Seeking Racial and Ethnic Parity in Preschool Outcomes: An Exploratory Study of Public Montessori Schools vs. Business-as-Usual Schools

Angeline S. Lillard[1], Xin Tong[1], and Paige M. Bray[2]
[1] University of Virginia, [2] University of Hartford

Keywords: *preschool, racial equity, child outcomes*

Abstract: Montessori pedagogy is a century-old, whole-school system increasingly used in the public sector. In the United States, public Montessori schools are typically Title I schools that mostly serve children of color. The present secondary, exploratory data analysis examined outcomes of 134 children who entered a lottery for admission to public Montessori schools in the northeastern United States at age 3; half were admitted and enrolled and the rest enrolled at other preschool programs. About half of the children were identified as White, and half were identified as African American, Hispanic, or multiracial. Children were tested in the fall when they enrolled and again in the subsequent three springs (i.e., through the kindergarten year) on a range of measures addressing academic outcomes, executive function, and social cognition. Although the Black, Hispanic, and multiracial group tended to score lower in the beginning of preschool in both conditions, by the end of preschool, the scores of Black, Hispanic, and multiracial students enrolled in Montessori schools were not different from the White children; by contrast, such students in the business-as-usual schools continued to perform less well than White children in academic achievement and social cognition. The study has important limitations that lead us to view these findings as exploratory, but taken together with other findings, the results suggest that Montessori education may create an environment that is more conducive to racial and ethnic parity than other school environments.

Racial inequality in the United States is a significant concern. One manifestation of the racial and ethnic opportunity gap is inequality in educational outcomes based on race in school (Reardon et al., 2019). Such differences are in place even before first grade, and they remain throughout schooling (Henry et al., 2020; Magnuson & Duncan, 2006; Paschall et al., 2018). Furthermore, it seems that schools exacerbate racial differences because the differences in learning rates between Black and Hispanic versus White students expand during the school years and contract in the summers (Haberman, 2010; Kuhfeld et al., 2021). Although U.S. public schools have, since their founding, been regarded as potentially addressing inequality by providing universal opportunities that eliminate prior differences (Mann, 1848/1961), in some ways schools may be engineered to continue inequality (Hammond, 2020); certainly racial inequity persists today, even after decades of efforts at its elimination via the conventional educational system (Jeynes, 2015). It is possible that a different pedagogical approach may address achievement gaps better than conventional pedagogy. Here we ask whether Montessori preschool may address the inequality in educational outcomes based on race at kindergarten better than other business-as-usual preschool programs.

Montessori Education

The Montessori education system has existed for more than 100 years and is now the most common alternative pedagogy (Debs, 2019; Debs et al., 2022), used in at least 600 public schools and at least 3,000 private U.S. schools and serving children from ages 3 to 18 (National Center for Montessori in the Public Sector, 2023; this census undercounts because not all schools provide data). Maria Montessori was a physician who first worked with atypically developing children in Rome and then with children from families with lower incomes. She eventually performed research on all inhabited continents to create a pedagogy she intended would help all children become flourishing, independent adults (Moretti, 2021). She explicitly addressed social inequity in her founding address to her first school (Montessori, 1967), and social reform aimed at supporting poor and disadvantaged people was a primary mission throughout her life (Trabalzini, 2011).

The Montessori system of education involves specific inputs. As laid out in a recent logic model (Culclasure et al., 2019), these inputs include the classroom features of mixed, 3-year age groupings in large classes with high child-to-teacher ratios; 3-hour uninterrupted work periods during which children may freely choose from a full set of specific, hands-on materials they have been taught to use; and well-trained teachers who carefully prepare and organize the environment for learning, provide small-group or individual instruction, observe all children carefully and assess them formatively, and engage in their own ongoing professional development. Montessori pedagogy emphasizes the classroom environment itself as another teacher; lessons using the Montessori materials in this environment are interconnected and given in a spiraling and successive curriculum (Preschlack, 2023). In addition to learning to carefully and objectively observe so they know how to support children's development, teachers are trained to deeply respect every child, the developmental process, and the interconnectedness of all life (Cossentino, 2009). This deep respect is reflected in a positive emotional climate and frequent and positive peer collaboration in Montessori classrooms (Lillard, 2017; Pottish-Lewis, 2021).

Montessori Pedagogy's Potential Impact on Racial Equity

Some aspects of the Montessori Method of educating children may mitigate racial differences in achievement, whereas other aspects may exacerbate them. One possible mitigator is that Montessori teacher training focuses on each child's individual development and is undergirded by a belief that every child has the potential to flourish in life if properly nurtured. As noted, teachers' attitudes toward all children are meant to undergo a personal transformation during training (Cossentino, 2009; Whitescarver & Cossentino, 2007). Teachers come to believe that all children will develop themselves not because a teacher teaches anything, but because the teacher provides an environment that enables concentration. In Montessori theory, it is children's own concentration—more than the teacher or lessons—that causes development, given a proper learning environment (Montessori, 2012). Once concentration happens, the teacher's job is to stay out of the way and not interfere. Social harmony is claimed to occur naturally within classrooms as the children in the class achieve concentration on their work (Montessori, 2012). In addition, Montessori viewed every child as equal at birth; in an introductory lecture to the last teacher-training course she gave in London in 1947, she said,

> *No matter to what race they belong, to which part of the world they are born, newborns are all alike*

> *. . . . There is another period when we are all alike and this is the period of childhood. All human beings follow the same laws when it comes to development. It is curious, but no matter whether they are Chinese, Indian, African, or European, children all start talking at the same time . . . .* (Montessori, 2012, pp. 4–5)

She lectured to future teachers that every child is a miracle and that teachers must focus on the "greatness of their powers" (Montessori, 2012, p. 6). Teachers are taught to observe (as a scientist observes) every child and to believe that every child is capable of great work (Cossentino, 2006) if teachers create conditions that will further the child's development. In this way, teachers' racial bias may be mitigated in Montessori training; they embrace every child as a miracle, and they focus on creating an environment to allow every child-miracle to unfold.

A second way that Montessori pedagogy may mitigate racial disparities in achievement is through its centralizing of self-determination (Lillard, 2019). Children choose what they do all day long (as long as their choices are constructive). If the "pedagogy of poverty" (Haberman, 2010, p. 81) is reinforced by restricting the access of children of color to challenging material in conventional schools, then giving children full access to materials in Montessori schools may free all children to develop to their fullest potential. As a corollary to the impact of self-determination, a teacher's belief that children may not be capable of doing the work is inert when children choose their own work. By contrast, in conventional schools, teachers' beliefs in students' abilities differ by children's race (Dee & Gershenson, 2017). Furthermore, because the materials are self-correcting, Montessori teachers do not tell a child they are wrong or have not worked carefully enough; children can see such things for themselves. With self-determination at its core, Montessori pedagogy "allows students to flex their cognitive muscles and become independent learners" (Hammond, 2020, p. 152), which is crucial for education equity.

However, there also are two aspects of Montessori education that may work against parity in racial achievement outcomes. One of these is differentiated instruction in the hands of teachers who may remain biased despite their training. Most Montessori teachers are White, whereas most students in public U.S. Montessori schools are children of color (Debs, 2016). If White teachers underestimate the intellectual capabilities of children of color (Dee & Gershenshon, 2017), then they may not give them lessons as readily, thereby impeding some children's progress in the individualized curriculum because children can use only the materials that they have been shown how to use. If children of color are limited by their teachers' biases, then the performance of world-majority children in Montessori classrooms could be worse, on average, than the performance of world-majority children in conventional schools, where children typically get large-group lessons with their classmates (Bassok et al., 2016).

Another aspect of Montessori education that may perpetuate inequality is the fact that it was designed by an Italian woman and her collaborators in the first half of the 20th century; many of its lessons may therefore ensconce a Eurocentric viewpoint that may fail to acknowledge alternative views. Although Montessori and Mario Montessori Sr., her son and collaborator, traveled extensively and spent seven years in India during and after World War II (Montessori, 2020), the potential for cultural hierarchy to pervade the curriculum and materials certainly exists. As Hammond (2020) stated, culturally responsive pedagogy "requires teachers to have the most useful analogies, illustrations, examples, and demonstrations that help make the content comprehensible to the student" (p. 157); the century-old Montessori materials and lessons may not speak to children of color.

Existing Research on Racial Outcomes of Montessori Education

Studies on the outcomes of Montessori education for world-majority children are not entirely consistent, and they have limitations. First, we review studies of elementary school–aged students that have shown that Montessori students had significantly better or similar outcomes than peers in comparison schools. One such study focused only on children in magnet schools, comparing the state test scores of Black or African American children in three urban public Montessori schools in North Carolina with those of students in three other magnet schools (Brown & Lewis, 2017). It found higher reading test performance and equal math test performance for students in Montessori schools. However, this study was small and limited to a few magnet schools. A much larger study of children who attended South Carolina public schools used participant matching for demographics and prior test scores and also controlled these factors (Culclasure et al., 2018); it also found a pattern of greater school-year growth in English

language arts (ELA) and social studies scores for Black children enrolled in the state's 23 public Montessori schools as compared with the children in other public South Carolina schools; however, Hispanic children's growth was not significantly different, nor were math or other scores for Black children. Thus, in this tightly controlled study, there was evidence of Montessori schooling benefiting Black children in elementary school in two subjects, but there was no general pattern of better performance for world-majority children.

