Child Development, September/October 2021, Volume 92, Number 5, Pages 2069–2088

Effects of Montessori Education on the Academic, Cognitive, and Social Development of Disadvantaged Preschoolers: A Randomized Controlled Study in the French Public-School System

Philippine Courtier, Marie-Line Gardes, and Jean-Baptiste Van der Henst
CNRS, INSERM, & Université de Lyon

Ira A. Noveck
CNRS & Université de Paris

Marie-Caroline Croset
Université de Grenoble

Justine Epinat-Duclos, Jessica Léone, and Jérôme Prado
CNRS, INSERM, & Université de Lyon

Previous research on Montessori preschool education is inconsistent and prone to analytic flexibility. In this preregistered study, disadvantaged preschoolers in a French public school were randomly assigned to either conventional or Montessori classrooms, with the latter being adapted to French public education. Adaptations included fewer materials, shorter work periods, and relatively limited Montessori teacher training. Cross-sectional analyses in kindergarten (N = 176; M_{age} = 5–6) and longitudinal analyses over the 3 years of preschool (N = 70; M_{age} = 3–6) showed that the adapted Montessori curriculum was associated with outcomes comparable to the conventional curriculum on math, executive functions, and social skills. However, disadvantaged kindergarteners from Montessori classrooms outperformed their peers on reading (d = 0.68). This performance was comparable to that of advantaged children from an accredited Montessori preschool.

Social and economic disparities have a tremendous impact on the academic, cognitive, and socioemotional development of children and adolescents. For example, in the United States and across Europe, socioeconomic status (SES) explains up to 15% of the variance in reading and math performance of high-schoolers (OECD, 2016). Disadvantaged children also lag behind more advantaged peers on measures of cognitive and socioemotional abilities (Farah, 2017). These SES-related differences emerge as early as age 4 and tend to grow over time (Jordan & Levine, 2009; OECD, 2016). Thus, it is increasingly believed that interventions aimed at reducing inequalities should focus on early childhood, a period of enhanced brain plasticity during which cognitive skills may be the most responsive to cognitive and social stimulation (Thomas & Knowland, 2009). This has led to a renewed interest in preschool programs as a way to reduce the influence of SES disparities on early childhood development (Campbell et al., 2014; Duncan & Magnuson, 2013).

Interest in early childhood education emerges from a large body of evidence showing that structured preschool programs can have a positive impact on school readiness and cognitive development of disadvantaged children (Campbell et al., 2014; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010). To date, policies have largely focused on structural factors such as teacher qualification (Jackson, Rockoff,

We owe a debt of gratitude to A. Gascher and all of the teachers from both schools. This research would not have been possible without their involvement and dedication to the project. We also thank A. Aguesse, L. Anguenot, M. Barilaro, C. Belanger, E. Bonassi, J. Bouvier, L. Brusset, A. Couderc, M. Coux, O. Debain, M. Debus, F. Desbois, M. Duclos, A. Dussol, J. Fassaya, M. Jacquel, Y. Jay, L. Longo, A. Martin, C. Murat, M. Patureau, J. Philip, M. Pottelain, J. Prost, M. Rengade, S. Reynaud, I. Rouine, L. Ruty, E. Schuld, S. Serhane, M. Sordet, S. Souici, M. Vigne, A. Vittoz, C. Wable and E. Zimmerman for their assistance in collecting the data. Finally, we are grateful to all of the children who participated in this study. This research was supported by funding from the Institut Carnot de l'Éducation Auvergne Rhône Alpes (ICE—AuRA PR06). The content is solely the responsibility of the authors and does not necessarily represent the official view of the funder.

Correspondence concerning this article should be addressed to Philippine Courtier or Jérôme Prado, Centre de Recherche en Neurosciences de Lyon, CH Le Vinatier, 95 bd Pinel, 69675 Bron Cedex, France. Electronic mail may be sent to philippine.courtier@gmail.com or jerome.prado@univ-lyon1.fr

0009-3920/2021/9205-0047
DOI: 10.1111/cdev.13575

& Staiger, 2014), class sizes (Chetty et al., 2011), or physical learning environment (Sabol, Soliday Hong, Pianta, & Burchinal, 2013). But preschool programs may also vary widely in terms of content and type of instruction (i.e., the curriculum) that children are exposed to (Jenkins et al., 2018). For instance, preschool curricula might differ with respect to pedagogical tools and materials (e.g., concrete vs. abstract), role of the teacher (e.g., leading activities vs. supporting children-led activities), organization of instruction (e.g., small or whole group vs. individualized), and use of assessments (e.g., explicit vs. implicit). Yet, relatively little is known about the preschool curricula that are the most effective, particularly for low SES children (Jenkins et al., 2018).

The Montessori Preschool Curriculum

One particular preschool curriculum that is currently experiencing a surge in popularity across many Western countries is the Montessori method. In the United States alone, Montessori programs have expanded from a small number of private schools in the 1970s to approximately 4,500 schools nowadays (including more than 500 public schools; Debs, 2016). The Montessori approach is named after Maria Montessori, an early 20th-century Italian physician, who developed a pedagogical method initially aimed at helping children with intellectual disability as well as typically developing preschoolers from low-income families in Rome. Montessori education revolves around the idea that children "best learn by absorbing and interacting with different aspects of their environment, as opposed to being directly taught specific knowledge and skills" (Ackerman, 2019, p. 2). As such, Montessori education relies on a specific organization and physical environment, with mixed-age classrooms that are open and have well-defined spaces for different parts of the curriculum (e.g., language, math; Lillard & McHugh, 2019a). It also involves *specific pedagogical tools, including a set of highly organized multisensory and auto-corrective materials that are presented by highly trained teachers and are designed to promote learning* through action and manipulation at the child's own pace (Lillard & McHugh, 2019a). Because Montessori pedagogy aims to promote self-directed learning, children are allowed to freely choose their activities. Accordingly, the role of Montessori teachers is more to guide and support each child's learning trajectory than to directly teach content (Lillard & McHugh, 2019b).

Overall, Montessori preschool education involves methods that often differ from those found in most *conventional* preschool classrooms. Indeed, conventional classrooms tend to have same-aged children who are placed in small or whole group, and who make use of a more diverse set of play materials (Lillard, 2012). Yet, there have been recent claims that the Montessori preschool curriculum is more consistent with principles of learning and development than conventional curricula (Lillard, 2019; Marshall, 2017). For example, children in Montessori classrooms systematically learn to read using multisensorial activities and a phonetic approach, consistent with the literature on both embodied cognition (Kontra, Goldin-Meadow, & Beilock, 2012; Pouw, van Gog, & Paas, 2014) and reading acquisition (Castles, Rastle, & Nation, 2018). The Montessori math curriculum also stresses the importance of understanding the correspondence between numerical symbols and quantities, in keeping with studies showing that early symbolic and arithmetic knowledge predict later math competence (Jordan, Kaplan, Ramineni, & Locuniak, 2009; Merkley & Ansari, 2016). Because Montessori classrooms are highly organized and involve a relatively strict set of rules and principles, it has also been argued that Montessori education may promote the growth of executive functions (Diamond & Lee, 2011; Lillard, 2019). Finally, open and multiage Montessori classrooms may offer an environment with rich, collaborative, and varied social interactions that may foster early socioemotional skills (McClellan & Kinsey, 1997). Thus, there are reasons to believe that Montessori preschool education may lead to greater academic, cognitive, and social outcomes than conventional education in young children.

Previous Research on Montessori Preschool Education

However, research evaluating the impact of Montessori education is scarce (see Appendix S1 for a list of recent studies). Findings are also inconsistent. For instance, while some studies have found that children in Montessori classrooms have better reading (Lillard, 2012; Lillard & Else-Quest, 2006), math (Chisnall & Maher, 2007; Denervaud, Knebel, Hagmann, & Gentaz, 2019; Laski, Vasilyeva, & Schiffman, 2016; Lillard & Else-Quest, 2006), executive functions (Denervaud et al., 2019; Lillard, 2012; Lillard & Else-Quest, 2006), and social skills (Lillard, 2012; Lillard & Else-Quest, 2006; Lillard et al., 2017) than children in conventional classrooms, other studies have found no advantage on similar measures of math (Chisnall & Maher, 2007; Lillard, 2012; Mix, Smith, Stockton, Cheng, & Barterian, 2017), executive functions (Denervaud et al., 2019; Lillard & Else-Quest, 2006; Lillard et al., 2017), and social skills (Lillard, 2012; Lillard et al., 2017). Figure S1 summarizes the number of previously reported effects in favor or against Montessori preschool education as a function of the domain.

