!" The box was closed, and a doll was brought out. The experimenter added, "Peter has never seen inside this BandAid box. Now here comes Peter. So, what does Peter think is in the box? BandAids or a pencil? [Target question]. Did Peter see inside this box? [Memory question]" Children who answered both questions correctly were given two points and moved up to the Hidden Emotion task. Children who answered either question incorrectly received 0 points and moved down to the Perceptual Access task.

For the Perceptual Access task, a nondescript wooden chest with a closed drawer containing a small shell was brought out, and children were asked, "What do you think is inside the drawer?" Regardless of what children guessed, the experimenter said, "Let's see," and opened the drawer. "There's really a shell inside!" She closed the drawer and asked, "Okay, what is in the drawer?" as a memory check, and then produced a doll. "This is Polly. Polly has never seen inside this drawer. Now here comes Polly. Does Polly know what is in the drawer? Did Polly see inside the drawer?" Children who answered "no" to both questions were scored 1; otherwise the score was 0.

Children who passed the False Belief task, and thus were very likely to understand the Perceptual Access task, went on to the Hidden Emotion task after False Belief. For Hidden Emotion, first children were shown the sheet with happy, neutral, and sad faces on it, and were asked to point to the sad and happy faces. Then they were shown the first doll, and the experimenter said, "This is Joey. Today is Joey's birthday. He really wants a bicycle for his birthday, and he thinks his uncle is going to give him a bicycle." Then the second doll was produced. "This is Joey's uncle. His uncle is coming to give him his birthday present. Look, his uncle is giving him a ball for his birthday." The sheet of faces was presented again, and children were asked, "Which picture shows how Joey really feels when he gets the ball?" and then, "Which picture shows what Joey's *face* will look like when he gets the ball?" A correct response was to point to the sad face for how Joey really feels, and the happy one for what his face looks like. A point for being correct on this task was added to the 2 points acquired for being correct on False Belief (because it was assumed that such a child would also be correct on Perceptual Access), and so a child who passed Hidden Emotion received a total of 3 points for the Theory of Mind scale.

2.3.3. Social problem solving

An object acquisition story from the Social Problem Solving Task (Rubin, 1988), one of the tasks used to assess social competence in the NICHD's seminal study of early child care (see also Spivak & Shure, 1974), was also used here. Children's performance on this task in kindergarten has been related to sociometric status in first grade (Rubin & Daniels-Beirness, 1983) and to teacher ratings of social skills (Rubin & Clark, 1983). A prior study had found that children responded similarly on each of three stories, so only one story was used here (Lillard & Else-Quest, 2006). For this task, children were shown a picture of two children of their same race and gender and told the children were their same age. One of the children in the picture was reading and the other looking on. Children were told, "(Reader) has been looking at this book for a long time and (Onlooker) really wants to look at the book. What could (Onlooker) do or say so he/she could have a look at the book?" Children's responses were recorded by hand, and then children were asked, "What else could he/she do or say?" and finally, "What if it was you? What could you do or say so you could have a look at the book?"

We did not use the standard method for scoring children's responses. Instead, based on prior research, we coded the number of references to sharing or fairness (for example, references to how they might trade items or how the reader had had 10 minutes with the book and now could the other person have 10 minutes). Three undergraduate coders were trained to recognize such responses by rating at least 60 responses collected in other studies until their level of agreement with coding of the author was over 90%; then they went on to code the responses collected in this study, with two coders and the author coding every response. Inter-rater reliability, both between the two undergraduate coders and for each coder with the author, was very high ($r > \text{or} = .95$ on 100% of children's responses), and differences were resolved by discussion. The number of sharing and fairness or justice strategies children used among their three responses was summed for a score ranging from 0 to 3.

2.3.4. Reading, vocabulary, and math

Reading, vocabulary, and math were assessed with Form A of the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001; hereafter WJ III). All three WJ III tasks were administered and scored according to the manual with one exception. Because at some of the Montessori schools children first learned cursive letters rather than print, at those schools the initial WJ III testing book pages (the letter identification items) were overlaid with cursive letters. This adaptation has been used in other research (Lillard & Else-Quest, 2006). Scores from all tasks were raw scores.

Reading was assessed with the Letter-Word Identification task in which one has to identify letters and increasingly difficult words. The manual reports a one-year test–retest reliability coefficient of .92 for

children ages 4–7; split-half reliability coefficients are .97, .98, and .99 for 3-, 4-, and 5-year-olds, respectively. Vocabulary was assessed with the Picture Vocabulary task (McGrew & Woodcock, 2001), which has children identify pictures. The manual does not report test-retest reliabilities for this task, but split-half reliability coefficients are .84, .81, and .76 for 3-, 4-, and 5-year-olds. Math achievement was examined via the Applied Problems subtest. This task involves simple counting, addition, and subtraction, reading clock faces, and reporting and calculating coin values. The manual reports a one-year test–retest reliability coefficient of .92 for children ages 4–7, and split-half reliability coefficients of .92, .94, and .92 for 3-, 4-, and 5-year-olds.

2.4. Procedure

Recruitment and other study procedures were approved by the university's internal review board.

2.4.1. Recruitment

Children were recruited from seven different schools, five Montessori and two Conventional. The Montessori schools were selected initially, and for convenience: their geographical locations made them easily accessible for testing. All schools that were invited to participate (via a letter of invitation) elected to do so. Letters were sent home to parents in all the classrooms serving 3- to 6-year-old children. The letter was accompanied by informed consent and basic demographic information forms. Included on the latter form for Montessori parents was the question, "If Montessori were not available, what school would your child attend?" The two most commonly selected private schools were then approached for participation as Conventional programs, and both agreed to participate. About half of all recruited parents from all schools consented.

2.4.2. Testing

Children were tested individually in a quiet area at their school. Five research assistants (3 undergraduate and 2 postgraduate) were trained to test children. Task procedures were practised with at least three children in the laboratory during piloting until the author was satisfied that the assistant was competent to run the procedures in the schools.

Test sessions took 15–25 minutes. The experimenter engaged each child in a few minutes of casual conversation until the child seemed comfortable, then administered the tasks in a set order: Sticker Delay, Theory of Mind Scale, Reading, Wind-Up Delay, Math, HTKS, Social Problem-Solving, and Vocabulary.