Snyder et al. (2022) conducted a nationwide study, examining proficiency levels on third-grade and eighth-grade state tests at Montessori schools (N = 191 schools) in the 10 U.S. states or regions (i.e., Washington, DC, metropolitan area) with the most public Montessori schools, as compared with proficiency levels of their districts (after removing the Montessori schools' scores). They found that public Montessori school students classified as Hispanic and as African American were, as groups, significantly more proficient on state ELA tests than were children attending all other public schools in their districts. On state math tests and compared with their third-grade counterparts in other district schools, African American children performed better, and Hispanic children performed similarly. In this study, even more than in the two just described, better performance may reflect factors outside of schooling itself because the Montessori schools were likely a parent choice (i.e., involving a special application process), and individual child-level data were unavailable. Snyder et al. (2022) attempted to address the issue of extraneous influences by examining differences in proficiency levels in eighth grade while controlling for proficiency levels in third grade. For Black and Hispanic children, the differences in eighth-grade proficiency levels controlling for third-grade proficiency levels were significantly greater for Montessori schools than for those in the rest of their districts' public schools, in both ELA and math. However, students who remained in Montessori schools until eighth grade may have been students who were particularly likely to thrive there.

These three studies suggest that Montessori pedagogy may reduce racial inequality to some degree during the elementary school years, particularly for Black children. Only one study has examined race and ethnicity in preschool. Ansari and Winsler (2014) compared children enrolled in HighScope programs to those in modified Montessori programs in Miami-Dade County, Florida; the Montessori programs were modified in that they had only one age group. Ansari and Winsler found that Hispanic children showed more academic development in Montessori programs than in HighScope programs by the end of kindergarten; these advantages held through third grade (Ansari & Winsler, 2020) but were not observed for Black children in the modified Montessori program at either time point. However, given the racial segregation in Miami-Dade County (Ansari & Winsler, 2014), children of different races were living in different neighborhoods and attending different schools. Because Hispanic children in the study were at different Montessori schools from the Black children in the study, it is possible that the different schools' quality undergirded the different results by race. Another possibility relates to cultural differences in parents' communication style. Black parents tend to use more directive language with children (Miller, 1996; Miller & Hoogstra, 1992). Montessori teachers are trained to use respectful language; in White culture, "respectful" can sometimes be interpreted to mean less direct. Because it differs from many Black children's home language, indirect language may be less effective for Black children. By this reasoning, young Black children in Montessori environments may be less apt to thrive, and the fact that older Black children appear to thrive in Montessori programs may suggest that cultural adaptation occurs on the part of the children or their teachers in public elementary schools.

In sum, some suggestions propose that children of color may thrive in Montessori public schools more than in other public schools, but many of these data are at the elementary level. The sole preschool study suggests that Montessori pedagogy may benefit Hispanic children, but in that study, among other issues, the Montessori program was modified.

In fact, fidelity is at issue in all the studies just reviewed; the fidelity of the Montessori programs was either not well documented or was known to include key modifications. Montessori programs vary widely in fidelity (Daoust, 2004; Daoust & Murray, 2018; Murray & Daoust, 2023), and outcomes can vary accordingly (Lillard, 2012; Lillard & Heise, 2016). In the Miami-Dade County study comparing Montessori programs with HighScope programs, for example, the Montessori program lacked the 3-year age grouping required for high-fidelity Montessori pedagogy (Lillard & McHugh, 2019a); instead, each classroom included only 4 year olds. In the South Carolina study by Culclasure et al. (2018), fidelity in some schools was rated low on a rubric that was designed for the study. A second problem, also noted previously, is that public Montessori schools are typically

choice schools (Culclasure et al., 2018), meaning that parents have chosen Montessori schools among an array of options. Although Brown and Lewis (2017) did compare Montessori schools with other choice (i.e., magnet) schools, we cannot know if characteristics of parents who choose public Montessori schools differ in ways that may directly cause different outcomes. In the South Carolina study (Culclasure et al., 2018), this concern is mitigated but not eliminated by examining year-over-year gains. Thus, the claimed Montessori effect in all of these studies may be an effect of parents who choose Montessori schools, rather than an effect of the pedagogy.

The Current Study

The study described here addresses problems in prior studies with secondary analysis of data from an existing study (Lillard et al., 2017). In this study, the participants were children in high-fidelity Montessori schools who had been admitted by a lottery. The lottery-admission criterion addresses the issue of possible differences in the children being created by differences in parents who choose Montessori schools for their children. This is because the parents of children in the control group (i.e., those who had not been selected in the lottery) had also made the choice for their children to attend the same Montessori schools. In the Lillard et al. (2017) study, children in Montessori schools performed better over time on early academic measures as well as on a test of social cognition, they were more likely to persist in the face of challenge, and they performed somewhat better on tests of executive function at age 4. Lower-income children were particularly affected—positively so—by Montessori education.

Initial results from the prior study did not address race because "the income achievement gap, which is larger than the racial achievement gap, is present by kindergarten, and persists at that high level throughout school" (Lillard et al., 2017, p. 4; Reardon, 2011). This failure to consider race as an independent variable reflected a view that the root of racial disparities in achievement is income disparities that coincide with race (Magnuson & Duncan, 2006).

The present secondary analyses focus on race because race itself is also an important factor in differences in achievement (Burchinal et al., 2002; Reardon, 2016). The most pertinent analyses, given national concern about racial differences in educational outcomes, address whether inequality in educational outcomes based on race exist in Montessori schools to the same degree as in control schools (i.e., the schools children attended when they were not selected by lottery placement in the Montessori schools). In the original study, the participating children were identified by a parent or guardian as African American, Asian, White, Hispanic, multiracial, or other. African American, Hispanic, and multiracial peoples are historically marginalized in the United States, and thus were the groups of most interest in a study addressing inequality in educational outcomes based on race, such as the present study. Although these groups have very different histories in the United States, no single group was sufficiently numerous for reliable analysis as a separate group, so they were combined. Children identified as Asian were not included in the current study because our analyses focused on groups that have historically faced structural inequity and obtained lower performance scores in school (Reardon et al., 2019). In addition, we omitted one child from the study whose parents declined to identify any ethnicity. Because our numbers were still small even when the groups were combined, we consider our analyses to be merely exploratory.

The study focus is academic achievement by race; the current study also examines executive function and theory of mind, which are predictive of academic achievement (Blair & Razza, 2007; Robson et al., 2020). The three outcomes that will be examined are discussed next in the context of existing literature regarding race.

Academic Achievement

As noted, several studies have found inequality in educational outcomes based on race, which is widely considered an opportunity gap (Reardon, 2011, 2016; Reardon et al., 2019). This gap may be caused by schools in which Black, Hispanic, and multiracial children are enrolled offering fewer opportunities (e.g., reading specialists or good library collections) or by fewer opportunities being afforded to Black, Hispanic, and multiracial children than White children within the same schools. At issue is whether the differences in educational outcomes based on race for Black, Hispanic, and multiracial versus White children in Montessori preschools are the same size as the difference seen in children in control preschools.

Executive Function

Executive function refers to the prefrontal processes that allow us to make plans, inhibit one behavior in preference for another, and hold and manipulate information in our minds (Miyake et al., 2000). Several

studies have suggested that executive function in young African American children may sometimes be less developed than in White children (e.g., Blair et al., 2011; Little, 2017); differences in academic achievement may be related to differences in executive function (Nesbitt et al., 2013) because self-regulation predicts academic achievement (Robson et al., 2020). Although reasons for delays in executive function in children of color are unclear, one suggestion is that higher levels of family stress associated with racism interfere with prefrontal development (Hackman & Farah, 2009).