Furthermore, prior studies have five flaws that limit the generalization of their findings (Marshall, 2017). First, parents who voluntarily choose a Montessori preschool for their children are likely to differ from other parents with regard to their involvement in education. Those parents who choose Montessori are arguably not representative of most families. One way to confront this would be to (a) randomly assign students to either a Montessori or a conventional classroom and; (b) investigate children whose parents do not specifically choose a Montessori preschool. However, most prior studies have not used a random assignment of children (thus confounding effects of pedagogy with parental involvement; for a review see Ackerman, 2019). Additionally, the two studies that did use a lottery design (Lillard & Else-Quest, 2006; Lillard et al., 2017) only included children of parents who opted for enrollment in a Montessori preschool (raising doubts about the representativeness of the study's sample).

Second, because most Montessori preschools are tuition-based, they largely enroll children from high-income families. Even public Montessori preschools tend to attract families with a higher income than other preschools (Culclasure, Fleming, & Sprogis, 2018; Debs & Brown, 2017). As a result, most prior studies have examined the effect of Montessori education on high SES children, which again makes the findings difficult to generalize to lower income children (but see Lillard & Else-Quest, 2006; Lillard et al., 2017).

Third, preschools greatly vary in the way they implement Montessori education, some being more faithful to the pedagogy than others. This factor may explain some of the inconsistencies observed in prior studies. Yet, an objective measure of the fidelity of implementation of the Montessori pedagogy has only been collected in a few studies (Lillard, 2012; Yen & Ispa, 2000). Likewise, previous studies have mainly been conducted in the United States (with some exceptions, e.g., Chisnall & Maher, 2007; Denervaud et al., 2019), where there is great heterogeneity in preschool programs and curricula. Thus, it can be difficult to evaluate to what extent conventional classrooms may adhere to a homogenous curriculum.

Fourth, prior studies employed frequentist statistics to compare Montessori to conventional classrooms, which can make it difficult to know whether a lack of difference between groups may support the null hypothesis (i.e., no difference between Montessori and preschool education; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008) or may reflect a lack of power.

Finally, prior studies often involve a number of different variables that may not always be analyzed in the same way (e.g., similar measures may be averaged in one study and not in another; Lillard & Else-Quest, 2006; Lillard et al., 2017). This raises questions about analytical flexibility and makes it difficult to parse out exploratory from confirmatory results (Simmons, Nelson, & Simonsohn, 2011). Yet, to our knowledge, no prior study on Montessori education has preregistered its hypotheses and methods.

Current Study

While addressing the shortcomings of prior studies listed above, this study takes advantage of a unique experimental context in which the Montessori curriculum was implemented in some (but not all) classrooms of a disadvantaged public preschool located in the Lyon area in France. All children from that preschool were randomly assigned (by the school) to either Montessori or conventional classrooms. Over the course of 4 years, children were tested on a range of tasks assessing language, mathematical, executive, and social abilities. In a cross-sectional experiment (Experiment 1), these measures were used to compare the skills of children from Montessori classrooms to those in conventional classrooms at the end of kindergarten. In a longitudinal experiment (Experiment 2), the same measures were used to track the progress of children from preschool entry to the end of kindergarten (i.e., 3 years of education).

We preregistered hypotheses (based on the cognitive literature and prior studies evaluating Montessori education), measures, and analytic strategy for both experiments. Thus, this study reflects a primarily confirmatory effort. Data were analyzed using both frequentist and Bayesian statistics. Children from conventional classrooms followed the French preschool curriculum, administered by teachers with the same academic training. This ensured homogeneity in the control group (see Appendix S3). Children from Montessori classrooms followed the Montessori preschool curriculum, which was adapted to the constraints of public education in France. Such adaptations can be unavoidable because public education will often impose pedagogical and material constraints on programs. To ensure that our findings would be reproducible and generalizable to other contexts, we took the following three measures. First, we observed and reported the way Montessori education was implemented in public classrooms (see Appendix S3). Second, we quantified the fidelity of implementation of the Montessori pedagogy in the public preschool using a scale developed based on the Montessori curriculum and previous research. Finally, we compared the fidelity of implementation of the Montessori method in the public preschool to that of private Montessori preschools in the Lyon area. We also compared the performance of low SES children from the public preschool to the performance of high SES children from a private Montessori preschool accredited by the French affiliate of the Association Montessori Internationale (AMI). The AMI, which is the organization Maria Montessori and her son founded to carry on her work, is widely believed to have the highest standards for Montessori teacher training and implementation (Lillard, 2012).

Experiment 1

Method

Participants

Participants were 196 five- to six-year-olds in kindergarten recruited from two preschools in the Lyon area (France). One preschool was part of the 3-year public preschool system (*écoles maternelles*) attended by most 3- to 6-year-olds in France (https://www.insee.fr/fr/statistiques/2383587). This public preschool was located in a neighborhood in which the median equivalized disposable income is €15,208, which is largely below the national median equivalized disposable income (i.e., €20,809, https://www.insee.fr/fr/statistiques/4195239). About 33% of individuals living in this area have an income that is below the French poverty threshold. As such, the school is part of the French "Reinforced Priority Education Network," a network of schools in which students and teachers benefit from various types of support aimed at reducing inequalities and enhancing the academic achievement of underprivileged students. Enrollment in public preschools, like everywhere in France, is reserved for children living in the neighborhood and cost-free. Three classrooms in this public preschool followed an adaptation of the Montessori curriculum to the context of the French public education system (i.e., *Montessori-public* group), whereas five or six classrooms (depending on the year) followed the conventional French public curriculum (i.e., *Conventional-public* group). All children included in the present analyses (see below) were assigned to one of these classrooms by the school when they first enrolled into the public preschool at age 3. This assignment was fully randomized except for children with disabilities and staff children (those were excluded from the study).

The other preschool (i.e., *Montessori-private* group) included in the study was a private preschool from a neighborhood in which the median equivalized disposable income is €30,020, which is well above the French national annual median income (see above). Only 8% of individuals living in this area have an income that is below the poverty threshold. This private preschool received an official accreditation from the Association Montessori de France, which is the French affiliate of the AMI. As such, it followed a rigorous and high-fidelity Montessori curriculum. The school does not receive any public funding and entirely relies on tuitions from parents (€5,100/year).

Data were collected from three cohorts of children in these two schools at the end of their kindergarten year, that is, June 2017, June 2018, and June 2019. The experiment was approved by the local school board and was performed in accordance with the ethical standards established by the Declaration of Helsinki. Parents gave their written informed consent and children gave their assent to participate in the experiment. Out of the 210 children, who had been enrolled in their respective schools for 3 years, 14 did not participate in the study because their parents did not give their consent. From the original sample of 196 children, children were excluded if they were not fluent in French ($N = 3$), had a diagnosed disability ($N = 4$), changed pedagogy at some point ($N = 12$), and were related to the staff ($N = 1$). Therefore, all children included in the analyses had received either 3 years of conventional education (which is standard in the French public preschool system) or 3 years of Montessori education (which is also standard in Montessori education). Our final sample comprised 176 children from age 5 to age 6 (86 females). There were 45 children in the Montessori-private group, 53 in the Montessori-public group, and 78 in the Conventional-public group. Detailed sample demographics are shown in Table S1.

Participants from the three groups did not differ with respect to age ($BF_{01} = 14.29$, $F(2, 173) = 0.20$, $p = .82$) and gender ($BF_{01} = 12.7$, $\chi^2(2, N = 176) = 1.07$, $p = .59$). Background information was also collected from parents of a subset of children (see Appendix S2 for detailed information). Consistent with the demographics of the school areas (see above), the SES of children from the public preschool was relatively low, with an average monthly household income between €1,500 and €3,000. In contrast, the SES of children from the private preschool was higher, with an average monthly household income between €3,000 and €5,000. Data on participant ethnicity were not collected because the collection of such data is in principle illegal in France and would require an exceptional waiver from government agencies that we did not seek.