2.4.3. Classification of the Montessori classrooms into programs

Although programs were chosen by a subjective judgment of implementation type (Classic or Supplemented), an objective measure was used for confirmation. To objectively classify the Montessori classrooms, four trained observers (two undergraduate and two post-graduate) conducted 5-minute snapshot observations of material use in the classrooms. These observations were typically completed between 9:30 and 11 a.m., with the only stipulation being that it was not completed when children were engaged in a whole class group activity. Three observers were trained to recognize classic Montessori materials by studying the classic Montessori materials in catalogs; the fourth observer had AMI Primary training so knew the materials. Observers entered the classroom during the morning work time twice during the fall and twice during the spring and noted what every child in a classroom was doing by marking a prepared list of common Montessori materials and some common supplemental activities. Observations were conducted by starting on one side of the classroom and marking down each child's activity, then moving towards the center and marking the activity of each child there, and then moving to the other side. Sometimes children moved; each child was counted only once, at their first activity during the observation. Observers checked with the author after conducting an observation when they had seen a material they did not recognize and its categorization was determined through discussion; the author consulted with trained Montessori teachers on rare occasions when needed. Four snapshot observations were made in each Montessori classroom, spread across the year.

Percentages of children present who were engaged in classic Montessori work were averaged across snapshots for each classroom. Interrater reliability was obtained for 20% of observations by a second observer, and agreement on the numbers of children engaged in broad categories of activities (i.e.,

Montessori, Supplemented, can't tell, and off-task) was highly reliable, kappa = .93. Note that in specifying that a child was using a Montessori material, there was no allowance for whether a child was using the material correctly. Incorrect use of the material could of course produce noise in the data, but training all observers on correct usage of each piece of material was beyond the scope of this study.

3. Results

First, Montessori classrooms were classified according to the measure of fidelity (the proportion of children using Montessori materials, averaged across the classroom observations). This step was followed by unconditional Hierarchical Linear Modeling (HLM) to examine whether nesting by classroom was necessary. Because it was not, ANOVAs determined whether programs differed on relevant demographic factors. Next, differences in fall scores were examined for the outcome variables of interest.

With these preliminary analyses in hand, the main research questions were addressed with analyses of gains across the school year. Most of these analyses were done with ANOVAs (using SPSS Version 19), which are appropriate for interval-scaled data. Two of the variables were less aligned with an interval scale: Applied Problems and Theory of Mind. Applied Problems goes from simple addition to clock and coin problems to word problems; there is no sense in which the amount of gain in ability required to pass any two successive problems is the same in scale as the amount of gain required to pass some other two problems. This applies to the Theory of Mind Scale as well: although children pass the tasks in a consistent order, the developmental difference underlying a 0 to 1 gain on the scale might not be equal in magnitude to the difference underlying a gain from 2 to 3. Because of this, for these two tasks nonparametric analyses were also used (Siegel & Castellan, 1988). Finally, regression analyses were conducted to determine whether within the set of all Montessori classrooms, the fidelity measure predicted children's gains.

3.1. Preliminary analyses

3.1.1. A priori classroom differences

The unconditional HLM revealed intraclass correlations of less than .11 for all variables of interest, well below the standard cut-off of .25 that would suggest nested modeling was needed (Guo, 2005).[3] Further, classrooms were demographically matched with respect to key parent variables: ANOVAs indicated no significant differences across the 18 classrooms either in parent education, $F(17,167) = 1.29$, $p = .20$, or family income, $F(16,1240) = 1.21$, $p = .27$. Parent education is a particularly important predictor of child outcomes (NECCRN, 2003) and its equivalence across samples is important.

3.1.2. Classroom classification

In three Montessori classrooms, the percentage of children using Montessori materials ranged from 95%–100%, and in the other nine Montessori classrooms, the range was 38–56%. A one-way ANOVA showed that there were significant differences in the proportion of children using Montessori materials across classrooms, $F(11, 52) = 9.57$, $p < .001$, and follow-up Tukey's tests showed that all three classrooms at one school differed significantly from every classroom at the other four schools (all $ps < .05$), and none of the classrooms at the other four schools differed significantly from each other ($ps > .9$). The percentage of children using Montessori materials thus appeared to be a suitable objective measure by which to classify Montessori programs as Classic or Supplemented.

3.1.3. Potentially confounding classroom variables

Preliminary analyses showed no correlation between years of teaching experience and children's gains across the school year on any of the measures: more experienced teachers were not associated with children showing more or less gain. In addition, within the Supplemented Montessori sample, *t*-tests showed no significant relations between teacher training (AMI vs. AMS) and student outcomes. Thus, teacher training alone was not a significant predictor variable.

[3] Intraclass correlations for the classroom level were as follows: Letter-Word Identification, .05; Picture Vocabulary, <.001; Applied Problems, <.001; Social Problem Solving, <.001 Theory of Mind, .08; and Head Toes Knees Shoulders, .03.

Table 2
Means (and SDs) in fall and spring and mean gains across program types.

Measures	Program type								
	Classic Montessori			Supplemented Montessori			Conventional		
	Fall	Spring	Gain	Fall	Spring	Gain	Fall	Spring	Gain
Letter-Word Identification**	14.08 (8.28)	25.36 (13.38)	11.28 (7.97)	12.07 (10.39)	17.81 (12.83)	5.61 (5.09)	16.45 (9.07)	22.35 (12.10)	5.90 (5.07)
Picture Vocab**	16.50 (3.43)	19.00 (3.05)	2.92 (2.03)	17.18 (3.29)	18.19 (3.09)	0.95 (1.95)	19.00 (3.05)	20.13 (3.29)	1.08 (2.22)
Applied Problems^{t}	15.78 (4.49)	20.36 (4.36)	4.58 (3.12)	15.71 (4.92)	18.88 (5.48)	3.06 (2.97)	18.00 (3.7)	21.53 (4.79)	3.54 (2.39)
Head-Toes-Knees-Shoulders**	22.06 (14.34)	35.78 (4.70)	13.72 (11.42)	17.30 (13.86)	24.47 (12.27)	7.22 (10.99)	24.93 (11.78)	32.60 (8.21)	7.67 (7.95)
Social Problem Solving*	.22 (.54)	.56 (.94)	.33 (1.01)	.42 (.74)	.43 (.75)	.01 (.97)	.35 (.86)	.30 (.61)	−.05 (.68)
Theory of Mind	1.50 (0.78)	1.89 (0.92)	0.39 (0.84)	1.25 (0.79)	1.47 (0.89)	0.22 (0.81)	1.63 (0.93)	1.75 (0.95)	0.13 (1.27)

Note. Differences in gains favoring Classic Montessori over one or both other types of program: $^{*}p = .05$; $^{**}p < .02$; $^{t}\ p < .05$ by Kruskal–Wallis rank test.