Theory of Mind

Theory of mind refers to a key aspect of social understanding, specifically appreciating that others have mental states that reflect how they construe the world and that drive their behavior. Along with being related to social competence (Wellman, 2011), theory of mind predicts academic achievement (Blair & Razza, 2007; Lecce et al., 2017). Several important developments in theory of mind occur in the preschool years, when children first understand that people may have divergent desires and perceptions and, later, that people can have divergent beliefs. There is a dearth of information about the performance of different racial and ethnic groups in the United States on theory of mind tests; most studies have used majority-White samples and had insufficient subgroup numbers to examine outcomes by race or ethnicity (e.g., Weimer & Guajardo, 2013). However, three studies did provide data on the performance of different racial and ethnic groups in the United States on theory of mind tests. Curenton (2003) tested a sample of African American and European American children enrolled in Head Start programs. Controlling for language proficiency, Curenton found lower performance on the contents version of the false belief test among African American children than White children. In a contents false belief test, crayons are placed in a Band-Aid box and children are asked what a naive person (i.e., someone who had never seen inside the box) would think was in the box—in other words its contents. Curenton found no racial differences in performance on two other standard theory of mind tests. The contents false belief finding replicated a previous study in which a mainly African American sample performed less well on the contents false belief test than is typical for predominantly White samples (Holmes et al., 1996). A more recent study using a full five-part Theory of Mind scale (Wellman & Liu, 2004), with a sample described as predominantly children of color, found they passed all tasks on the scale at an older age on average relative to other studies with predominantly White samples (Baker et al., 2021). In sum, although few theory of mind studies have addressed race in a U.S. context, those that have suggest that the development of theory of mind in children of color may occur somewhat later, at least on specific tests, than in White children; here, we ask whether there is parity in this development for children of different races who attend Montessori schools.

In sum, the goal of the present study was to analyze an existing dataset to determine whether high-fidelity Montessori preschool environments are places of greater racial parity than business-as-usual preschools for academic achievement, executive function, and theory of mind development.

Method

Participants

Participants were 134 children with an average age of 41.16 months; *SD* = 3.30, range = 33.8–48.7 at their first testing point in the fall of their first year of prekindergarten (PK3, or prekindergarten at age 3 years) (See Table 1). Seventy-two children were male and 62 were female; 53 children were identified by their parents or guardians as White and 81 as either African American (n = 23), Hispanic (n = 27), or multiracial (n = 31). Of the nine multiracial participants whose parents specified what "multiracial" meant, six children were Hispanic/Latino and White, two were African American and White, and one was African American and Hispanic. The average household income in the full sample was $70,022 (*SD* = $45,550; range = $0–$200,000). Average maternal education included some college (6.67, *SD* = 1.2, range = 2–8: where 2 = ninth grade, 5 = high school diploma, 8 = graduate school; see Appendix).

Lottery and Control-Group Schools

The children's parents or guardians had entered them in a lottery to enter the PK3 program at one of two high-fidelity urban public Montessori schools in the northeastern United States in one of the 4 years spanning 2010–2013. The fidelity of the schools was indicated by their being recognized by Association Montessori Internationale of the United States (i.e., AMI/USA), the American branch of the association Maria Montessori founded in 1929 with the aim of maintaining and developing her pedagogy. AMI/USA has a formal recognition program for schools that have AMI-trained teachers and that apply the pedagogy according to specific

Table 1

Average and Standard Deviation of Age, Household Income, and Maternal Education and Numbers of Each Race by School Type

Variable	Montessori group (*SD*)	Control group (*SD*)
Age at fall test in months	41.45 (3.21)	40.87 (3.38)
Household income	$72,795 (41,553)	$67,165 (49,490)
Maternal education	6.72 (1.31)	6.62 (1.11)
Race (*n*):		
White	33	20
Hispanic	11	16
Black	12	11
Multiracial	12	19

Note. For maternal education, 2 = ninth grade, 5 = high school diploma, and 8 = graduate school.

standards. The lottery was random except for sibling and staff preferences and preferences for children who live in the neighborhood; no staff children were included in the study, and only two siblings were. Omitting the siblings (i.e., students whose families had been enrolled through previous years' lotteries) did not affect results. There was also one crossover (i.e., noncomplier) child in the control group who had been admitted to one of the two Montessori schools but did not attend. Excluding this child also did not change results. The fact that both of the schools were magnet schools and thus were in low-income neighborhoods but admitted a fixed percentage of children from outside of the neighborhood means that, ideally, our study enrollment could have incorporated the information about what lottery categories (or blocks) the children were in. Unfortunately, when the study was conducted, no information was available regarding neighborhood-preference lottery blocks; this threat to validity is discussed further in the Limitations section.

All children's parents had specified one of two Montessori schools as their first choice. Among the lottery-waitlisted children, only those who went to another type of school (i.e., not another Montessori school) were included in the study; thus, the study used a treatment-on-the-treated design.

Control Schools

The control participants were in 51 different schools when they were 3 years old, including other magnet schools (e.g., a Reggio magnet school, a science specialty school), childcare centers such as Bright Beginnings, and cooperative schools. Thirty-one control children were in urban schools, and 35 were in suburban schools. Twenty-two control children were in public schools, and 14 of these were in a public magnet school. Thirty-seven children were in private schools or day-care centers (roughly half urban, half suburban), and seven were in urban Head Start programs. At the time of the study, all public early childhood programs in the state in which the study took place were required to satisfy National Association for the Education of Young Children (NAEYC) accreditation standards and be a member of the state's early childhood professional registry. This state also required an early childhood teaching credential that entailed either (a) being a graduate of an approved (public state) higher education program or teaching experience or (b) a degree from an unapproved institution and 12 credits in early childhood education. No further information on the control children's schools is available.

Measures

Measures used in the study addressed children's early academic achievement, executive function, and theory of mind.

Academic Achievement

Academic achievement was measured with four Woodcock-Johnson IIIR subtests (McGrew & Woodcock, 2001): Picture Vocabulary, Letter Word, Applied Problems, and Calculation. These tests are widely used in the field and have been normed on nationally representative samples of children ages 4 and older. Some Letter Word test stimuli were modified to reflect that Montessori classrooms teach cursive letters: The early items in which children identify letters were overlaid with cursive letters for the Montessori participants. The Calculation subtest was administered only to children who reached item 19 on the Applied Problems test. The Applied Problems and Calculation raw scores were

summed to create a math score. In the original study, the Math, Letter Word, and Picture Vocabulary scores loaded on a common factor and were highly correlated ($r > .80$); to reduce the number of comparisons, these scores were combined (by adding z scores) for an overall academic achievement score for each child (for another prominent study using such a strategy, see Lipsey et al., 2017).

Executive Function

Two tests measured executive function: Head-Toes-Knees-Shoulders, or HTKS (Ponitz et al., 2009), and Design Copy (Korkman et al., 2007). HTKS is an opposites game in which children have to touch the opposite of a specified location; the experimenter explains the test ("When I say touch your toes, I want you to touch your head") and then gives a series of commands. Children are given 2 points for immediately touching the opposite location, 1 point for starting to touch the wrong location and then switching to the right location, and 0 points if they touch the designated location (e.g., touch their head when told to touch their head). Children who do well on the Head-Toes portion have Knees and Shoulders added to the command set. There are 10 commands in each section, so the possible scores range from 0 to 40.

Design Copy is a subtest from the Visuospatial Processing section of the neuropsychological assessment NEPSY-II; it was administered and scored in the standard manner (Korkman et al., 2007). Children were shown a 4 x 4 grid with geometric or other shapes in each box of the top row and the third row. The first box had a vertical line; the experimenter showed children how to copy the line in the box below it, saying, for 3- and 4-year-olds, "See this line? I will draw one here." The experimenter then pointed to the second figure and the second box in the second row and said, "Now you draw one here," pointing to the second figure (i.e., a horizontal line) and the box below it. When children were in kindergarten, and for the remaining items, the experimenter simply pointed to the top figure and then the box below, saying, "Copy this one here."

This sequence continued until a child failed to copy three consecutive figures, or for 16 items. Raw scores ranged from 0 to 16. An independent coder coded a randomly selected subset of children at each test period, and interrater reliabilities across the two coders were excellent: $r = .98$ (28 children at Time 1), $r = .97$ (23 children at Time 2), $r = .95$ (15 children at Time 3), and $r = .91$ (21 children at Time 4). To reduce the number of comparisons, the scores on HTKS and Design Copy were converted to z scores and summed for an executive function score. A second rationale for combining the two scores is that single measures of executive function are less reliable than composite measures created from more than one test (Willoughby et al., 2011).

Theory of Mind

Theory of mind was measured using the Theory of Mind scale (Wellman & Liu, 2004). The scale has good psychometric properties (Beaudoin et al., 2020; see their Supplementary Table 2). Four consecutive tests from the scale were used; children's scores on each of the four theory of mind tests were summed for the scale score and also examined separately.

Each short vignette in this scale measures an aspect of understanding others' minds and is presented either with small dolls and other objects or with pictures. For the test entitled Diverse Beliefs, children were shown a doll and pictures of different locations and then asked where they thought an object was (e.g., the doll's cat)—in the bushes or in the garage. After children responded, they were told the doll thought her cat was in the other location; children were then asked where the doll would look for her cat. The correct answer was where the doll (not the child) thought it was.

For the test assessing children's understanding of knowledge access, children were shown a doll and a doll-sized cupboard and then were asked what was inside the cupboard. The children were then shown the contents of the cupboard (e.g., a ladybug) and were asked what the doll, who had never seen inside the cupboard, would think was inside.