A full description of the characteristics of both Montessori and conventional classrooms, as well as a description and results of the scale used to assess the fidelity of implementation of Montessori education, is given in Appendix S3 and Table S2. Briefly, both the Montessori-private and Montessori-public groups scored relatively high (i.e., above 80% fidelity) on the *characteristics* and *activities* scales. However, due to budgetary constraints, Montessori materials were found to be less numerous in the Montessori-public group than in the Montessori-private group (i.e., 60% vs. 91% fidelity). A complete list of the materials present in the Montessori-public group is given in Appendix S3. Furthermore, to be allowed to operate within the French public education system, the Montessori classrooms in the public school ultimately differed from those in the private school in two respects. First, the public preschool had fewer hours of instruction per day than the private preschool due partly to daily gym requirements in the French Public schools. For these reasons, daily work periods were shorter in the Montessori-public group than in the Montessori-private group (1.5–2 hr vs. 2.5 hr, respectively). Second, unlike teachers in the Montessori-private group, none of the teachers in the Montessori-public school were trained in an AMI center at the beginning of the experiment. They only held a degree from a conventional teacher's college (as is minimally required in France) and were initially only largely self-trained in Montessori education (see Appendix S3 for details about this training). By the end of the study, however, one teacher in the Montessori-public group completed AMI training. Therefore, the Montessori curriculum might be considered to be of lower fidelity in the Montessori-public than in the Montessori-private group. However, Montessori classrooms from the public preschool were still far more faithful to Montessori education than the conventional classrooms, which scored very low on our fidelity scale (see Table S2).

Preregistration

Measures and analytic strategy were preregistered via the Open Science Framework (OSF) at https://osf.io/z2cre. Two additions and a set of modifications were made to the anticipated analyses since the project was preregistered. First, frequentist analyses are presented in addition to the preregistered Bayesian statistics. Second, we decided to include the Montessori-private group (which was not preregistered) in Experiment 1 because it provides a useful benchmark against which performance of the Montessori-public and Conventional-public groups can be assessed. Thus, one-way analyses of variance (ANOVAs) were used instead of the preregistered *t*-tests. Finally, some changes were made to the calculation of a few scores to be consistent with prior studies and for the sake of parsimony (these are detailed in Tables 1 and 3).

Materials

Children were tested on language, math, and social competences, as well as on executive functions. The tests assessing language and math skills are described in Table 1. All these tests except the Tokens test had been used in prior research. The Tokens test was developed to measure the overall quantitative knowledge of children. It was based on math standards for children in kindergarten in France (Ministère de l'Education Nationale & de la Jeunesse et des Sports, 2015). The tests assessing executive functions and social skills (also used in prior research) are described in Tables 2 and 3.

In a subset of participants (i.e., children from the last 2 cohorts), we also measured well-being in school with an adaptation of the Feeling about School (FAS) scale (Valeski & Stipek, 2001). The test measured children's subjective perception of four different areas of their school experience: their competence in reading, their competence in mathematics, their relationship with their teacher, and their general attitude toward school. A fifth area was added for the purpose of this study: their relationship with their peers. For each area, children answered three questions (e.g., "How much do you

Table 1
Descriptions of Tests Assessing Language and Math Abilities

Skill	Test	Description	Scoring
Language			
Vocabulary	Evaluation du Langage Oral (Khomsi, 2001)	Determine which of four pictures corresponds to a word given by the experimenter	Number of items correctly answered (0–20)
Phonological awareness	Evaluation des fonctions cognitives et Apprentissages (Phonologie subtest; Billard & Touzin, 2012)	Repeat pseudo-words said aloud by the experimenter. There were four two-syllable pseudo-words, four three-syllable pseudo-words and two four-syllable pseudo-words to repeat	Number of correctly repeated syllables (0–28)
Reading	Evaluation Des fonctions cognitives et des Apprentissages (Lecture subtest; Billard & Touzin, 2012)	Decode letters, then digraphs, increasingly difficult words, and finally sentences. The test was stopped if a child was not able to read any of the letters or any of the digraphs	Number of correctly completed items (0–70)
Pragmatics	Scalar task (Stiller, Goodman, & Frank, 2015)	Make pragmatic inferences by constructing contextually derived "ad-hoc" implicatures, using sets of pictures with contrasting features	Proportion of plausible choices (non-errors) that are revealing of one-feature item choices, i.e., [pragmatic/(pragmatic + logical) responses]
Math			
Math problem solving	Woodcock-Johnson III (Applied Problems subtest; Woodcock, Mather, McGrew, & Wendling, 2001)	Respond to 63 increasingly difficult math problems. The first items involved counting, simple subtraction and addition, clock reading and calculating with coins. Items then progress onto verbal problems. The test is stopped after participants make six consecutive errors	Number of correctly completed items (0–63)
Counting knowledge	Counting task (adapted from Lipton & Spelke, 2005)	Count twice to the highest number known	Highest number reached
Quantitative knowledge	Tokens	Perform 10 quantitative tasks. Children have to recognize the numerosity of a given set, create a set given a specific numerosity, solve simple non-symbolic arithmetic problems, create a set of the same numerosity as another distant set, compare the numerosities of two sets, recognize the numerosity of a set created by the addition of two other sets, recognize the numerosity of a set created by the subtraction of one set from another, recognize number symbols, recognize the position of a number within a sequence, and use ordinal information to identify a position in a sequence	Overall score from 0 to 10

Note. Scores were all consistent with preregistration, with two exceptions. In the counting and tokens tasks, the counting knowledge score was dissociated from the quantitative knowledge score to facilitate interpretation. In the scalar task, we focused on a global score for the sake of parsimony.

like going to school?") with a five-point Likert scale (e.g., from "I don't like it at all" to "I like it a lot"). Before starting the test, children had three practice questions to get familiarized with the scale. Raw scores were the mean total scores for each area on the Likert scale.

Table 2
Descriptions of Tests Assessing Executive Functions

Skill	Test	Description	Scoring
Short-term and working memory	Corsi Block Tapping task (Corsi, 1972)	Repeat a sequence shown by the experimenter by touching blocks glued onto a wooden board. The sequence progressively increases in difficulty. The sequence is first repeated in the same order as the experimenter, then in reverse order. For each sequence, there are four chances to succeed	Maximum number of blocks correctly repeated forward (score = 0–9) for short-term memory and maximum number of blocks correctly repeated backward for working memory (score = 0–6)
Self-regulation	Head Toes Knees Shoulders task (Ponitz, McClelland, Matthews, & Morrison, 2009)	Perform a gesture that is opposite to the one shown by the experimenter (e.g., touch the toes when asked to touch the head and the other way around). The test is made progressively more difficult by adding knees and shoulders commands. Children get two points every time they directly performed the correct action, one point if they have to self-correct their action and no point if they perform an incorrect action	Sum of points (0–52)
Planning	Evaluation Des fonctions cognitives et des Apprentissages (Planification subtest; Billard & Touzin, 2012)	Complete three mazes of progressive difficulty. Each completed maze is worth 10 points. Children have a maximum of 120 s to complete each maze and lose one or two points depending on their completion time. The test is stopped if a child scored 0 at a maze	Sum of the points for all the mazes completed (0–30)

Note. Scores were all consistent with preregistration.

Procedure

Children were tested individually in a quiet room of their preschool. The tests were administered by different experimenters (graduate students, research assistants, and undergraduate students) in five sessions of approximately 15–20 min. The order of sessions was randomized on a child-to-child basis. No feedback was given to children during testing. Children, teachers, and parents were also not informed about the results during the 3 years of the study. Finally, teachers and parents were blind to the tests until the end of data collection.

Analyses

Data were analyzed using both frequentist and Bayesian statistics. First, test scores were entered in a series of one-way frequentist ANOVAs with the between-subject factor Group (Conventional-public, Montessori-public, Montessori-private). To account for the number of tests, the main effect of group was only considered significant at a Bonferroni-corrected significance threshold of $p = .002$ (i.e., 0.05/24 tests). In case of a main effect of group, the analysis of variance (ANOVA) was followed up by independent *t*-tests between groups. Frequentist statistics, however, cannot provide evidence for a null hypothesis. In other words, a nonsignificant effect of the group for a given test would not mean that a group difference does not exist. Therefore, we turned to Bayesian statistics to estimate the strength of evidence (i.e., the Bayes factor, BF) for the alternate hypothesis of a difference between groups (H1) versus the null hypothesis of no difference between groups (H0) for each test.