An ANOVA revealed a significant difference in the ages of children across program types, $F(2, 169) = 3.67$, $p = .03$. Children in the Conventional programs were significantly older on average (59.2 months at fall test) than children in the Supplemented Montessori programs (54.0 months), Tukey's HSD = 5.16, $p = .02$, so age was used as a covariate in linear analyses. There were no age differences between the Classic Montessori programs (55.6 months) and either of the other two types: Tukey's HSD was not significant between Classic and Supplemented ($p = .71$) and between Classic and Conventional ($p = .28$). As shown in Table 1, gender was distributed similarly across the three types of programs, $\chi^2(2) = 2.87$, $p = .24$.

3.1.4. Fall scores

Only one variable was different across programs at the start of the study (see Table 2): Children attending Conventional programs had higher fall vocabulary scores: $M = 19.00$, $SD = 3.05$; the Supplemented Montessori had the next highest, $M = 17.18$, $SD = 3.29$, and the Classic Montessori had the lowest, $M = 16.50$, $SD = 3.43$, $F(2, 169) = 4.50$, $p = .01$, partial $\eta^2 = .05$. Simple Planned Contrasts (Rosenthal & Rosnow, 1985) showed a significant difference of 1.8 ($p < .01$) between the Conventional and Classic Montessori programs. Because vocabulary score is highly correlated with intelligence scores (McGrew & Woodcock, 2001), subsequent analyses included fall Picture Vocabulary scores as a covariate. The Sidak procedure was used to control for Type I error in this and all Planned Contrasts reported here.

3.2. School-year gains

Analyses of covariance (controlling for fall age, fall score on that outcome, test interval, and fall Picture Vocabulary score) revealed differences in school year gains by program type for four of the six outcome variables. One-tailed significance levels were applied for tasks on which differences were found in a prior study: Letter-Word, Applied Problems, and Social-Problem Solving (Lillard & Else-Quest, 2006).

3.2.1. Letter-Word Identification

Children in Classic Montessori programs gained an average of 11.28 points ($SD = 7.97$), whereas children in Supplemented Montessori programs gained an average of 5.61 points ($SD = 5.09$) and children in Conventional programs gained an average of 5.90 points ($SD = 5.07$) across the school year. These gain scores were significantly different across programs, $F(2, 168) = 5.02$, $p < .01$, partial $\eta^2 = .06$.[4] Simple Planned Contrasts showed a significant difference between Classic Montessori and both other types of program (difference = 4.47, $p = .002$ between Classic and Supplemented Montessori, with a difference of 5.19, $p = .006$ between Classic Montessori and Conventional).

3.2.2. Picture Vocabulary

Children in Classic Montessori programs gained an average of 2.92 points ($SD = 2.03$) in Picture Vocabulary scores, children in Supplemented Montessori programs an average of 0.95 ($SD = 1.95$), and children in Conventional programs an average of 1.08 ($SD = 2.22$), $F(2, 168) = 4.41$, $p = .01$, partial $\eta^2 = .05$. Simple Planned Contrasts showed a significant difference of 1.07, $p = .04$, between the Classic Montessori and Supplemented Montessori programs. The Conventional programs did not differ from Classic Montessori.

3.2.3. Applied Problems

On the Applied Problems task, children in Classic Montessori programs gained an average of 4.58 points ($SD = 3.12$), children in Supplemented Montessori programs gained an average of 3.09 points ($SD = 2.98$), and children in Conventional programs gained an average of 3.53 points ($SD = 2.36$). Because raw scores on the Applied Problems task do not reflect an interval scale, the Kruskal–Wallis test was conducted, revealing a significant difference ($p = .02$) whereby children in Classic Montessori programs gained more in scores over the year than the children in Supplemented Montessori programs, Mann–Whitney $U = 1220$, $p = .01$.

[4] Effect sizes for partial η^2 are interpreted as follows: *small* = 0.01, *medium* = 0.06, and *large* = 0.14 (Kittler, Menard, & Phillips, 2007; see also Cohen, 1988; Olejnik & Algina, 2000).

3.2.4. Head-Toes-Knees-Shoulders

Children in Classic Montessori programs gained an average of 13.72 points ($SD = 11.42$), whereas children in Supplemented Montessori programs gained on average 7.34 points ($SD = 10.92$), and those in Conventional programs 7.85 points, $SD = 7.85$, $F(2, 167) = 18.12$, $p < .001$, partial $\eta^2 = .18$. Simple Planned Contrasts showed the significant differences were between the Classic Montessori and both the Supplemented Montessori (11.03, $p < .001$) and Conventional programs (7.60, $p = .004$).

3.2.5. Social problem solving

Children in the Classic Montessori programs increased references to sharing and justice by an average of 0.33 points ($SD = 1.01$), whereas children in the Supplemented Montessori programs did not change, $M = 0.01$, $SD = 0.96$, and those in the Conventional programs declined in the number of such references, $M = -0.07$, $SD = 0.69$. The analysis indicated a small-medium effect size, $F(2, 170) = 2.30$, $p = .03$, partial $\eta^2 = .03$; Simple Planned Contrasts showed a significant difference between the Classic Montessori and the Conventional programs (0.54, $p = .04$).

3.2.6. Theory of mind

Children in the Classic Montessori programs gained an average of 0.39 points (of a possible 3, $SD = 0.84$) on the Theory of Mind scale, children in Supplemented Montessori programs an average of 0.26 points ($SD = 0.84$), and children in Conventional programs an average of 0.12 points ($SD = 1.25$). These differences were not statistically significant.