For the contents false belief test (described earlier), children were shown a standard box (e.g., a Band-Aid box) and, after the children agreed that they thought the box would contain Band-Aids, they were shown that it actually contained crayons. The children were then asked what a doll who had never seen inside the box would think was in it.

Only children who passed the contents false belief test by saying that a person would think the Band-Aid box contained Band-Aids were given the final theory of mind test, the appearance reality emotion test. For this test, participating children were given a scenario in which a child received a disappointing gift. To pass the test, participating children had to report that the child who received a disappointing gift would pretend to be happy in front of the giver while feeling sad inside. This test is given

only to children who successfully complete the false belief test because it is highly unusual for a child who has been unsuccessful on the false belief test to pass the appearance reality emotion test (Wellman, 2014). Because there is a maximum of four tests, each of which is either passed (for a score of 1) or failed (for a score of 0), Theory of Mind scale scores in this study range from 0 to 4.

Procedure

Children were tested individually by trained experimenters on the battery of measures on each of four occasions: in the fall soon after they matriculated (September–December), approximately six months later, and then approximately 12 and 24 months after that. Most children were tested in their school or day care; some were tested in a local library. All children were tested in English. The study methods are described in more detail in Lillard et al. (2017).

Power

Given the sample sizes here, using Cohen's *d*, a power of .8, and the standard alpha of .05, the minimum Cohen's *d* is 0.69 for the Montessori group and 0.76 for the control group. These effects are quite large for field research in schools (Kraft, 2020), so our study is underpowered; this is a second reason why we consider the study to be only exploratory.

Analytic Approach

The research question addressed in this analysis was whether racial disparities that exist in business-as-usual preschools also exist in Montessori preschools. We first examined whether socioeconomic status, the education of the mother and father, and racial and gender balances differed across Montessori and control groups. Next, to address the primary research questions regarding whether racial disparity is less apparent in Montessori programs, the data file was split between Montessori and control groups. We conducted two longitudinal latent growth curve analyses on each variable, the first to determine whether the slope of change across the preschool years differed for White versus Black, Hispanic, and multiracial children in Montessori schools and the second to determine whether the slope of change differed for children in these groups in the control schools. These analyses were followed by simple *t* tests examining whether there were racial group differences at any time point for the focal variables within each school group. Differences at single time points were deemed less interesting than patterns of difference; hence, we report results reflecting clear directional patterns. Analyses were performed via Mplus (Version 8.4) and R software (Version 4.2).

Results

The Montessori and control groups were not significantly different in terms of racial or gender category (as determined by chi-square tests), nor did they differ in age, household income, or mother's highest level of education at baseline (using *t* tests). Although not significantly different, the racial composition was not identical (possibly suggesting some compromise in the random assignment, due either to not taking neighborhood preferences into account or to differentials in the choice to participate in the study by condition). For this reason, race was accounted for in the analyses. Because our samples were small and therefore more prone to spurious effects, we also controlled for gender and maternal education (which is highly related to income) in analyses where possible.

Children were not clustered in classrooms (as they would be had we used hierarchical modeling) because (a) for the control children, typically only one child was in a classroom (indeed, only one child was typically in each control *school*) and therefore there were no clusters; and (b) for the treatment children, the classroom composition changed markedly each year as 33% of the children were replaced by a new set of children. There also was teacher and assistant turnover in the 11 classrooms involved in the study. Because of this instability, it did not make sense to us to cluster sets of children within Montessori classrooms.

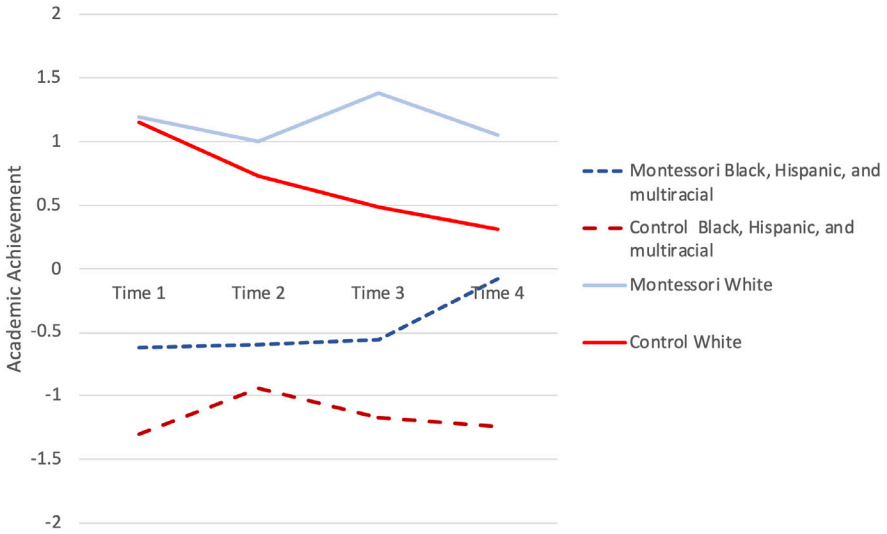
There was sample attrition during the study: From the first test point to the fourth test point, the Montessori group decreased from 68 children to 57, and the control group decreased from 66 children to 61. The primary cause of attrition was parents moving out of the area; because moving out of the area is (in study terms) a random event (rather than caused by a systematic variable related to Montessori education), the missing data were viewed as missing at random. Missing data were managed using full information maximum likelihood estimation.

Academic Achievement

Latent growth curve analyses were performed on data from each school group, controlling for baseline score (Time 1) at the intercept and for baseline score,

Figure 1

Academic Achievement z Scores Across Time by School Type and Racial Group



gender, and maternal education when examining the slope. Details are provided in the Appendix.[1] For both groups, as expected, test point affected the intercept in that Black, Hispanic, and multiracial children's academic achievement was lower when they first began school. Thereafter, for children in Montessori schools (i.e., the treatment group), test point was not significantly related to the slope of academic achievement. However, it was in the control group, with a beta of -0.243 (p = .026).

This pattern in academic achievement was reiterated using t tests. Significant differences in White versus Black, Hispanic, and multiracial children in the control group were seen at all four test points. In the treatment (i.e., Montessori school) group, significant differences between White versus Black, Hispanic, and multiracial children were present at the first three test points, but the difference was not significant by the end of kindergarten. This pattern is shown in Figure 1, in which the lines of the Montessori group begin to close from the 4-year-old prekindergarten (PK4) year to the kindergarten year (i.e., the third to the fourth test point), with the Black, Hispanic, and multiracial children's z scores improving for treatment children, whereas the control children's lines remained separate and did not improve relative to the sample. In fact, the achievement z scores of the Montessori Black, Hispanic, and multiracial group approached those of the control group's White children by the spring of the kindergarten year.

Theory of Mind

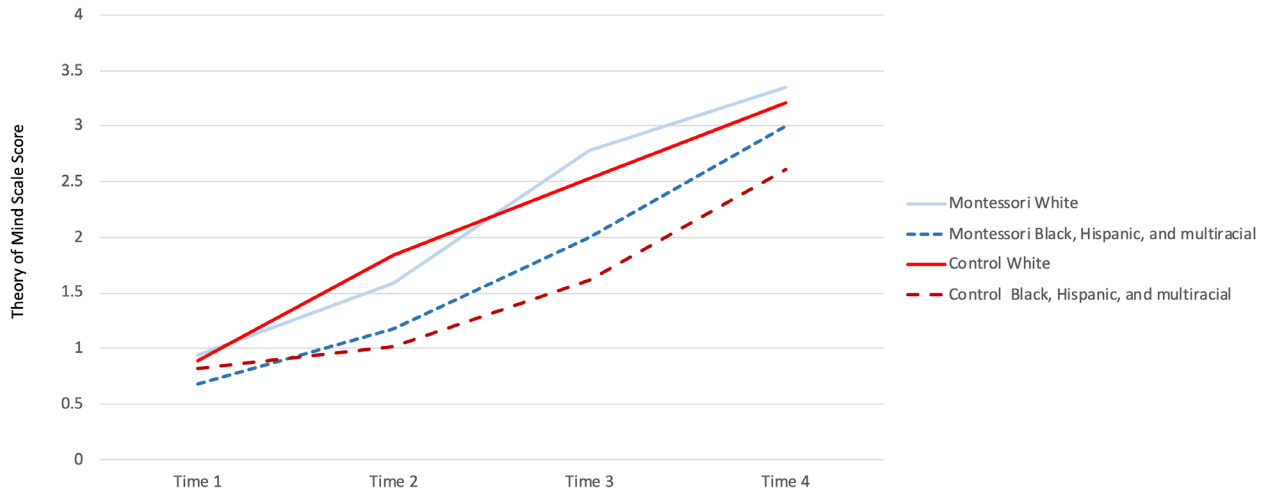
The same latent growth curve analysis was performed on the total Theory of Mind scale score and revealed no racial differences in the slope of theory of mind development in either the Montessori group or the control group. Details are provided in the Appendix. Although the latent growth curves were not significant, model fit was not ideal. Using an alternative analytic method, t tests showed significant racial group differences at all spring test points in the control group: White children in the control group scored higher than Black, Hispanic, and multiracial children at each spring test point. No pattern of racial difference was observed in the Montessori group, as Figure 2 shows.