Test scores were entered into a series of Bayesian ANOVAs with the between-subject factor Group (Conventional-public, Montessori-public, Montessori-private). Following Jeffreys (1961), a $\mathrm{BF} < 3$ was considered anecdotal evidence, a $3 < \mathrm{BF} < 10$ was considered substantial evidence, a $10 < \mathrm{BF} < 30$ was considered strong evidence, a $30 < \mathrm{BF} < 100$ was considered very strong evidence, and a $\mathrm{BF} > 100$ was considered extreme evidence that our data are more likely under the alternate than the

Table 3
Descriptions of Tests Assessing Social Skills

Skill	Test	Description	Scoring
Sharing	Dictator Game (Benenson, Pascoe, & Radmore, 2007)	Share stickers with classmates. Specifically, children first choose their 10 favorite stickers among a choice of 30. They are then told that, unfortunately, not enough stickers are available for all of their classmates. But participants could donate any number of their stickers to their classmates (the choice is made while the experimenter is looking away)	Number of stickers given (0–10)
Distributive justice	Resource allocation task (inspired from Huppert et al., 2018)	Distribute candies to characters in situations of inequality. Three situations are presented (resource inequality between a rich and a poor character, unequal contribution to a common good between a hard-working and a lazy character, and a power inequality between a dominant and a subordinate character). At the end of each situation, children are told that both characters love candies and they have to distribute 4 candies to the characters the way they want	Number of candies allocated to the disadvantaged characters in each of the three conditions (0–4) and in total (0–12)
Social competence	Social Problem Solving Task-Revised (object acquisition item; Rubin, 1988)	Solve a problem involving a social component. Two cartoon children are shown, one reading a book and the other standing behind him/her (characters gender matched the gender of the participants). Children are told that the character reading the book has been reading it for a very long time and that the other character wants to have the book. Participants have to imagine what the character could say or do to get the book, what else the character could say or do to get the book, and what they could themselves say or do if they want a book that another child is already using. The same story was used at each testing session	Number of responses with reference(s) to justice/sharing (0–3; Lillard, 2012), as well as the flexibility (i.e., level of novelty) of the second response in relation to the first response (0–3; Rubin & Clarke, 1983)
Theory of mind	Wellman and Liu (2004) task	Understand the mental state of a character. The task involves 5 stories measuring the complexity of theory of mind understanding (diverse desire, diverse belief, diverse knowledge, false belief, hidden emotion). The test is stopped after two failed stories	Number of stories completed successfully (0–5)

Note. Scores were all consistent with preregistration, with two exceptions. In the dictator game, the number of stickers given was used rather than the number of stickers kept to be consistent with previous studies (Benenson et al., 2007). In the resource allocation task, the proportion of children who allocated the first candy to the disadvantaged characters was not included for the sake of parsimony.

null hypothesis (i.e., BF_{10}) or under the null hypothesis than the alternate hypothesis (i.e., BF_{01}). Given that previous studies provide inconsistent evidence for the effects of Montessori education, we did not choose an informative prior but rather used the default Cauchy distribution prior centered on the null with a width of 0.707 to quantify the evidence for the alternate hypothesis (that there is a difference between groups). An effect size of 0 as prior was used to quantify evidence for the null hypothesis (that there is no difference between groups). In case of more than anecdotal evidence for a main effect of Group, the ANOVA was followed up by Bayesian independent *t*-tests (also with default priors). All analyses were performed with the JAMOVI software (The Jamovi Project, 2019).

Data Availability

The data that support the findings of Experiment 1 are available via the OSF at https://osf.io/cbr97/ .

Results and Discussion

Frequentist Results

Results from the frequentist one-way ANOVAs are indicated in Table 4 (see Table S3 for exact mean performance on each test).

First, a significant main effect of Group was observed on the reading test (see Figure 1A). Follow-up independent *t*-tests revealed that children from the Conventional-public group had lower reading scores than children from the Montessori-private group ($t(120) = 4.37$, $p < .001$, $d = 0.82$) and children from the Montessori-public group ($t(128) = 3.83$, $p < .001$, $d = 0.68$). There was no difference between the two Montessori groups ($t(96) = 0.42$, $p = .67$, $d = 0.09$). Therefore, within the public preschool (in which classroom assignment was randomized), the reading skills of children from Montessori classrooms were higher than the reading skills of children from conventional classrooms. The size of this effect ($d = 0.68$) is considered very large in education interventions (Kraft, 2020).

A similar pattern was obtained on the scale measuring the subjective perception of reading competence, for which a main effect of group was also observed (see Figure 1B). That is, children from the Conventional-public group reported being less competent with reading than children from both the Montessori-private group ($t(67) = 2.90$, $p = .005$, $d = 0.71$) and the Montessori-public group ($t(74) = 4.04$, $p < .001$, $d = 0.93$). There was no difference between the two Montessori groups with respect to this perceived competence with reading ($t(61) = 0.92$, $p = .36$, $d = 0.23$). Thus, not only were children from Montessori public classrooms better readers than children from conventional public classrooms, they were also aware of their relatedly high competence.

Second, there was a main effect of Group on the tests measuring math problem solving and quantitative knowledge (see Figures 2A and 2B). However, follow-up analyses indicated that this effect was entirely driven by the higher scores of children from the Montessori-private group. Specifically, children from the Montessori-private group had higher math problem solving and quantitative knowledge scores than children from both the Montessori-public group (math problem solving: $t(96) = 4.50$, $p < .001$, $d = 0.91$; quantitative knowledge: $t(96) = 4.68$, $p < .001$, $d = 0.96$) and the Conventional-public group (math problem solving: $t(120) = 4.60$, $p < .001$, $d = 0.86$; quantitative knowledge: $t(118) = 3.15$, $p = .002$, $d = 0.60$). There was no difference between the two groups within the public school (math problem solving: $t(128) = 0.12$, $p = .91$, $d = 0.02$; quantitative knowledge: $t(126) = 1.82$, $p = .07$, $d = 0.33$). Thus, there was no evidence that Montessori education was associated with greater math skills, as greater performance of children from the Montessori-private group may be attributable to parental differences between the public and the private school (e.g., income or involvement in education).

Third, there was a main effect of Group on the measure of working memory (see Figure 2C). Similar to math skills, however, the effect was entirely driven by the Montessori-private group. *T*-tests revealed that children in the Montessori-private group had higher working memory skills than children in both the Montessori-public group ($t(96) = 5.07$, $p < .001$, $d = 1.03$) and the Conventional-public group ($t(120) = 4.57$, $p < .001$, $d = 0.86$). There was no difference between the two groups within the public school ($t(128) = 0.36$, $p = .72$, $d = 0.06$). Thus, there was no evidence that Montessori education was associated with greater executive functions than conventional education.

The main effect of Group was significant in none of the other tests (see Table 4).

Bayesian Results

Results from the Bayesian one-way ANOVAs are indicated in Table 4. Of the 24 tests, 20 led to at least substantial evidence for either the null (BF_{01}) or the alternative (BF_{10}). There was at least substantial evidence for a lack of difference between the

Table 4
Experiment 1: Main Effect of Group (Montessori-Public, Conventional-Public, Montessori-Private) for Each Test in the Frequentist and Bayesian ANOVAs

Skill	*df*	*F*	*p*	η_p^2	BF_{10}	BF_{01}
Language						
Vocabulary	2, 172	5.77	.004	.06	**7.9**	0.13
Phonological awareness	2, 172	3.22	.042	.04	0.95	1.08
Reading	2, 172	**10.70**	**< .001**	**.11**	**10.70**	0.09
Pragmatic skills	2, 167	0.93	.40	.01	0.14	**7.14**
Math						
Math problem solving	2, 172	**13.28**	**< .001**	**.13**	**> 100**	< 0.01
Counting knowledge	2, 169	0.38	.69	< .01	0.08	**12.5**
Quantitative knowledge	2, 169	**10.19**	**< .001**	**.11**	**> 100**	< 0.01
Executive functions						
Short-term memory	2, 172	3.81	.02	.04	1.52	0.65
Working memory	2, 172	**13.76**	**< .001**	**.14**	**> 100**	< 0.01
Self-regulation	2, 172	4.95	.008	.05	**4.19**	0.23
Planning	2, 101	1.59	.21	.03	0.32	**3.13**
Social abilities						
Sharing	2, 173	0.48	.62	.01	0.09	**11.11**
Distributive justice (overall)	2, 173	1.68	.19	.02	0.25	**4**
Distributive justice (resources)	2, 173	0.60	.55	.01	0.1	**10**
Distributive justice (common good)	2, 173	1.52	.22	.02	0.22	**4.55**
Distributive justice (power)	2, 173	2.10	.13	.02	0.35	2.86
Social competence (justice)	2, 173	1.68	.19	.02	0.25	**4**
Social competence (flexibility)	2, 173	4.62	.01	.05	**3.30**	0.30
Theory of mind	2, 173	2.62	.08	.03	0.54	1.85
Feeling about school						
Teacher relationship	2, 101	0.82	.44	.02	0.17	**5.79**
Peer relationship	2, 101	1.27	.28	.02	0.25	**3.95**
Reading competence	2, 101	**9.90**	**< .001**	**.16**	**> 100**	< 0.01
Math competence	2, 101	0.62	.54	.01	0.14	**6.72**
General attitude	2, 101	0.66	.52	.01	0.15	**6.58**