3.3. Prediction of Montessori scores from fidelity

The third set of analyses examined whether the proportion of children in one's classroom who were using Montessori materials at snapshots predicted spring scores, after taking children's fall scores and

Table 3
Stepwise regression on spring scores entering fall scores, inter-test interval, and proportion of children using Montessori materials.

		Model 1		Model 2		Model 3	
	ΔR^2	*b (SEb)*	β	*b (SE b)*	β	*b (SE b)*	β
Head-Toes-Knees-Shoulders spring score							
Step 1	.39**						
Fall score		.53 (.06)	.63**	.51(.06)	.60	.49(.06)	.58
Step 2	.03*						
Test interval				2.47 (1.02)	.17*	−1.62(1.44)	−.11
Step 3	.06**						
Montessori work						4.41 (1.15)	.37**
Total R2	.48**						
Letter-Word Identification spring score							
Step 1	.78**						
Fall score		1.21 (.06)	.89**	1.22 (.05)	.89**	1.19 (.05)	.87**
Step 2	.02**						
Test interval				2.48(.65)	.15	.43(.96)	.03
Step 3	.01**						
Montessori work						2.19(.76)	.16**
Total R2	.82**						
Picture Vocabulary spring score							
Step 1	.61**						
Fall score		.78(.05)	.78**	.80 (.05)	.81**	.80 (.05)	.81**
Step 2	.06**						
Test interval				1.04(.21)	.25**	.45(.32)	.11
Step 3	.02*						
Montessori work						.61(.25)	.19*
Total R2	.69**						

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

inter-test interval into account. The proportion of children using Montessori materials was entered into the regression at the third step, with fall score at the first step, and inter-test interval at the second step. These analyses showed that degree of materials usage within Montessori classrooms was a significant predictor for three of the six outcome variables: Head-Toes-Knees-Shoulders, Letter-Word Identification, and Picture Vocabulary (see Table 3). Thus, even within the set of Montessori programs, the more children in one's class engaged in Montessori work, the more one gained over the school year on these three variables.

4. Discussion

Although Montessori is a fairly common alternative form of schooling, there is little research on its outcomes, and what research exists is inconsistent. This study sought to examine whether different child outcomes are associated with differences in implementation fidelity. Across a variety of outcomes, children attending Classic Montessori programs gained significantly more across the school year than did demographically similar children in Supplemented Montessori programs and Conventional programs in schools that their parents said they would enroll the children in were Montessori programs unavailable. Importantly, at the beginning of the school year, there was no advantage for children in Classic Montessori programs, suggesting that unobserved selection effects are not at play. The larger gains were accrued on outcomes ranging from executive function and social problem-solving to common academic achievement measures. These findings are important because early advantages in these areas predict success in school and social relations (Baron-Cohen, 2000; Mischel et al., 1989; Pianta, La Paro, Payne, Cox, & Bradley, 2002; Razza, 2005; Watson et al., 1999). In addition, among the Montessori programs, the percentage of children in each classroom using Montessori materials predicted executive function, early reading, and vocabulary. For the other variables, this measure of fidelity was not predictive. Closer examination of what children were doing in each classroom might be necessary to determine why these variables and not the others were predicted by the measure.

From this study, Montessori education appears to reflect a pattern noted in the fidelity of implementation literature. In this literature, programs last to the degree that they are flexible and adaptive (Goodman & Steckler, 1989). Montessori has significant representation in America after 100 years, whereas many other alternative school programs that existed 100 years ago are no longer very visible—for example, one rarely sees Froebel or Pestalozzi or Dewey or Dalton Plan schools today (Beatty, 1995; Cunningham & Duffy, 1996; Edwards, 1991). American Montessori education's very adaptability, including openness to supplementary materials, might be key to its endurance. On the other hand, the fidelity of implementation literature also shows that lower fidelity is associated with less positive outcomes (O'Donnell, 2008), and this study indicates this might also be the case for Montessori. It is ironic that the very element that might enhance survival of a program might also reduce its effectiveness, although it is important to note that the supplemented Montessori programs were generally on a par with the highly regarded conventional comparison programs studied here. In the section that follows, I discuss each result in turn, examining some elements of the Classic Montessori program that I hypothesize might have contributed.

4.1. Executive function

The most notable result in this study was the school-year gain in executive function observed in children in the Classic Montessori program. In prior studies, children in this age range in conventional public preschool programs gained about 2 to 5 points on this task over a school year (Ponitz et al., 2008, 2009), whereas children in the private Supplemented and Conventional programs in this study gained 7 to 8 points. By contrast, children in the Classic Montessori program gained almost 14 points, and the statistical tests indicated a large effect size (Cohen, 1988). This gain in executive function is of practical importance, because children with stronger executive function skills in kindergarten are concurrently and subsequently more academically (Blair & Razza, 2007; Bull et al., 2008) and socially competent (Riggs et al., 2006). Increasingly executive function skills are seen as key not only to school readiness but to success in life (Blair, 2002; Diamond & Lee, 2011).

Theoretically, using Montessori materials would seem to exercise many aspects of executive function. For example, one of the first Montessori materials with which a child is presented is the Pink Tower, a set

of 10 wooden pink blocks ranging in size from a 1-cm cube to a 10-cm cube, with each cube 1-cm larger on each face than the previous one. In using this material, the children's task is to carry the cubes one by one from a display to a rug that they have previously rolled out on the floor, then rebuild the tower. To do this task entails planning. Second, each time a child chooses a block, she or he must do so with reference to its relation to all the other blocks: Is there another one in between the size of this one and the last one placed on the tower? This step requires working memory. Third, the child must inhibit the prepotent tendency to grab the closest block, and fourth, the child must pay strict attention to how he or she places each block on the one below it, creating a symmetrical tower. After building the tower, the child takes it down, returns the blocks to their stand by the shelves (in the proper order), and then tightly rolls up the rug and returns it to its place. This step requires flexibility and task switching. Consider the difference between this and engaging with ordinary blocks. With ordinary blocks, one can do anything, without necessarily having any set plan, and one does not have to think about the blocks in relationship to each other. A preschool might not have a requirement that children put items away right after use (instead, there often is a single clean-up time right before going home), and there may well be no set way to arrange blocks when returning them to their place (often, they get put haphazardly into a large basket or box). The executive function demands are much reduced, and this difference in executive function demands applies across many other activities as well.