Because prior research had shown racial differences particularly on one test (i.e., contents false belief), we ran Mann-Whitney U tests (appropriate for 0–1 data) to examine possible racial differences at each test point

[1] Although the sample size is relatively small for growth curve analyses, children were randomly assigned to the Montessori group or the control group. Remedies (e.g., controlling for covariates) were also undertaken to strengthen the statistical conclusion validity. Although Bayesian methods in conjunction with informative priors perform better with small sample sizes, they may produce incorrect conclusions when the prior information is incorrect (Shi & Tong, 2017). For our analysis, we tried Bayesian methods with noninformative priors; the results were the same as our current results. It is difficult to find informative priors and check whether they are accurate. Because Bayesian methods are less familiar to most researchers, we did not report the results from the Bayesian approach.

Figure 2

Theory of Mind Scale Scores Across Time by School Type and Racial Group



for each test on the scale. In the control group, White children were more apt than Black, Hispanic, and multiracial children to answer correctly on the false belief test in the spring of both their PK3 and their PK4 years, or Times 2 and 3 ($U = 290.0, 260.0$; $p < .001$, $p = .001$, respectively), as well as the knowledge access test at those same time points ($U = 267.5, 312.0$; $p = .002, .024$, respectively). The hardest test on the scale, the appearance reality emotion test, also trended to difference at the end of the kindergarten for children in the control group ($U = 109.5$, $p = .076$). By contrast, for children in the Montessori group, the knowledge access test showed a racial group difference when they first started school (Time 1): $U = 404.0$, $p = .016$. The significance of that difference was reduced by the end of PK3 (Time 2) in the Montessori group ($U = 412.0$, $p = .051$); thereafter (Times 3 and 4), the difference in the knowledge access test scores of White children versus Black, Hispanic, and multiracial children was not significant in the Montessori group.

Executive Function

The same latent growth curve analysis was performed on the executive function composite and revealed no differences in racial group performance in either type of school. Again, details are provided in the Appendix. For executive function, *t* tests at each time point also showed no patterns of differences.

Discussion

Education in the United States has long been viewed as a mechanism that may level economic outcomes by providing opportunities to all children. Current assertions and developing mainstream understandings of how implicit bias can affect opportunities in schools run counter to this long-held view (Hammond, 2020). The present exploratory secondary data analyses add to a body of existing research that suggests Montessori education may be a mechanism for creating more equal outcomes for Black, Hispanic, and multiracial children.

The first finding is related to racial differences in academic achievement. In both samples, when children began school at age 3, there were differences by race, with Black, Hispanic, and multiracial children scoring lower than White children. These differences remained throughout preschool for Black, Hispanic, and multiracial children in the control group; for Black, Hispanic, and multiracial children in Montessori classrooms, scores were similar by the end of preschool, and the racial difference in academic achievement was no longer significant. This finding is consistent with existing literature that showed smaller racial test score gaps for children in Montessori programs compared with other school programs (Brown & Lewis, 2017; Culclasure et al., 2018; Snyder et al., 2022), as well as better performance among Hispanic children in modified Montessori programs versus HighScope programs (Ansari & Winsler,

2020). This finding is also consistent with qualitative research reporting on interviews with adults ages 25 to 40 who attended a predominantly Black Montessori preschool as young children. Although there was no control group in this mixed-methods study, these adults were highly successful: 92% had an undergraduate degree (compared with less than 40% of Americans in general), and 25% also had postgraduate degrees (Lillard et al., in press).

There are limitations in all of these studies, but if the results are valid and reliable, what might be responsible for the finding that gaps in performance of different racial groups remained steady across preschool for the control group but lessened over time for children attending public Montessori schools? Because all parents of the children in the present study had selected a Montessori school for their child, it seems unlikely that the findings in this study can be attributed to preexisting differences in Montessori parents versus control parents (cf. Todd & Wolpin, 2007). Another possibility is that different schools have different resources. Children in the Montessori group were at the same two public schools, distributed across 11 Montessori preschool classrooms. By contrast, children in the control group were at 51 different schools at the start of the study. It is possible that Black, Hispanic, and multiracial children who were not admitted to Montessori schools attended lower-quality schools than did White children who were not admitted to Montessori schools and that those lower-quality schools then exacerbated differences over time.

Unfortunately, little information about the schools attended by control children was collected, but it is possible that different schools contributed to the different levels of performance seen in the present study. Although some research has found that school inputs have little effect after family inputs are accounted for (e.g., Todd & Wolpin, 2007), certainly preschool quality is known to have effects (Yoshikawa et al., 2013). However, studies that use hierarchical linear modeling to control features at the classroom level (where resources or classroom quality are the same) still find inequality in educational outcomes based on race (Quinn & Cooc, 2015). Finally, even within the same conventional schools, although differences in levels of performance by race decrease somewhat, there are still differences (Singham, 2003). High-quality preschool does reduce inequality in educational outcomes based on race (Friedman-Krauss et al., 2016). The quality of the public control schools in the present study was likely similar to that of the public Montessori schools in several respects, in that public early childhood programs in the study state were required to satisfy NAEYC accreditation standards and to be members of the state's professional registry; teachers also were required to have specific credentials. The private control schools may explain the difference, in that perhaps White children in the control group were more likely to attend high-quality private schools than were Black, Hispanic, and multiracial children in the control group; on average, however, private and public school attendance does not render different achievement outcomes (Pianta & Ansari, 2018). In sum, it is possible that lower quality in schools attended by Black, Hispanic, and multiracial children in the control group explains our findings, but there are reasons to think this is not the full explanation. A meta-analysis of the inequality in educational outcomes based on race showed that curriculum can reduce the gap (Jeynes, 2015), and it is possible that Montessori pedagogy is one such curriculum.

Thus, we next consider the possibility that the difference in educational outcomes based on race across Montessori schools and control schools stems from features of Montessori pedagogy not present in most control preschool programs. Most preschool programs are teacher driven, not learner centered (Bassok et al., 2016). We know from many years of research that teacher expectations can be a significant predictor of student learning in conventional school environments (Good et al., 2018). In contrast to conventional teachers (Dee & Gershenson, 2017), Montessori teachers may be less likely to hold lower expectations for global majority children, although we know of no research that supports that conjecture. However, it is possible that, even if the expectations of Montessori teachers and non-Montessori teachers were equally biased, those biases may have less influence on student outcomes in the Montessori system, for reasons discussed in the Introduction. For example, this failure to negatively influence children could be caused by the different ways teachers interact with children and give feedback in each system. Montessori pedagogy offers a prepared environment that supports agency or learner autonomy (Montessori, 2012). In Montessori programs, children are given initial lessons with materials, but thereafter they learn from using the materials. The teacher's role is to make that initial connection, but children then seek to master the materials on their own. Corroborating the possibility that teacher bias has less impact in Montessori classrooms because teachers interact with children differently there, in the

Lillard (in press) interview study referred to previously, one alumnus said,

> *The Montessori environment let me know that I could identify what it is that I'm interested in, capitalize on those things, learn those materials, perfect those materials at my own pace, and then move forward on to the next project because that's where Montessori [school] always was.* (p. 16)

Another former student said,

> *Not only did [Montessori schooling] give me autonomy over what I was learning about, and the pace at which I learned, but it also in turn allowed me to feel mastery of it . . . . We were self-led. We had to figure it out for ourselves. I mean, we were given a lesson, but then we were sent off to get to work and I think that is just . . . that is so important. . . . [The teacher] was always available for help and we were encouraged to ask questions and get help, but at the end of the day it was on us, we were the ones who were taking charge of our own learning and we had to engage with whatever it was in the classroom that was at our level at that time.* (Lillard et al., in press, pp. 16–17)

In sum, perhaps Black, Hispanic, and multiracial children close outcome gaps over time in Montessori environments because teachers do not inadvertently provide feedback in ways that reinforce those gaps. One reason they may not provide such feedback is because Montessori pedagogy entails self-directed learning with a set of hands-on materials designed to teach, rather than learning that is achieved through teacher interaction with students. In this way, Montessori pedagogy shrinks achievement gaps because it frees children to capitalize on their own capabilities.

There is a third possibility that Montessori education closes achievement gaps more than business-as-usual schools do: Teacher-child relationships, in theory, may be stronger in Montessori settings than in non-Montessori settings, in part because of the one-on-one instruction that attends to a child's specific learning needs. Other researchers have shown that stronger teacher-child relationships predict, in particular, reading achievement for African American preschoolers (Burchinal et al., 2002). Although we know of no studies examining the strength of teacher-child relationships in Montessori education, it is the case that Montessori children are typically with the same teacher for 3 years (rather than the typical 1 year in most schools), providing an opportunity for stronger bonds. In addition, Montessori teachers are counseled to behave toward children in ways that may foster strong relationships (Lillard, 2017); for example, misbehaving children are not punished with a time-out but are instead asked to stay very close to the teacher until they learn to control themselves. Montessori teachers are also counseled to be warm and sensitively responsive (Lillard & McHugh, 2019b); such interactions are associated with stronger school-readiness skills (Pianta et al., 2020).