Note. Frequentist statistics: The Bonferroni-corrected significance threshold was $p < .002$. Significant results are in bold. η_p^2s represent effect sizes that can be considered small (.01), medium (.06) or large (.14; Cohen, 1988). Degrees of freedom vary because sample sizes change depending on the measure (some participants were absent or refused to complete certain tests). Bayesian statistics: BFs > 3 are indicated in bold. BF_{10} indicates the strength of the evidence for the alternative (there is a group difference), whereas BF_{01} values indicate the evidence for the null (there is no group difference). ANOVAs = analyses of variance.

groups in the tests measuring pragmatic skills, counting knowledge, planning, sharing, distributive justice (overall score as well as resources and common goods subtests), social competence (reference to justice), and all measures of FAS except subjective perception of reading competence.

In contrast, there was at least substantial evidence for a difference between the groups in the tests measuring vocabulary, reading, math problem-solving, quantitative knowledge, working memory, self-regulation, social competence (flexibility of the response), and subjective perception of reading competence (see Table 4). In the test measuring reading (see Figure 1A), follow-up *t*-tests indicated extreme evidence that children from the Conventional-public group had lower scores than children from both the Montessori-private groups ($BF_{10} > 100$) and the Montessori-public group ($BF_{10} > 100$). There was also substantial evidence for a lack of difference between the Montessori-public and the Montessori-private groups ($BF_{01} = 4.35$). Therefore, Bayesian analyses not only show that children from the Montessori-public group have higher reading skills than children from the Conventional-public group, they also indicate that children from the Montessori-public group have similar reading skills than children from the Montessori-private group. Considering the gap in reading achievement typically observed between children from different SES backgrounds (Buckingham, Wheldall, & Beaman-Wheldall, 2013), this lack of difference is noticeable.

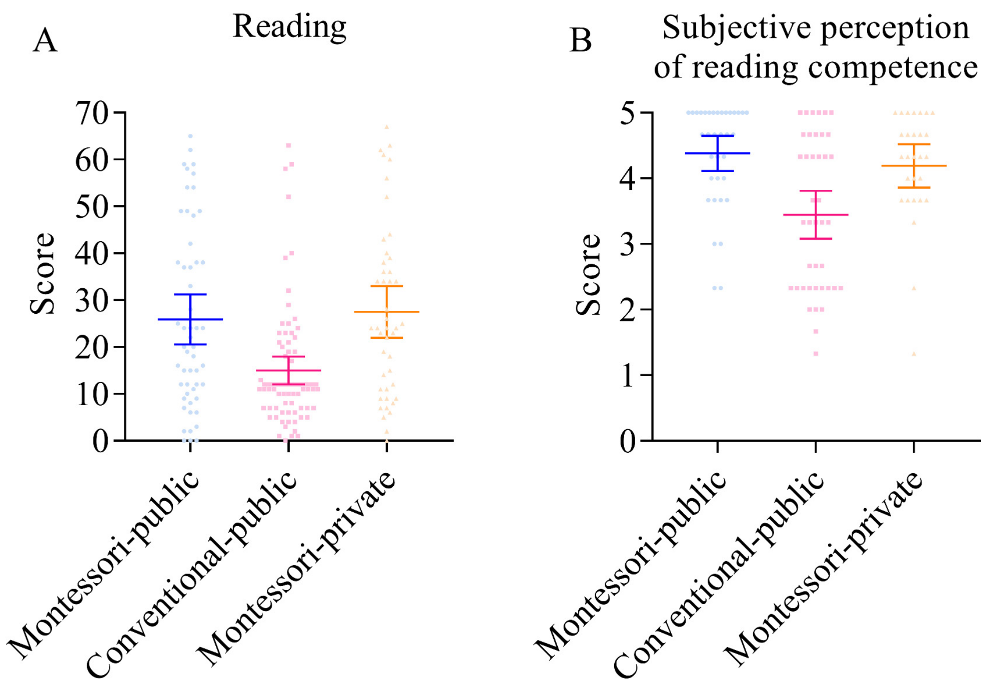

Figure 1. Experiment 1: Distribution of scores in the two measures for which the average score was higher in children from the Montessori-public group than from the Conventional-public group. (A) Reading. (B) Subjective perception of reading competence. Each blue, pink, and orange dot represents the score of a child in the Montessori-public, Conventional-public, Montessori-private group (respectively). Error bars represent 95% confidence intervals. [Color figure can be viewed at wileyonlinelibrary.com]

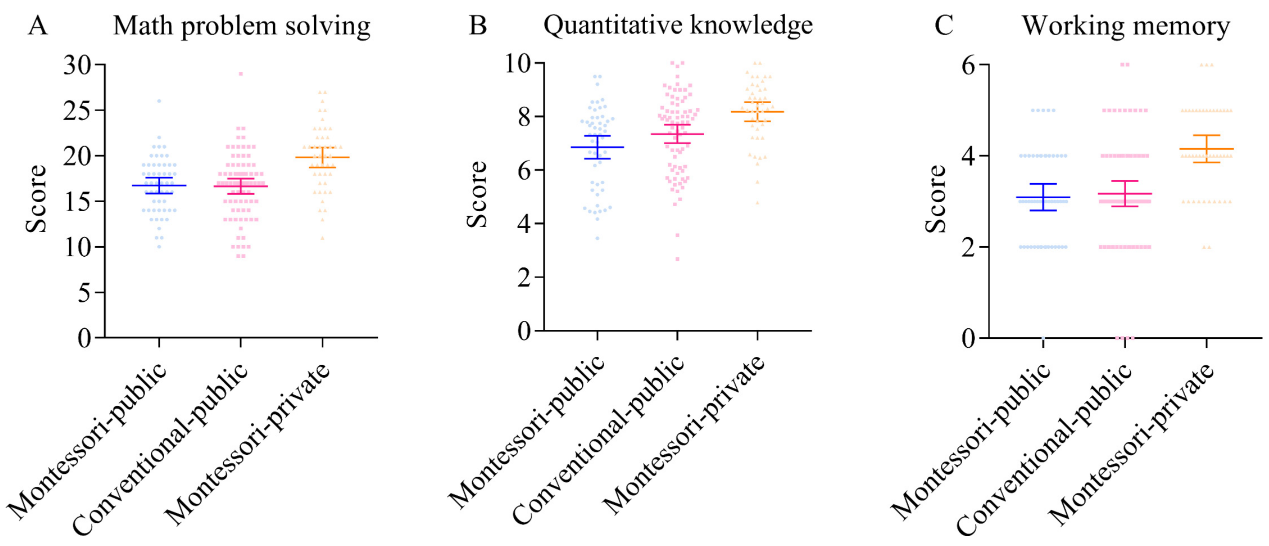

Figure 2. Experiment 1: Distribution of scores in the three measures for which the average score was higher in children from the Montessori-private group than in the other groups. (A) Math problem solving. (B) Quantitative knowledge. (C) Working memory. Each blue, pink, and orange dot represents the score of a child in the Montessori-public, Conventional-public, Montessori-private group (respectively). Error bars represent 95% confidence intervals. [Color figure can be viewed at wileyonlinelibrary.com]

Similarly, in the scale measuring subjective perception of reading competence (see Figure 1B), there was substantial evidence for a lower score in the Conventional-public group than in the Montessori-private group ($BF_{10} = 7.99$) and extreme evidence for a lower score in the Conventional-public group than in the Montessori-public group ($BF_{10} > 100$). There was anecdotal evidence for the lack of difference between the two Montessori groups, ($BF_{01} = 2.70$). Therefore, Bayesian analyses indicated that children from the Montessori-public group perceived being more competent in reading than children from the Conventional-public group, though there was weak evidence that children from the Montessori-public and the Montessori-private groups had the same perception.