4.2. Reading and vocabulary

Children in the Classic Montessori programs also gained the most in early reading and vocabulary, with effect sizes of .05 and .06. Although not nearly as large as the effect size for executive function, these are considered medium effect sizes (Cohen, 1988). In raw terms, the gains in Letter-Word Identification scores for the children in Classic Montessori programs were twice what they were for the other two groups (11 vs. 5–6). The latter gains just exceeded those seen in a large scale public prekindergarten study in Oklahoma (Gormley, Gayer, Phillips, & Dawson, 2005), in which the gain of about 5 points rendered a large effect size of .80 compared to no preschool. Thus, it seems that the score gain seen in children in Classic Montessori programs, which was twice that, is certainly meaningful. Early gains of this magnitude would be expected to have lasting benefits to children and society (Camilli, Vargas, Ryan, & Barnett, 2010; Pianta, Barnett, Burchinal, & Thornburg, 2009).

One possible reason for the gains in Letter-Word Identification is the materials used to teach writing and reading in Montessori programs, and the fact that the children were using those materials more often in Classic than in Supplemented programs. The conventional school programs in this study taught reading by teaching children to identify letters and words at circle time; they also provided plentiful literacy materials (e.g., magnetic letters, books, and writing workbooks, the latter of which were also available in Supplemented programs). Montessori reading instruction begins by having children trace sandpaper letters with their index finger while reciting the phonetic sounds, and then also putting words together with a cardboard Moveable Alphabet; these materials are used before children learn to read. The phonetic approach to reading is well-supported (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Haptics or using the hand to trace letters has been much less studied, but findings from embodied cognition would suggest it would be helpful (see Lillard, 2005), and at least one study finds tracing letters improves their learning (Bara, Gentaz, & Cole, 2007). For Picture Vocabulary, emphasis on nomenclature in Montessori might play a role. Programs with abundant supplementary materials would have fewer children engaged in the Montessori nomenclature exercises (for example, creating labels and placing them besides the objects they name).

4.3. Social problem-solving

Children in Classic Montessori programs showed larger gains in using more mature social problem solving strategies, particularly ones centering on justice and taking another person's goals into consideration. For example, children in Classic Montessori programs were more likely to offer object acquisition strategies like, “Can I please look at that book while you are looking at another book you might want to look at?” and “Maybe we could each have turns for 10 minutes?” One possible reason why children in Classic Montessori programs might show greater school-year gains in the use of these strategies is the

provision of only one set of each Montessori material, which might create frequent needs to consider how to resolve conflicts over limited resources. Social competence in preschool is associated with better outcomes in social and academic domains (Capage & Watson, 2001; Ladd et al., 1999; Malecki & Elliott, 2002; Welsh et al., 2001), and thus this finding seems to be of practical significance.

4.4. Theory of mind

One domain in which gains were not significant was Theory of Mind. This is interesting in part because theory of mind is related to executive function (Carlson, Moses, & Claxton, 2004), so much so that some consider it an executive function task (Blair & Razza, 2007), and performances on both types of task (a similar Executive Function task and False Belief) were higher in Montessori children in the low-income lottery study (Lillard & Else-Quest, 2006). For these middle-income children, program type made no difference in performance on Theory of Mind tasks.

4.5. Math

The Applied Problems task showed significant program differences with a nonparametric rank test, appropriate because score intervals for this task are not uniform. The extent and sophistication of the Montessori math materials might be expected to result in differences, and they have been found in prior research (Dohrmann et al., 2007; Lillard & Else-Quest, 2006). A consideration for why only the rank test revealed differences concerns the Applied Problems task itself. For children in Conventional and in Classic Montessori programs, the spring scores averaged 20 to 21 points. At a score of 19, the Applied Problems task presents its first clock face, and at item 24, it begins to present pictures of coins. If a child has not had a specific lesson on decoding these cultural artifacts, the child will not pass these items no matter how adequate the child's basic math skills. In between these items, Applied Problems embarks on word problems with extraneous information: a child must know what to ignore. This ignoring of extraneous information is a sophisticated skill that could well stump young children who have not specifically been taught that adults sometimes give problems with extraneous information. A second concern with the Applied Problems task is that its early problems assess superficial computation, and although the Montessori program was associated with better simple computation skills in low-income populations, more conceptual tasks might be needed to reveal differences at higher levels of the task, such as were achieved in middle-income samples here.

4.6. Fall scores

Looking across the data set generally, the equality of fall scores is puzzling. If Classic Montessori does produce more school-year gains, one might expect 4- and 5-year-old children, in their second and third years in these programs, to be different at the fall test as well. We see three possibilities. First, the scores of children in Classic Montessori programs were higher in the fall than those of children in other programs, but when combined with the scores of the 3-year-old children, the averages are not different. Unfortunately, the sample sizes for some narrow age groups were too small to render meaningful statistical results, but children in Classic Montessori programs scored slightly higher on Letter-Word Identification and Head-Toes-Knees-Shoulders at 4 and 5 but not at 3 years of age. A second possibility is that the older children's scores really are equal in the fall, implying that school-year gains are not sustained over summer vacation. A meta-analysis of studies of the impact of summer vacation on achievement test scores found that at least older students (4th to 9th graders) generally lose about a month's portion of their school-year gain score over the summer, although middle-income children do not show a loss in reading scores (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). A third but related possibility that would explain equal means at older ages, but only in this study, stems from fall test date: by necessity children in the Classic Montessori were tested earlier in the school year than children in the other programs. This potential confound was controlled for statistically in terms of its possible impact on gain scores, but it might have masked fall differences in the older groups if being less well adjusted to school after summer vacation leads to lower scores. Further research could address this by including larger numbers of children at each age level so each age level can be examined separately, and by testing the same children at the end of one school year and then repeatedly at the same time for the first few months of the next school year to see

whether there is a difference in summer reversion for children at different age levels in different types of school program. Regardless, in this study differences in gains across programs were observed, taking fall scores and test–retest interval into account.