Differences in theory of mind for different racial and ethnic groups were not seen in the latent growth curve analyses, which admittedly were underpowered, but differences were seen both overall and on two of the subtests that comprise the overall Theory of Mind scale score (i.e., the knowledge acquisition and false belief tests). What may account for these differences? One possibility is that the 3-year age groupings in Montessori classrooms, which provide opportunities for learning about others' minds, are not achieved as often in the programs in control schools because many of those classrooms were likely single age or had at most 2-year groupings (e.g., Foster et al., 2020). Supporting this possibility, a Chinese study (Wang & Su, 2009) found that only children (i.e., children with no siblings) had more advanced understanding of false belief when they were in preschool classrooms with 2-year age spans than when they were in classrooms with children who were all the same age. Considering family contexts, children who have one or more siblings who are close to the child's own age, and with whom they can interact, have a more advanced theory of mind than children whose siblings are much older or younger or than children who lack siblings (McAlister & Peterson, 2013). The ability to interact with other children who are not of the same age (but are not too much older or younger) may proffer opportunities to develop social understanding among all children (Lillard & Eisen, 2017). According to one hypothesis, then, the racial differences in theory of mind disappear in Montessori schools because all children have social experiences in the classroom that spur theory of mind development.

Another possibility for why racial differences are mitigated in Montessori schools is related to the didactic apparatus itself and to the specific understandings tested in the Theory of Mind scale. In Montessori classrooms,

there are many *Sensorial activities*, which include activities meant to educate the senses of touch, smell, and hearing. When one engages in these activities, one sometimes wears a blindfold to accentuate the sense. Another standard exercise uses the Mystery Bag (or stereognostic bag), which is a bag full of little objects into which children insert their hand to feel for the correct object. For the Theory of Mind scale's knowledge access test, children know what is inside a cupboard (or drawer, etc.) but have to acknowledge that a doll who had not seen inside the cupboard would not know its contents. At school entry, there was a racial performance difference on this test in the Montessori sample, but that difference disappeared by the PK4 year. By contrast, there was no initial difference in successfully completing this test among the control sample, but there was at the later test points. It is conceivable that Montessori children's experience with Sensorial exercises, blindfolds, and the Mystery Bag helped their understanding of knowledge access. Theories concerning how a theory of mind develops in children maintain that the component understanding (e.g., knowledge access, contents false belief) are hierarchical, such that each understanding builds on the previous ones. Thus, children who understand perceptual access early also develop false belief understanding early, and then appearance reality emotion understanding early as well (Wellman, 2014). Thus, two possibilities for the different performance patterns on the theory of mind tests are that the differences stem from children interacting with peers of slightly different ages in Montessori classrooms or from specific Montessori didactic materials that help them learn about minds, or both.

Limitations

Although our findings are consistent with some other literature (e.g., Brown & Lewis, 2017; Culclasure et al., 2018), we view them as preliminary for several reasons. First, we did not have access to lottery information that enabled us to determine whether a child was admitted because they had preference due to residing in the neighborhood. We understood that both lotteries (in neighborhood and out) were competitive. Using lottery-waitlisted children as one's control group equalizes the treatment and control groups in one important way: All children in the study have a parent who entered them in a lottery to attend a Montessori school and thus are equal on any characteristics that go along with that. Nevertheless, we ideally could have also had information about who was admitted because of neighborhood preference and considered those children as a separate lottery pool as further basis for equalizing the treatment and control groups. Although racial representation was not significantly different within our small sample, across our groups it was not even: White children were overrepresented in the Montessori sample. Although we controlled for this difference in analyses, it is a reason for caution regarding the results.

Another limitation is that we know little about the alternative programs in which the control children were enrolled. Ideally, we would have had more information about the control children's experiences. It is possible that, in the control sample, the Black, Hispanic, and multiracial children attended lower-quality preschools than did White children; if so, that may explain the different patterns of performance observed in the present study. Further research should examine features of the control schools. However, we do know something about those features because all public prekindergarten programs in the test state must comply with NAEYC standards, as well as specific training standards, and there are reasons to think the Montessori curriculum itself may be responsible for the different patterns of performance seen in the present study.

Another limitation is that all children in this study participated in a lottery to enter a high-quality preschool program. It is unclear whether the results found in the present study would apply to children whose parents or guardians did not enter them in such a lottery.

Another limitation is that children of different ethnic backgrounds were grouped together to create sufficient sample sizes. The life experiences of African American children and Hispanic children and their families are different, and, although this grouping was necessary for analysis, further work using larger samples should examine separate outcomes for different racial groups. Another limitation of the small sample, besides not having sufficient representation to examine each race separately, is that, particularly for the control group, the model fit for theory of mind and executive function was less than ideal, making the results less reliable. However, differences in theory of mind were also revealed by *t* tests.

Finally, both of the Montessori schools in this study were recognized by AMI/USA for their high level of fidelity at the time of the study. Therefore, they adhered to strict implementation criteria, which included that all teachers were trained by AMI and that all teachers had Montessori materials. It does not necessarily mean that

every teacher implemented the Montessori program with fidelity, but it is a fairly good indicator of fidelity. Many schools call themselves Montessori schools but do not adhere to Montessori's pedagogy at a high level of fidelity. We do not know whether the results found here generalize to other Montessori schools or even to these study schools at another time.

Conclusion

The study found that, while children in the control group showed gaps in academic outcomes and theory of mind by race, consistent with the existing literature, children who had won the lottery to enter high-quality Montessori preschools did not show such gaps by the end of preschool (although they did show gaps initially). Although it is possible that these results stem from children in the control group attending different schools, the results may also be caused by features of the Montessori system, including self-directed learning, mixed-age groups, and specific didactic exercises. Limitations in the design of this study—including not having complete information about lotteries, a small sample, and uneven racial representation—temper the strength of our conclusions, and we hope the findings will spur further research into the possibility that Montessori education may help close racial opportunity gaps.

Funding Acknowledgment

During preparation of this manuscript Tong was supported in part by National Science Foundation grant no. SES-1951038, and Lillard was supported by grants from the Wend Collective and the Institute of Education Sciences, U.S. Department of Education; the latter was through Grant R305A180181 to the American Institutes for Research and Angeline Lillard. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Author Information

† Corresponding Author

Angeline S. Lillard† is professor of psychology at University of Virginia. She can be reached at asl2h@virginia.edu. https://orcid.org/0000-0001-9697-6611

Xin Tong is Associate Professor of Psychology at University of Virginia.
https://orcid.org/0000-0003-3050-1554

Paige M. Bray is associate professor and director of early childhood education and Montessori studies at University of Hartford.

References

Ansari, A., & Winsler, A. (2014). Montessori public school pre-K programs and the school readiness of low-income Black and Latino children. *Journal of Educational Psychology*, *106*(4), 1066–1079. https://doi.org/10.1037/a0036799

Ansari, A., & Winsler, A. (2020). The long-term benefits of Montessori pre-K for Latinx children from low-income families. *Applied Developmental Science*, 252-266. https://doi.org/10.1080/10888691.2020.1781632

Baker, E. R., Huang, R., Battista, C., & Liu, Q. (2021). Theory of mind development in impoverished US children and six cross-cultural comparisons. *Journal of Applied Developmental Psychology*, 76, 101314. https://doi.org/10.1016/j.appdev.2021.101314

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, I(4), 1–31. https://doi.org/10.1177/2332858415616358

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, 10, Article 2905. https://doi.org/10.3389/fpsyg.2019.02905

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., Kivlighan, K. T., Fortunato, C. K., & Investigators, F. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development*, 82(6), 1970–1984. https://doi.org/10.1111/j.1467-8624.2011.01643.x

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(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Brown, K. E., & Lewis, C. W. (2017). A comparison of reading and math achievement for African American third grade students in Montessori and other magnet schools. *The Journal of Negro Education*, *86*(4), 439–448. https://doi.org/10.7709/jnegroeducation.86.4.0439

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, *40*(5), 415–436. https://doi.org/10.1016/S0022-4405(02)00107-3

Cossentino, J. (2009). Culture, craft, & coherence: The unexpected vitality of Montessori teacher training. *Journal of Teacher Education*, *60*(5), 520–527. https://doi.org/10.1177/0022487109344593

Cossentino, J. M. (2006). Big work: Goodness, vocation, and engagement in the Montessori method. *Curriculum Inquiry*, *36*(1), 63–92. https://doi.org/10.1111/j.1467-873X.2006.00346.x

Culclasure, B., Fleming, D. J., & Riga, G. (2018). *An evaluation of Montessori education in South Carolina's public schools*. The Riley Institute at Furman University. https://www.furman.edu/wp-content/uploads/sites/195/rileypdfFiles/MontessoriOverallResultsFINAL.pdf

Culclasure, B. T., Daoust, C. J., Cote, S. M., & Zoll, S. (2019). Designing a logic model to inform Montessori research. *Journal of Montessori Research*, *5*(1), 35–74. https://doi.org/10.17161/jomr.v5i1.9788

Curenton, S. M. (2003). Low-income preschoolers' false-belief performance. *The Journal of Genetic Psychology*, *164*(4), 411–424. https://doi.org/10.1080/00221320309597887

Daoust, C. J. (2004). An examination of implementation practices in Montessori early childhood education (Publication No. 3146824) [Doctoral dissertation, University of California–Berkeley]. ProQuest Dissertations and Theses Global.