In the tests measuring vocabulary, math problem solving, quantitative knowledge, working memory, and self-regulation, there was at least substantial evidence that children from the Montessori-private group had higher scores than children from both the Montessori-public group (all BF_{10}s > 3.83) and the Conventional-public group (all BF_{10}s > 8.14). Within the public preschool, there was also substantial evidence that the two groups did not differ on tests of vocabulary ($BF_{01} = 4.55$), math problem solving ($BF_{01} = 5.26$), and working memory ($BF_{01} = 5$; see Figure 2) as well as anecdotal evidence that the two groups did not differ on tests of quantitative knowledge ($BF_{01} = 1.18$) and self-regulation ($BF_{01} = 1.64$). Thus, Bayesian analyses indicated that Montessori education was not associated with better outcomes than conventional education in these tests. Again, differences between the private and the public groups may be attributable to parenting differences as much as differences in curriculum.

Finally, in the test of social competence, there was substantial evidence that children from the Conventional-public group were more flexible in their response than children from both the Montessori-private group ($BF_{10} = 3.63$) and the Montessori-public group ($BF_{10} = 6.66$). There was also a substantial difference for a lack of difference between the two Montessori groups ($BF_{01} = 4.76$). This result might suggest that children from Montessori classrooms may be less used to finding ways to deal with conflictual situations (e.g., they typically do not ask for materials when another child uses it but rather wait for it to be available). However, it needs to be interpreted with caution because it was not predicted and not observed in frequentist statistics.

Overall, results from both frequentist and Bayesian analyses indicate that Montessori education does not lead to wide-ranging benefits in all measures of academic, cognitive, and social skills. However, the Montessori curriculum clearly appears to promote reading skills in kindergarten-aged children. Because Experiment 1 was a cross-sectional study that focused on children at the end of kindergarten, it remains unclear (a) when a difference in reading performance may emerge between children from Montessori and conventional classrooms during the preschool period and (b) whether any difference between pedagogies may be observed before kindergarten. By tracking the progress of children from preschool entry to the end of kindergarten in the public preschool, Experiment 2 aimed to answer these questions.

Experiment 2

Method

Participants

Participants were a subset of public-school students from Experiment 1 whom we were able to follow from preschool entry to the end of kindergarten (i.e., 3 years of education). Initially, 119 students were tested during their first year of preschool. From the original sample of 119 children (69 in Conventional-public and 50 in Montessori-public), 23 left the school at some point during the study (12 in Conventional-public and 11 in Montessori-public) and 11 did not have parental consent for testing at all time points (7 in Conventional-public and 4 in Montessori-public). We also excluded children who were not fluent in French (2 in Conventional-public and 1 in Montessori-public), changed pedagogy at some point (11 left Conventional-public for Montessori-public), and were related to the staff (1 in Montessori-public). Our final sample was therefore composed of 70 students (33 participants in the Montessori-public group and 37 participants in the Conventional-public group). Sample demographics are detailed in Table S4. Participants from the two groups did not differ with respect to age ($BF_{01} = 3.85$, $t(66) = 0.30$, $p = .77$) and gender ($BF_{01} = 3.32$, $\chi^2(2, N = 70) = 0.09$, $p = .77$). Participants whose parents filled our home environment questionnaire (23 parents from the Conventional-public group and 11 parents from the Montessori-public group) indicated that the subgroups did not differ in household income ($BF_{01} = 2.86$, $t(31) = 0$, $p = 1$), number of siblings ($BF_{01} = 1.96$, $t(32) = 1.01$, $p = .32$), and languages at home ($BF_{01} = 1.85$, $t(32) = 1.09$, $p = .28$).

Preregistration

Measures and analytic strategy were preregistered via the OSF at https://osf.io/pabz3. Those were faithful to the preregistration, except for the additional use of frequentist analyses and the slight changes in some scores (see Tables 1 and 3).

Materials

The tests used at each time point of Experiment 2 are identical to those described for Experiment 1 (see Tables 1–3). The only exceptions are (a) the test measuring planning skills and (b) the scale measuring FAS. We introduced those at a later stage in the study and could only collect longitudinal data on a small subset of participants. Therefore, these two tests were not analyzed in Experiment 2. In addition, only nine subtests of the test measuring quantitative knowledge are analyzed in Experiment 2 (compared to 10 subtests in Experiment 1) because the last subtest was also added at a later stage during the study.

Procedure

Testing procedures and test sessions were the same as in Experiment 1. Data were collected over three time points from two cohorts of children. The first time point was at the beginning of their first year of preschool (i.e., in October 2015 for cohort #1 and in October 2016 for cohort #2). The second time point was halfway during their second year of preschool (i.e., January 2017 for cohort #1 and January 2018 for cohort #2), The third and final time point was at the end of their kindergarten year (third year of preschool; i.e., June 2018 for cohort #1 and June 2019 for cohort #2). At each time point, five to seven experimenters collected the data.

Analyses

As in Experiment 1, longitudinal data in Experiment 2 were analyzed using both frequentist and Bayesian statistics. First, test scores were entered in a series of 2×3 frequentist ANOVAs with the between-subject factor Group (Conventional-public, Montessori-public) and the within-subject factor Time (Year #1, Year #2, Year #3). To account for the number of tests, effects were only considered significant at a Bonferroni-corrected significance threshold of $p = .003$ (i.e., 0.05/18 tests). In case of an interaction between Group and Time, the ANOVA was followed up by independent *t*-tests between groups.

Bayesian statistics were further used to estimate the strength of evidence (i.e., BF) for the alternate hypothesis of a difference in development between groups (H1) versus the null hypothesis of no difference in development between groups (H0) for each test. Test scores were entered into a series of Bayesian ANOVAs with the between-subject factor Group (Conventional-public, Montessori-public, Montessori-private) and the within-subject factor Time (Year #1, Year #2, Year #3). Default priors were used. All analyses were performed with the JAMOVI software (The Jamovi Project, 2019).

Data Availability

The data that support the findings of Experiment 2 are available via the OSF at https://osf.io/4knh2/.

Results

Frequentist Results

Average performance by test, group, and time point is indicated in Table S5. Results from the frequentist 2×3 ANOVAs with the between-subject factor Group and the within-subject factor Time indicated a main effect of Time for the large majority of the tests (see Table S6). This indicated that performance generally improved with age across both groups. There was a main effect of group for the reading test (see Table S7), which was qualified by an interaction between Group and Time (see Table 5). Follow-up *t*-tests revealed that whereas the two groups did not differ from one another at Year #1 ($t(63) = 1.74$, $p = .09$, $d = 0.43$) and Year #2 ($t(63) = 0.96$, $p = .34$, $d = 0.24$), children from the Montessori group outperformed children from the Conventional-public group at Year #3 ($t(63) = 4.32$, $p < .001$, $d = 1.07$; see Figure 3A). In other words, the benefits of Montessori education on the growth of reading skills were not visible before kindergarten. The interaction between Group and Time was significant in none of the other tests, indicating that Montessori education did not impact the growth of any other skills in this longitudinal sample.

Bayesian Results

As shown in Table S6, evidence for an effect of Time was at least substantial on most tests. This indicated that performance improved with age. There was no more than anecdotal evidence for a main effect of Group in all tests except for the social competence test (reference to justice), for which there was substantial evidence (see Table S7). This main effect, however, was entirely driven by a higher score for children from the Conventional-public than the Montessori-public group at Year #2 (see Table S5). More importantly, there was extreme evidence for an interaction between Group and Time for the reading test (see Table 5). Follow-up Bayesian *t*-tests revealed anecdotal evidence for a lack of a difference between the two groups at Year #1 ($BF_{01} = 1.11$), and Year #2 ($BF_{01} = 2.63$). However, there was extreme evidence that children from the Montessori-public group outperformed children from the Conventional-public group at Year #3 ($BF_{10} > 100$; see Figure 3A). Thus, in line with frequentist statistics, Bayesian analyses indicated that the large gap in reading skills between children from Montessori classrooms and children from conventional classrooms only emerged at the end of kindergarten.

The only other test in which there was more than anecdotal evidence for an interaction between Group and Time was the measure of short-term memory (see Table 5). As shown in Figure 3, this effect was mainly driven by a difference between the groups at baseline (i.e., Year #1). Indeed, Bayesian *t*-tests indicated anecdotal evidence that children from the Conventional-public group outperformed children from the Montessori-public group at Year #1 ($BF_{10} = 2.30$), whereas there was anecdotal evidence that the groups had a similar performance at Year #2 ($BF_{01} = 2.94$) and Year #3 ($BF_{01} = 2.32$). Finally, evidence for a lack of interaction between Group and Time was at least substantial for most of the other tests, including tests measuring vocabulary, phonological awareness, pragmatic skills, counting knowledge, sharing, distributive justice, and theory of mind. Therefore, with the notable exception of reading, there was evidence that the growth of academic, cognitive, and social skills was largely similar in children from Montessori classrooms and children from conventional classrooms. Results from both Experiment 1 and Experiment 2 are discussed in the following section.