4.7. Implications

These results are important for parents considering school choices, especially since Montessori education is increasingly offered in public school districts. Although Montessori education is growing, there is a paucity of research on its outcomes (Walsh & Petty, 2007), and what exists is contradictory. These results suggest a source of the inconsistencies could be implementation, indexed here by use of Montessori materials. These results could also spur research promoting better understanding of children's development in educational settings, because they suggest that some elements of classic Montessori preschool programs are particularly helpful to development. Whereas Classic Montessori was associated with larger gains than were seen in either Supplemented Montessori or Conventional programs, it still could be the case that even some Conventional programs could be improved by incorporating certain Montessori-like practices, such as use of haptic materials for early reading (rather than workbooks) and more executive-function-taxing activities. Exactly what those elements would be is a topic for further research.

This research also has important implications for teacher training. Implementing a program to a high standard of fidelity requires deep understanding of that program, which can take months or even years of study depending on the depth and complexity of the system. Teachers who fully understand the materials offered in a given program that supplies a broad, inclusive, and intertwined set, as does Montessori, should be less inclined to try to supplement those materials with others that do not fit into the set. Teachers who take less intensive and more superficial training courses might be more inclined to supplement materials, possibly resulting in less effective classrooms.

Finally, this research connects with a recent emphasis in school psychology on what is termed the 3-tiered model of intervention, where the top tier is individualized instruction for individual students, the middle is interventions for groups of students with special needs, and the bottom or universal tier is for students generally (Tilly, 2008). This research is aimed at that universal tier, examining the core curriculum of Montessori education and its impact on child outcomes, both academic and social-behavioral. Research like this across a variety of educational programs, from Steiner to Responsive Classrooms, looking at how universal instruction and the integrity of an individual program is related to child outcomes, is important as schooling practices become increasingly evidence-based. This research makes one contribution to this base.

4.8. Limitations

A significant weakness of this study is lack of random assignment to program. Several analyses were done to mitigate this concern: the conventional schools were selected as ones that parents who chose Montessori education would also have chosen; analyses showed no differences in parent education or income across classrooms; neither teacher experience nor type of Montessori teacher training was associated with child outcomes; and there were no differences in children's scores in the fall except on Picture Vocabulary, and this difference favored the Conventional programs and was subsequently used as a covariate. Despite these efforts, nonrandom designs cannot rule out third variables. Perhaps particular parents are more attracted to classic Montessori. Parents are known to have tremendous influence on child outcomes (NECCRN, 2004; Shonkoff & Phillips, 2000), and for obvious reasons: they confer genes to their children, provide the home environment where most children spend most of their time, select neighborhoods, and pick schools. The ultimate test for whether parents are responsible for the differences between schools programs found here would have children who were placed by lottery in these schools, as in Lillard and Else-Quest (2006) that found a similar pattern of differences comparing low-income children. Further studies using randomized controls are necessary to allow for causal conclusions.

A second limitation of this study is that we know little about what was actually happening in the classrooms that might have contributed to the results. Montessori classrooms were classified by the number of children using Montessori materials, but this classification could simply be an indicator variable for other differences that might have been more central contributors to children's outcomes. Further research must examine what aspects of the classic program are important to children's gains. For example, it is possible

that teachers who choose not to bring supplementary materials into their classroom also interact with children differently; other studies have shown that teacher interaction (particularly warmth and sensitivity, and asking thought-scaffolding questions) is a major factor in children's achievement (Hamre & Pianta, 2007; Mashburn et al., 2008). Further studies should expand classroom observations to include other features like styles of teacher interaction that might go along with eschewing supplementary materials, and which might be responsible for the larger gains seen here. An alternative but promising strategy would be to randomly assign supplemented Montessori programs to retain or be stripped of their supplementary materials and measure subsequent gains. A small pilot study did this, and found significantly larger gains in the stripped classrooms in Letter-Word Identification and Head-Toes-Knees-Shoulders over just 4 months (Lillard, unpublished raw data), but sample sizes were small.

A third limitation is that the sample here was well-educated, middle- to high-income, and mostly White, and one might argue that the findings are relevant only to this group. The Lillard and Else-Quest (2006) study and the Dohrmann et al. (2007) study described in the Introduction showed benefits of classic Montessori for low-income children as well. Dohrmann et al. (2007) showed that children who had previously gone to Milwaukee public Montessori schools, many of whom were on free lunch programs, scored higher on math and science composite score in high school. Lillard and Else-Quest (2006) showed that lower-income children who were randomly chosen to attend Montessori or went to other schools did better on a range of academic and social measures at ages 5 and 12. This study even showed benefits on many of the same outcomes as were observed in the middle-income children in the Classic Montessori programs studied here. A final limitation is that this study included only children from 3 to 6 years of age. The extent to which these early gains are carried over as children transfer into elementary school and beyond is an important question for further research.

4.9. Conclusion

The research presented here suggests that Classic Montessori is associated with significant gains in student achievement and development relative to Supplemented Montessori and highly regarded conventional school programs. What aspects of the Classic Montessori programs might have been responsible for the advantages is not clear; further research could shed light on this by randomly assigning supplemented programs to remove their additional materials and by more closely observing the micro level interactions of teachers and students in such environments.

As early childhood educators search for ways to improve the academic and social outcomes of children in American schools, Montessori education might be worthy of more consideration (Walsh & Petty, 2007). But Classic Montessori is the rarer form; a recent survey of 85 public Montessori schools in the United States (about a third of the total number) showed that only 28% of school heads strongly agreed with the statement, “Our school implements elementary education according to the original vision of Maria Montessori” (Murray & Peyton, 2009, p. 28). With any alternative educational program, fidelity is an important consideration. Variations could in principle improve or demote program quality; the present research suggests that a very common variation in Montessori programs, supplementing the core set of materials with ones typically used in conventional preschool programs, or something associated with their use, can demote it. School districts and parents considering Montessori education should be aware of these differences in implementation and their possible effects. Montessori education appears to reflect a pattern in the fidelity of implementation literature, whereby the very feature that might be responsible for its strong representation 100 years after its development—its adaptability—might also compromise its effectiveness.

References

American Montessori Society (2009). *School accreditation standards and criteria.* Retrieved July 2, 2011, from. http://www.amshq.org/School%20Resources/AMS%20%20School%20Accreditation.aspx

American Montessori Society (2011). *AMS and the Montessori Movement: A new emphasis.* Retrieved July 2, 2011, from. http://www.amshq.org/Montessori%20Education/History%20of%20Montessori%20Education/AMS%20and%20the%20Montessori%20Movement.aspx

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science*, *11*(6), 862–868.