Daoust, C. J., & Murray, A. (2018, March 21–25). *A comparison of elementary classroom practices based on Montessori certification*. Paper presented at the annual conference of the American Montessori Society, Denver, CO, United States.

Debs, M. C. (2016). Racial and economic diversity in U.S. public Montessori schools. *Journal of Montessori Research*, *2*(2), 15–34. https://doi.org/10.17161/jomr.v2i2.5848

Debs, M. C. (2019). *Diverse parents, desirable schools: Public Montessori in an era of school choice*. Harvard Education.

Debs, M., de Brouwer, J., Murray, A. K., Lawrence, L., Tyne, M., & von der Wehl, C. (2022). Global diffusion of Montessori schools: A report from the 2022 global Montessori census. *Journal of Montessori Research*, *8*(2), 1–15. https://doi.org/10.17161/jomr.v8i2.18675

Dee, T., & Gershenson, S. (2017). *Unconscious bias in the classroom: Evidence and opportunities*. Google's Computer Science Education Research. https://goo.gl/O6Btqi

Foster, T. J., Burchinal, M., & Yazejian, N. (2020). The relation between classroom age composition and children's language and behavioral outcomes: Examining peer effects. *Child Development*, *91*(6), 2103–2122. https://doi.org/10.1111/cdev.13410

Friedman-Krauss, A., Barnett, W. S., & Nores, M. (2016). *How much can high-quality universal pre-K reduce achievement gaps?* Center for American Progress. https://www.americanprogress.org/wp-content/uploads/sites/2/2016/04/NIEER-AchievementGaps-report.pdf

Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3-5), 99–123. https://doi.org/10.1080/13803611.2018.1548817

Haberman, M. (2010). The pedagogy of poverty versus good teaching. *Phi Delta Kappan*, *92*(2), 81–87. https://doi.org/10.1177/003172171009200223

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003

Hammond, Z. (2020). Looking at SoLD through an equity lens: Will the science of learning and development be used to advance critical pedagogy or will it be used to maintain inequity by design? *Applied Developmental Science*, *24*(2), 151–158. https://doi.org/10.1080/10888691.2019.1609733

Henry, D. A., Betancur Cortés, L., & Votruba-Drzal, E. (2020). Black–White achievement gaps differ by family socioeconomic status from early childhood through early adolescence. *Journal of Educational Psychology*, *112*(8), 1471–1489. https://doi.org/10.1037/edu0000439

Holmes, H., Black, C., & Miller, S. (1996). A cross-test comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology*, 63, 263–285. https://doi.org/10.1006/jecp.1996.0050

Jeynes, W. H. (2015). A meta-analysis on the factors that best reduce the achievement gap. *Education and Urban Society*, *47*(5), 523–554. https://doi.org/10.1177/0013124514529155

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Clinical and interpretive manual*. The Psychological Corporation.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kuhfeld, M., Condron, D. J., & Downey, D. B. (2021). When does inequality grow? A seasonal analysis of racial/ethnic disparities in learning from kindergarten through eighth grade. *Educational Researcher*, *50*(4), 225–238. https://doi.org/10.3102/0013189X20977854

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, *51*(2-3), 220–239, https://doi.org/10.1080/00273171.2015.1134306

Lecce, S., Caputi, M., Pagnin, A., & Banerjee, R. (2017). Theory of mind and school achievement: The mediating role of social competence. *Cognitive Development*, *44*, 85–97. https://doi.org/10.1016/j.cogdev.2017.08.010

Lillard, A. S. (2012). Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs. *Journal of School Psychology*, *50*(3), 379–401. https://doi.org/10.1016/j.jsp.2012.01.001

Lillard, A. S. (2017). *Montessori: The science behind the genius* (3rd ed.). Oxford.

Lillard, A. S. (2019). Shunned and admired: Montessori, self-determination, and a case for radical school reform. *Educational Psychology Review*, *31*(4), 939–965. https://doi.org/10.1007/s10648-019-09483-3

Lillard, A. S., & Eisen, S. (2017). Why Montessori is a facilitative environment for Theory of Mind: Three speculations. In V. Slaughter & M. de Rosnay (Eds.), *Theory of mind development in context* (pp. 57–70). Routledge.

Lillard, A. S., & Heise, M. J. (2016). Removing supplementary materials from Montessori classrooms changed child outcomes. *Journal of Montessori Research*, *2*(1), 17–27. https://doi.org/10.17161/jomr.v2i1.5678

Lillard, A. S., Heise, M. J., Richey, E. M., Tong, X., Hart, A., & Bray, P. M. (2017). Montessori preschool elevates and equalizes child outcomes: A longitudinal study [Original Research]. *Frontiers in Psychology*, *8*, Article 1783. https://doi.org/10.3389/fpsyg.2017.01783

Lillard, A. S., & McHugh, V. (2019a). Authentic Montessori: The *dottoressa*'s view at the end of her life, part I: The environment. *Journal of Montessori Research*, *5*(1), 1–18. https://doi.org/10.17161/jomr.v5i1.7716

Lillard, A. S., & McHugh, V. (2019b). Authentic Montessori: The *dottoressa*'s view at the end of her life, part II: The teacher and the child. *Journal of Montessori Research*, *5*(1), 19–34. https://doi.org/10.17161/jomr.v5i1.9753

Lillard, A. S., Taggart, J., Yonas, D., & Seale, M. N. (in press). An alternative to "no excuses": Considering Montessori as culturally responsive pedagogy. *Journal of Negro Education*. https://www.montessoripublic.org/wp-content/uploads/2021/03/Lillard.21.Taggart.Yonas_.Seale_.JNE_.pdf

Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevance of selected measures. *Journal of Educational Psychology*, *109*(8), 1084–1102. https://doi.org/10.1037/edu0000203

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school: Nationally representative evidence from the ECLS-K: 2011. *Educational Researcher*, *46*(2), 103–109. https://doi.org/10.3102/0013189X17698700

Magnuson, K. A., & Duncan, G. J. (2006). The role of family socioeconomic resources in the black–white test score gap among young children. *Developmental Review*, *26*(4), 365–399. https://doi.org/10.1016/j.dr.2006.06.004

Mann, H. (1848/1961). Twelfth annual report to the Massachusetts Board of Education (1848), together with the twelfth annual report of the secretary of the board. In I. Starr, L. P. Todd, & C. Mercle (Eds.), *Living American documents* (Vol. 1, pp. 154–157). Harcourt Brace. https://archives.lib.state.ma.us/handle/2452/204731

McAlister, A. R., & Peterson, C. C. (2013). Siblings, theory of mind, and executive functioning in children aged 3–6 years: New longitudinal evidence. *Child Development*, *84*(4), 1442–1458. https://www.jstor.org/stable/23469388

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Riverside Publishing.

Miller, P. J. (1996). Instantiating culture through discourse practices: Some personal reflections on socialization and how to study it. In R. Jessor, A. Colby, & R. A. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 183–204). The University of Chicago Press.

Miller, P. J., & Hoogstra, L. (1992). Language as tool in the socialization and apprehension of cultural meanings. In T. Schwartz, G. M. White, & C. A. Lutz (Eds.), *New directions in psychological anthropology* (pp. 83–101). Cambridge University Press.

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*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Montessori, M. (1967). *The discovery of the child* (M. J. Costello, Trans.). Ballantine.

Montessori, M. (2012). *The 1946 London lectures* (B. Scoppola & K. M. Baker, Eds.). Montessori-Pierson Publishing.

Montessori, M. (2020). *Maria Montessori writes to her grandchildren: Letters from India, 1939–1946* (C. Montessori, Trans.). Montessori-Pierson.

Moretti, E. (2021). *The best weapon for peace: Maria Montessori, education, and children's rights*. University of Wisconsin Press.

Murray, A. K., & Daoust, C. (2023). Fidelity issues in Montessori research. In A. K. Murray, E. T. Ahlquist, M. K. McKenna, & M. Debs (Eds.), *The Bloomsbury Handbook of Montessori Education* (pp. 199–208). Bloomsbury Academic.