General Discussion

Overall, our cross-sectional and longitudinal findings largely support the null hypothesis (i.e., no difference between the Montessori and the conventional preschool curriculum) in the majority of competences tested here. For example, we found that children in Montessori-private group outperformed children from the Conventional-public group (who followed the French national curriculum) on measures of reading, math skills, and

Table 5

Experiment 2: Interaction Between Group (Montessori-Public, Conventional-Public) and Time (Year #1, Year #2, Year #3) for Each Test in the Frequentist and Bayesian ANOVAs

Skill	*df*	*F*	*p*	η_p^2	BF_{10}	BF_{01}
Language						
Vocabulary	2, 132	1.55	.22	.02	0.23	**4.35**
Phonological awareness	2, 126	0.01	.99	< .01	0.08	**12.5**
Reading	2, 126	**22.26**	**< .001**	**.26**	**> 100**	< 0.01
Pragmatic skills	2, 124	0.03	.97	< .01	0.05	**20**
Math						
Math problem solving	2, 124	2.97	.06	.05	1.05	0.95
Counting knowledge	2, 126	0.62	.54	.01	0.11	**9.10**
Quantitative knowledge	2, 126	3.73	.03	.06	1.44	0.69
Executive functions						
Short-term memory	2, 126	5.13	.007	.08	**4.75**	0.21
Working memory	2, 126	0.70	.50	.01	0.37	2.70
Self-regulation	2, 130	2.09	.13	.03	0.4	2.50
Social abilities						
Sharing	2, 130	0.14	.87	< .01	0.04	**25**
Distributive justice (overall)	2, 130	0.21	.81	< .01	0.1	**10**
Distributive justice (resources)	2, 130	0.49	.61	.01	0.12	**8.33**
Distributive justice (common good)	2, 130	0.43	.65	.01	0.08	**12.5**
Distributive justice (power)	2, 130	0.17	.84	< .01	0.01	**100**
Social competence (justice)	2, 130	1.60	.21	.02	0.79	1.27
Social competence (flexibility)	2, 130	2.21	.11	.03	0.43	2.32
Theory of mind	2, 128	0.75	.48	.01	0.12	**8.33**

Note. Frequentist statistics: The Bonferroni-corrected significance threshold was $p < .003$. Significant results are in bold. η_p^2s represent effect sizes that can be considered small (.01), medium (.06) or large (.14; Cohen, 1988). Degrees of freedom vary because sample sizes change depending on the measure (some participants were absent or refused to complete certain tests). Bayesian statistics: BFs > 3 are indicated in bold. BF_{10} indicates the strength of the evidence for the alternative (there is an interaction between Group and Time), whereas BF_{01} values indicate the evidence for the null (there is no interaction between Group and Time). ANOVAs = analyses of variance.

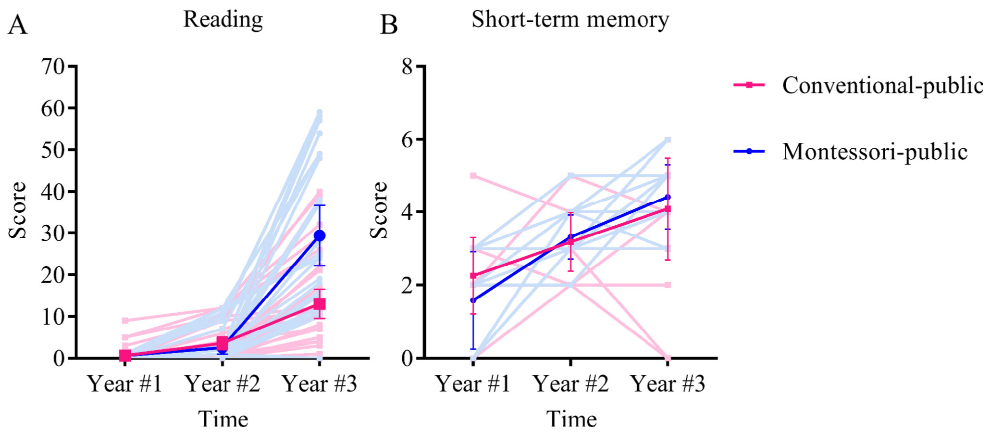

Figure 3. Experiment 2: Longitudinal changes in scores for the two measures in which an interaction between Group (Montessori-public, Conventional-public) and Time (Year #1, Year #2, Year #3) was observed. (A) Reading. (B) Short-term memory. Blue lines and pink lines represent change in the score of a child from the Montessori-public group and the Conventional-public group (respectively). Bold lines represent group average. Error bars represent 95% confidence intervals. [Color figure can be viewed at wileyonlinelibrary.com]

working memory. However, within the public school, the adapted Montessori pedagogy showed a difference only in reading. Because groups were only randomized within the public preschool, differences between the Montessori-private and the Conventional-public group could be explained by a range of variables, including parental involvement and SES. Thus, these findings cannot be taken as evidence for an effect of the Montessori curriculum. Rather, based on Bayesian evidence that children from the Montessori-public and the Conventional-public group are largely similar on measures of math, executive functions, and social skills, our preregistered randomized controlled study provides evidence that the adapted Montessori curriculum in the public preschool leads to gains in math skills, executive functions, or social skills that are similar to a conventional preschool curriculum.

There was, however, one notable exception to the lack of difference between pedagogies in the public preschool. By the end of kindergarten, children who experienced the adapted Montessori curriculum largely outperformed their peers who followed the French national curriculum on the measure of reading skill. This is consistent with studies that might have been methodologically less rigorous than this study but investigated a higher fidelity Montessori curriculum. These studies also observed a positive effect of Montessori education on early reading skills (Lillard, 2012; Lillard & Else-Quest, 2006). What aspects of the Montessori pedagogy may explain such a reading advantage? One possibility is that the acquisition of reading skills may benefit from general aspects of Montessori education, such as multiage classrooms that may promote peer learning (e.g., Bowman-Perrott et al., 2013; Puzio & Colby, 2013) and teaching that is more individualized than in conventional classrooms (e.g., Connor et al., 2009). Although such aspects of a Montessori classroom may indeed affect reading acquisition to some extent, it is difficult to explain why they would not also influence other learning domains (e.g., math learning). Therefore, the reading advantage of Montessori preschool education may be better explained by specific principles and materials underlying the acquisition of literacy skills in the Montessori preschool curriculum. We can see at least three aspects of the Montessori approach that are consistent with research.

First, children from Montessori classrooms start learning the phonetic sounds of letters at around age 3, so that these phonemes can later be blended into words (an approach that has been termed *synthetic phonics;* Castles et al., 2018). There is evidence that systematic phonics instruction is a particularly effective way to teach reading in alphabetic languages (Torgerson, Brooks, Gascoine, & Higgins, 2019). This is because the sounds of language are represented by letters or graphemes in an alphabetic system such as French (a language in which mappings between graphemes and phonemes are even more consistent than in English). Thus, learning to explicitly map sounds onto symbols will help children decode words, thereby accessing their meaning (Castles et al., 2018). This is supported by studies showing that reading comprehension is predicted by decoding skills in French children (including those from low SES background; Gentaz, Sprenger-Charolles, Theurel, & Colé, 2013;

Megherbi, Seigneuric, & Ehrlich, 2006), though other skills (such as listening comprehension and vocabulary) have also been shown to predict individual differences (Gentaz et al., 2013).

Second, an interesting characteristic of the Montessori literacy curriculum is that children learn to write words before they learn to read them. For instance, as soon as children begin to associate graphemes and phonemes, they begin to generate words by themselves. This is typically done first with a movable alphabet and then by writing letters and words on slates (and subsequently notebooks). Not only does this approach emphasize the fact that words can be generated by knowing the sound of individual letters or groups of letters, asking children to produce content may also improve memorization (Bertsch, Pesta, Wiscott, & McDaniel, 2007).