Bara, F., Gentaz, E., & Cole, P. (2007). Haptics in learning to read with children from low socio-economic status families. *British Journal of Developmental Psychology*, *25*, 643–663.

Baron-Cohen, S. (2000). Autism. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1167–1270). (2nd ed.). Cambridge, MA: MIT Press.
Beatty, B. (1995). *Preschool education in America: The culture of young children from the colonial era to the present.* New Haven, CT: Yale University Press.
Besancon, M., & Lubart, T. (2008). Differences in the development of creative competencies in children schooled in diverse learning environments. *Learning and Individual Differences*, *18*, 381–389, doi:10.1016/j.lindif.2007.11.009.
Blair, C. (2002). School readiness. *American Psychologist*, *57*, 111–127.
Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, *20*, 899–911.
Blair, C., Knipe, J., Cummings, E., Baker, D. P., Gamson, D., Eslinger, P., et al. (2007). A developmental neuroscience approach to the study of school readiness. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten* (pp. 149–174). Baltimore, MD: Brookes.
Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663.
Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, *15*, 253–268.
Bronson, M. (2001). *Self-regulation in early childhood: Nature and nurture.* New York, NY: Guilford Press.
Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228.
Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*, 579–620.
Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education & Development*, *12*, 613–628.
Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*, 299–319.
Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 40–49.
Cohen, J. (1988). *Statistical power for the behavioral sciences.* Hillsdale, NJ: Erlbaum.
Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*, 227–268.
Cossentino, J. (2005). Ritualizing expertise: A non-Montessorian view of the Montessori method. *American Journal of Education*, *111*, 211–244.
Cox, M. V., & Rowlands, A. (2000). The effect of three different educational approaches on children's drawing ability: Steiner, Montessori and traditional. *British Journal of Educational Psychology*, *70*, 485–503.
Cunningham, D., & Duffy, T. (1996). Constructivism: Implications for the design and delivery of instruction. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 170–198). New York: Simon & Schuster.
Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*, 1387–1388.
Diamond, A., & Lee, K. (2011). How can we help children succeed in the 21st century? What the scientific evidence shows aids executive function development in children 4–12 years of age. *Science*, *333*, 959–964.
Dohrmann, K. R., Nishida, T. K., Gartner, A., Lipsky, D. K., & Grimm, K. J. (2007). High school outcomes for students in a public Montessori program. *Journal of Research in Childhood Education*, *22*, 205–217.
Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology*, *36*, 161–174.
Dunn, J., & Hughes, C. (2001). "I got some swords and you're dead": Violent fantasy, antisocial behavior, friendship, and moral sensibility in young children. *Child Development*, *72*, 491–505.
Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350.
Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, *78*, 558–580.
Edwards, J. (1991). To teach responsibility, bring back the Dalton plan. *Phi Delta Kappan*, *72*(5), 398–401.
Epstein, A. S. (2006). *Essentials of active learning in preschool: Getting to know the High/Scope curriculum.* Ypsilanti, MI: High/Scope Press.
Erwin, P. (1993). Social problem solving, social behavior, and children's peer popularity. *The Journal of Psychology*, *128*, 299–306.
Espy, K., McDiarmid, M., Cwik, M., Stalets, M., Hamby, A., & Senn, T. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*, 465–486.
Farrant, B., Fletcher, J., & Maybery, M. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and the developmental role of language. *Child Development*, *77*, 1842–1853.
Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60.
Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. (2007). What brains are for: Action, meaning, and reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 221–240). Mahwah, NJ: Erlbaum.
Goodman, R. M., & Steckler, A. (1989). A model for the institutionalization of health promotion programs. *Family & Community Health*, *11*, 63–78.
Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*, 872–884.
Graue, E. (2006). The answer is readiness—Now what is the question? *Early Education & Development*, *17*, 43–56.
Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*, 361–396.
Guo, S. (2005). Analyzing grouped data with hierarchical linear modeling. *Children and Youth Services Review*, *27*, 637–652.
Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten* (pp. 49–84). Baltimore, MD: Brookes.