National Center for Montessori in the Public Sector. (2023). *Montessori census*. https://www.montessoricensus.org/about-the-montessori-census/

Nesbitt, K. T., Baker-Ward, L., & Willoughby, M. T. (2013). Executive function mediates socio-economic and racial differences in early academic achievement. *Early Childhood Research Quarterly, 28*(4), 774–783. https://doi.org/10.1016/j.ecresq.2013.07.005

Paschall, K. W., Gershoff, E. T., & Kuhfeld, M. (2018). A two decade examination of historical race/ethnicity disparities in academic achievement by poverty status. *Journal of Youth and Adolescence, 47*(6), 1164–1177. https://doi.org/10.1007/s10964-017-0800-7

Pianta, R. C., & Ansari, A. (2018). Does attendance in private schools predict student outcomes at age 15? Evidence from a longitudinal study. *Educational Researcher, 47*(7), 419–434. https://doi.org/10.3102/0013189X18785632

Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's school readiness skills across the pre-K year: Associations with teacher–student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology, 66*, 101084. https://doi.org/10.1016/j.appdev.2019.101084

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*(3), 605–619. https://doi.org/10.1037/a0015365

Pottish-Lewis, P. (2021). *Whose Montessori do you practice? A faithfully applied taxonomy of cosmic education*. Montessori360.

Preschlack, P. L. (2023). *The Montessori potential*. Chicago Press.

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educational Researcher, 44*(6), 336–346. https://doi.org/10.3102/0013189X15598539

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 91–116). Russell Sage Foundation.

Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(5), 34–57. https://doi.org/10.7758/rsf.2016.2.5.03

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology, 124*(4), 1164–1221. https://www.journals.uchicago.edu/doi/abs/10.1086/700678

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin, 146*(4), 324–354. https://doi.org/10.1037/bul0000227

Shi, D., & Tong, X. (2017). The impact of prior information on Bayesian latent basis growth model estimation. *SAGE Open, 7*(3). https://doi.org/10.1177/2158244017727039

Singham, M. (2003). The achievement gap: Myths and reality. *Phi Delta Kappan, 84*(8), 586–591. https://www.jstor.org/stable/20440430

Snyder, A., Tong, X., & Lillard, A. S. (2022). Standardized test performance in public Montessori schools. *Journal of School Choice, 16*(1), 105–135. https://doi.org/10.1080/15582159.2021.1958058

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, *1*(1), 91–136. https://doi.org/10.1086/526401

Trabalzini, P. (2011). Maria Montessori through the seasons of the "method." *The NAMTA Journal*, 36(2).

Wang, Y., & Su, Y. (2009). False belief understanding: Children catch it from classmates of different ages. *International Journal of Behavioral Development*, *33*(4), 331–337. https://doi.org/10.1177/0165025409104525

Weimer, A. A., & Guajardo, N. R. (2013). False belief, emotion understanding, and social skills among Head Start and non-Head Start children. *Early Education and Development*, *16*(3), 341–366. https://doi.org/10.1207/s15566935eed1603_3

Wellman, H. M. (2011). Developing a theory of mind. In U. Goswami (Ed.), *Handbook of cognitive development* (pp. 258–284). Blackwell.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tests. *Child Development*, *75*(2), 523–541. https://www.jstor.org/stable/3696656

Whitescarver, K., & Cossentino, J. (2007). Lessons from the periphery: The Role of dispositions in Montessori teacher training. *Journal of Educational Controversy*, *2*(2), 1–12. https://cedar.wwu.edu/jec/vol2/iss2/11/

Willoughby, M. T., Wirth, R., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of Experimental Child Psychology*, *108*, 414–435. https://doi.org/10.1016/j.jecp.2010.04.007

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., Zaslow, M. J., & O Brien, M. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development; Foundation for Child Development. http://www.srcd.org/sites/default/files/documents/washington/mb_2013_10_16_investing_in_children.pdf

Appendix

Descriptive Statistics

Variables	Montessori group			Control group		
	%		Missingness (%)	%		Missingness (%)
Race White	51		0	70		0
Gender male	54		0	53		0
	M	*SD*	Missingness (%)	*M*	*SD*	Missingness (%)
Mother's highest level of education	7.28	4.16	0	6.72	4.95	0
AA1	0.27	2.32	1	-0.34	2.28	2
AA2	0.20	2.34	1	-0.31	2.19	5
AA3	0.43	2.57	1	-0.56	1.82	2
AA4	0.45	2.29	16	-0.65	2.15	8
ToM1	0.81	0.87	1	0.86	0.70	2
ToM2	1.38	0.94	3	1.27	0.90	3
ToM3	2.37	1.14	1	1.88	1.11	2
ToM4	3.16	1.07	18	2.80	1.09	9
EF1	0.10	1.47	4	-0.17	1.33	8
EF2	0.17	1.63	4	-0.23	1.41	6
EF3	0.26	1.63	6	-0.29	1.37	8
EF4	0.23	1.35	16	-0.30	1.85	8

Note. Mother's highest level of education: 1 = eighth grade or less, 2 = ninth grade, 3 = tenth grade, 4 = eleventh grade, 5 = high school, 6 = some college, 7 = 4 years of college, 8 = graduate school; AA = Academic Achievement at Times 1, 2, 3, and 4, respectively (composite of *z* scores); ToM = Theory of Mind scale score at Times 1, 2, 3, and 4, respectively (range from 0 to 4); EF = executive function composite score at Times 1, 2, 3, and 4, respectively (composite of *z* scores).

Latent Growth Curve Model—Academic Achievement (AA)

Parameters	Montessori group			Control group		
	Estimate	*SE*	*p* value	Estimate	*SE*	*p* value
Intercept						
Race	1.799	0.510	0.000	2.187	0.455	0.000
Slope						
Race	-0.156	0.094	0.095	-0.243	0.109	0.026
Gender	-0.091	0.085	0.281	-0.151	0.085	0.076
Income	0.005	0.011	0.662	0.015	0.009	0.106

Note. Montessori group CFI = 1.0; TLI = 1.0; control group CFI = 0.99; TLI = 0.98.

Latent Growth Curve Model—Theory of Mind (ToM)

Parameters	Montessori group			Control group		
	Estimate	*SE*	*p* value	Estimate	*SE*	*p* value
Intercept						
Race	0.320	0.182	0.079	0.332	0.231	0.150
Slope						
Race	0.049	0.052	0.346	0.079	0.076	0.299
Gender	0.011	0.050	0.823	-0.076	0.055	0.172
Income	-0.002	0.006	0.754	0.004	0.006	0.548

Note. Montessori group CFI = 1.0; TLI = 1.0; control group CFI = 0.74; TLI = 0.63.[1]

Latent growth curve model—Executive Function (EF)

Parameters	Montessori group			Control group		
	Estimate	*SE*	*p* value	Estimate	*SE*	*p* value
Intercept						
Race	0.531	0.353	0.132	0.543	0.399	0.110
Slope						
Race	0.001	0.093	0.989	-0.051	0.119	0.667
Gender	-0.004	0.057	0.945	-0.115	0.085	0.177
Income	0.001	0.011	0.936	0.006	0.009	0.483

Note. Montessori group CFI = 1.0; TLI = 1.0; control group CFI = 0.85; TLI = 0.80.[1]

[1] Because the fit indices for ToM and EF models for the control indicated that the two linear growth curve models did not have a good fit, we tried to fit nonlinear growth curve models for this subpopulation to analyze the change of ToM and EF. Given the number of time points and the limited sample size, we could fit only a latent basis growth curve model or a quadratic growth curve model. The latent basis model either did not converge (for EF) or had a similar fit as the linear growth curve model (for ToM). Although the quadratic growth curve models converged and had better fits (CFIs > 0.9, TLIs < 0.9), there were warning messages in Mplus that the latent variable covariance matrix was not positive definite, indicating the model specification was not appropriate for the data.

All model-fit indices are sensitive to sample size. As Lai and Green (2016) discussed, the fit indices by design evaluate the model fit from different perspectives, the cutoff values for the indices are arbitrary, and the meaning of "good" fit and its relationship with fit indices are not well understood. These problems are all the more salient for small samples. Given the relatively small sample size of our data, even if we fit the quadratic growth curve models, we cannot reach a consistent conclusion based on different model-fit indices (e.g., CFI > 0.9, TLI < 0.9). In fact, for EF in the control sample data, when we fit a quadratic model instead of a linear model, CFI increased from 0.85 to 0.92, but TLI decreased from 0.80 to 0.76. The linear growth curve model is parsimonious and consistent with the models for the Montessori group. Therefore, we decided to report the results from the linear growth curve models, although the fit indices are a bit less than the good fit value 0.9. We would like to note that it is a limitation that the linear growth curve models do not fit the EF and ToM data of the control group as well as they fit the data of the Montessori group.