Hojnoski, R. L., Margulies, A. S., Barry, A., Bose-Deakins, J., Sumara, K. M., & Harman, J. L. (2008). Analysis of two early childhood education settings: Classroom variables and peer verbal interaction. *Journal of Research in Childhood Education*, *23*, 193–209.
Hongwanishikul, D., Happeny, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*, 617–644.
Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*, 27–50.
Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233–253.
Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test–retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, *41*, 483–490.
Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036.
Humphryes, J. (1998). The developmental appropriateness of high-quality Montessori programs. *Young Children*, *53*, 4–16.
Karnes, M., Shewedel, A., & Williams, M. (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. 133–171). Hillsdale, NJ: Erlbaum.
Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765–778.
Kittler, J. E., Menard, W., & Phillips, K. A. (2007). Weight concerns in individuals with body dysmorphic disorder. *Eating Behaviors*, *8*, 115–120.
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*, 637–658.
Kramer, R. (1976). *Maria Montessori: A biography*. New York, NY: Putnam.
Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373–1400.
Lillard, A. S. (2005). *Montessori: The science behind the genius.* New York, NY: Oxford University Press.
Lillard, A. S. (2011a). Materials: What belongs in a Montessori primary classroom? Results from a survey of AMI and AMS teacher trainers. *Montessori Life*, *22*, 18–32.
Lillard, A. S. (2011b). Mindfulness practices in education: Montessori's approach. *Mindfulness*, *2*, 78–85, doi:10.1007/s12671-011-0045-6.
Lillard, A. S., & Else-Quest, N. (2006). The early years: Evaluating Montessori education. *Science*, *313*, 1893–1894.
Lillard, A. S., & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics*, *128*, 644–649, doi:10.1542/peds.2008-1536.
Lopata, C., Wallace, N. V., & Finn, K. V. (2005). Comparison of academic achievement between Montessori and traditional education programs. *Journal of Research in Childhood Education*, *20*, 5–13.
Malecki, C., & Elliott, S. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*, 1–23.
Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*, 732–749.
McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471–490.
McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*, 947–959.
McDonald Connor, C., Son, S., Hindman, A., & Morrison, F. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, *43*, 343–375.
McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual.* Itasca, IL: Riverside Publishing.
Miller, L. B., & Bizzell, R. P. (1984). Long-term effects of four preschool programs: Ninth- and tenth-grade results. *Child Development*, *55*, 1570–1587.
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.
Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*, 933–938.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
Montessori, M. (1966). The secret of childhood (M. J. Costello, Trans.). New York, NY: Ballantine.
Montessori, M. (1967). The discovery of the child. (M. J. Costello, Trans.). New York, NY: Ballantine.
Montessori, M. (1989a). Creative development in the child I (R. Ramachandran, Trans.). Madras, India: Kalakshetra Press.
Montessori, M. (1989b). Creative development in the child II (R. R, Trans.). Madras, India: Kalakshetra Press.
Montessori, M. (1989c). *The child, society, and the world: Unpublished speeches and writings.* (Vol. 7). Oxford, England: Clio.
Montessori, M. (1997). *The California lectures of Maria Montessori, 1915.* Oxford, England: Clio.
Murray, A., & Peyton, V. (2009). Public Montessori elementary schools: A delicate balance. *Montessori Life*, *20*, 26–30.
National Early Child Care Research Network (2002). Child-care structure–>process–>outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, *13*, 199–206.
National Early Child Care Research Network (2003). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development*, *74*, 1639–1662.
National Early Child Care Research Network (2004). Multiple pathways to early academic achievement. *Harvard Educational Review*, *74*, 1–28.
Odom, S. L. (2009). The tie that binds. *Topics in Early Childhood Special Education*, *29*, 53–62.
O'Donnell, M. (2007). *Continuum library of educational thought: Maria Montessori.* (Vol. 7). London, England: Continuum.
O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241–286.
Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development*, *22*, 165–184.
Peterson, C. C., & Wellman, H. M. (2009). From fancy to reason: Scaling deaf and hearing children's understanding of theory of mind and pretence. *British Journal of Developmental Psychology*, *27*, 297–310.
Pettit, G., Dodge, K., & Brown, M. (1988). Early family experience, social problem solving patterns, and children's social competence. *Child Development*, *59*, 107–120.
Piaget, J. (1970). Science of education and the psychology of the child (D. Coltman, Trans.). New York, NY: Orion Press.
Pianta, R. C., Barnett, W., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education. *Psychological Science in the Public Interest*, *10*, 49–88.
Pianta, R. C., Cox, M. J., & Snow, K. L. B. (Eds.). (2007). *School readiness and the transition to kindergarten in the era of accountability*. Baltimore, MD: Brookes.
Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, *102*, 225–238.
Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes: Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, *23*, 141–158, doi:10.1016/j.ecresq.2007.01.004.
Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*, 605–619, doi:10.1037/a0015365.
Povell, P. (2009). *Montessori comes to America.* Lanham, MD: University Press of America.
Rathunde, K. R., & Csikszentmihalyi, M. (2005a). Middle school students' motivation and quality of experience: A comparison of Montessori and traditional school environments. *American Journal of Education*, *111*, 341–371.
Rathunde, K. R., & Csikszentmihalyi, M. (2005b). The social context of middle school: Teachers, friends, and activities in Montessori and traditional school environments. *The Elementary School Journal*, *106*, 59–79.
Raver, C. C., Garner, P. W., & Smith-Donald, R. (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten* (pp. 121–148). Baltimore, MD: Brookes.
Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*, 31–74.
Razza, R. A. (2005). Relations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *66*(5-B), 2859.
Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology*, *27*, 300–309.
Rodriguez, L., Irby, B. J., Brown, G., Lara-Alecio, R., & Galloway, M. M. (2005). *An analysis of second grade reading achievement related to pre-kindergarten Montessori and transitional bilingual education.* Mahwah, NJ: Erlbaum.
Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance.* Cambridge: Cambridge University Press.
Rubin, K. (1988). *The Social Problem-Solving Test-Revised.* Waterloo, Ontario: University of Waterloo.
Rubin, K., & Clark, M. L. (1983). Preschool teachers' ratings of behavioral problems: Observational, sociometric, and social-cognitive correlates. *Journal of Abnormal Child Psychology*, *11*, 273–286.
Rubin, K., & Daniels-Beirness, T. (1983). Concurrent and predictive correlates of sociometric status in kindergarten and grade 1 children. *Merrill-Palmer Quarterly*, *29*, 337–351.
Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influcences on the development of executive attention. *Proceedings of the National Academy of Science*, *102*, 14931–14936.
Salovey, P., & Mayer, J. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*, 185–211.
Schwarz, J. C., Schrager, J. B., & Lyons, A. E. (1983). Delay of gratification by preschoolers: Evidence for the validity of the choice paradigm. *Child Development*, *54*, 620–625.
Shapiro, D. (2007, June). How many Montessori schools? *Jola Newsletter*, *2*, 1.
Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development.* New York, NY: National Academy Press.
Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York, NY: McGraw-Hill.
Spivak, G., & Shure, M. B. (1974). *Social adjustment of young children.* San Francisco, CA: Jossey-Bass.
Tager-Flusberg, H. (2003). Exploring the relationship between theory of mind and social-communicative functioning in children with autism. In B. Repacholi, & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 197–212). New York, NY: Psychology Press.
Thorell, L., Lindqvist, S., Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*, 106–113.
Tilly, W. D., III (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. *Best practices in school psychology*, *Vol. V*. (pp. 17–36)Bethesda, MD: National Association of School Psychologists.
Toner, I. J., Holstein, R. B., & Hetherington, E. M. (1977). Reflection-impulsivity and self-control in preschool children. *Child Development*, *48*, 239–245.
Walsh, B. A., & Petty, K. (2007). Frequency of six early childhood education approaches: A 10-year content analysis of early childhood education journal. *Early Childhood Education Journal*, *34*, 301–305, doi:10.1007/s10643-006-0080-4.
Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental Psychology*, *35*, 386–391.
Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: Bradford Books/MIT Press.
Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*, 1075–1081.
Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *83*, 780–792.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541.
Welsh, M., Parke, R., Widaman, K., & O'Neil, R. (2001). Linkages between children's social and academic competence a longitudinal analysis. *Journal of School Psychology*, *39*, 463–482.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside Publishing.