Third, activities in the Montessori literacy curriculum initially center on sensory-motor interactions. For instance, children trace letters on sandpaper with their fingers, mimic actions they read, or manipulate figurines of objects to associate them with their written names. Prior studies suggest that such a haptic (i.e., tactile and kinesthetic) component may promote the development of reading skills. For example, Bara, Gentaz, Colé, and Sprenger-Charolles (2004) designed a training intervention for French preschoolers aimed at teaching letters and letter-sound correspondences (each of seven sessions was dedicated to learning the sound of a letter). Although exploring the letters visually was associated with some improvements in knowledge of the alphabetic principle (measured through pseudoword decoding), these gains were amplified when the visual exploration was associated with a haptic exploration of letters (i.e., children explored the letter with their finger). In a subsequent study, Bara, Gentaz, and Colé (2007) further found that visual and haptic exploration of letters improved letter recognition, phoneme identification, and later on decoding skills in disadvantaged preschoolers. This may be because sensory-motor interactions enrich the sensory input of information into memory (Shams & Seitz, 2008). It is also possible that perceptual and motor simulations contribute to conceptual processing. For example, concrete objects may reactivate motor information about how they have been manipulated (Creem-Regehr & Lee, 2005), which may promote conceptual processing and development (Mounoud, Duscherer, Moy, & Perraudin, 2007). This is broadly supported by studies showing the benefits of multimodal (especially visual and haptic) perception on object recognition in children and adults (Bara et al., 2004; Fredembach, de Boisferon, & Gentaz, 2009).

Although several aspects of the Montessori literacy curriculum may not be unique to Montessori preschools, there is a variability of practices in conventional classrooms, and some aspects—such as reliance on synthetic phonics, haptic modality, and writing—might not be as systematic as in Montessori classrooms. For example, the French national literacy curriculum clearly emphasizes phonemic awareness instruction and a phonics approach to teach children how to read (Ministère de l'Education Nationale & de la Jeunesse et des Sports, 2020). However, teachers from conventional classrooms may not necessarily use synthetic phonics but may also introduce children to grapheme-phoneme correspondences by decomposing whole words into parts (an approach that has been termed *analytic phonics*; Castles et al., 2018). Activities involving learning letters and the alphabetic principle might also not be as prevalent in conventional classrooms as they are in Montessori classrooms (though such activities are emphasized in the French literacy curriculum).

An important finding from this study is that the reading skills of children from Montessori classrooms of the public preschool were comparable to the reading skills of children from the Montessori private preschool. Children following the adapted Montessori curriculum were also aware of their reading competence and reported being as competent as children from the Montessori private preschool. This is noteworthy because there was a clear difference in SES between the public and the private school. France is also one of the OECD countries in which the relation between SES and school achievement (including literacy skills) is the strongest (OECD, 2016). Thus, literacy appears to be one domain where Montessori preschool education may have the potential to reduce the early SES gap.

Although we believe that this study is unique in its context and experimental design, it has limitations. For example, assessors of the children were not blind to which group the children were in. This might have introduced some bias during testing (though this would not explain why reading measures differed whereas other measures did not). The sample size was also constrained by the number of children attending the schools during the study period. Nonetheless, Bayesian analyses revealed substantial to decisive evidence for either H0 or H1 in the vast majority of our measures, indicating that our analyses remained sensitive and that a lack of difference between groups may not necessarily stem from a lack of power.

Finally, we acknowledge that the Montessori curriculum evaluated here was adapted to the context of French public education. For example, daily work periods were shorter than in accredited Montessori schools, teachers were trained in a conventional teacher's college (and lacked accredited Montessori training), and the materials in the public program were less abundant. This may raise issues regarding whether our findings could be replicated in a different context. To help understand how the curriculum evaluated here differs from what could be considered a high-fidelity Montessori curriculum (Lillard, 2012), we developed an openly available scale measuring the fidelity of implementation of the Montessori pedagogy in the public preschool. Remarkably, there was relatively little difference between Montessori classrooms of the public preschool and private Montessori preschools of the Lyon area in terms of characteristics and activities of children (see Table S2). However, there were fewer materials in public than private classrooms (see Appendix S3 for a complete list of the materials present in classrooms of the Montessori-public group), which again suggests that the fidelity of implementation of the Montessori curriculum in the public preschool was lower than in previous studies that might have been methodologically less rigorous but focused on higher fidelity AMI-affiliated schools (Lillard, 2012; Lillard & Else-Quest, 2006). More generally, the lack of advanced materials in some areas (e.g., math) might have also affected children's learning and motivation in the Montessori-public group. Overall, our results call for preregistered studies investigating high-fidelity Montessori education in randomized controlled designs.

In sum, this preregistered, randomized controlled study shows that the advantage of an adapted Montessori curriculum over a conventional preschool program (i.e., the French national curriculum) was limited to literacy. We believe that this finding has two important implications for early childhood education. First, our study shows that a structured preschool curriculum (such as the one used in France) may lead to gains that are largely comparable to an adapted Montessori program across a wide range of academic, cognitive, and social skills. Second, the reading advantage of children from Montessori classrooms indicates that the Montessori literacy curriculum is particularly well suited to the development of early reading skills and can be implemented even by teachers with limited training. Not only are several features of this curriculum supported by cognitive research, but its introduction in preschool does not appear to impact other learning domains. Thus, greater implementation of principles underlying the Montessori literacy curriculum in preschool programs might be beneficial to the acquisition of early reading skills. A specificity of our study is that it was conducted (a) in France and (b) on children from low SES backgrounds. To some extent, our results may thus generalize previous studies that showed a similar reading advantage in preschoolers from variable SES backgrounds in the United States (e.g., Lillard, 2012; Lillard & Else-Quest, 2006). Together with these studies, our findings suggest that the Montessori literacy curriculum may have the potential to reduce early literacy inequalities among children.

References

Ackerman, D. J. (2019). The Montessori preschool landscape in the United States: History, programmatic inputs, availability, and effects. *ETS Research Report Series*, *2019*, 1–20. https://doi.org/10.1002/ets2.12252

Bara, F., Gentaz, E., & Colé, P. (2007). Haptics in learning to read with children from low socio-economic status families. *British Journal of Developmental Psychology*, *25*, 643–663. https://doi.org/10.1348/026151007X186643

Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, *19*, 433–449. https://doi.org/10.1016/j.cogdev.2004.05.003

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, *28*, 168–175. https://doi.org/10.1016/j.evolhumbehav.2006.10.003

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201–210. https://doi.org/10.3758/BF03193441

Billard, C., & Touzin, M. (2012). *Assessment of cognitive functions and learning skills from ages 4 to 11*. Isbergues: Ortho Édition.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, *42*, 39–55. https://doi.org/10.1080/02796015.2013.12087490

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education*, *57*, 190–213. https://doi.org/10.1177/0004944113495500

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, *343*, 1478–1485. https://doi.org/10.1126/science.1248429

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*, 5–51. https://doi.org/10.1177/1529100618772271

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics*, *126*, 1593–1660. https://doi.org/10.1093/qje/qjr041

Chisnall, N., & Maher, M. (2007). Montessori mathematics in early childhood education. *Curriculum Matters*, *3*, 6–28. https://doi.org/10.18296/cm.0086

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. *Child Development*, *80*, 77–100. https://doi.org/10.1111/j.1467-8624.2008.01247.x

Corsi, P. M. (1972). *Memory and the medial temporal region of the brain*. Unpublished doctoral dissertation, McGill University.

Creem-Regehr, S. H., & Lee, J. N. (2005). Neural representations of graspable objects: Are tools special? *Cognitive Brain Research*, *22*, 457–469. https://doi.org/10.1016/j.cogbrainres.2004.10.006

Culclasure, B., Fleming, D., & Sprogis, A. (2018). *An evaluation of Montessori education in South Carolina's Public Schools: 2015–2016 educator survey results summary*. Greenville, SC: The Riley Institute at Furman University.

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

Debs, M., & Brown, K. E. (2017). Students of color and Public Montessori Schools: A review of the literature. *Journal of Montessori Research*, *3*, 1–15. https://doi.org/10.17161/jomr.v3i1.5859

Denervaud, S., Knebel, J.-F., Hagmann, P., & Gentaz, E. (2019). Beyond executive functions, creativity skills benefit academic outcomes: Insights from Montessori education. *PLoS One*, *14*, https://doi.org/10.1371/journal.pone.0225319

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, *333*, 959–964. https://doi.org/10.1126/science.1204529

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, *27*, 109–132. https://doi.org/10.1257/jep.27.2.109

Farah, M. J. (2017). The neuroscience of socioeconomic status: Correlates, causes, and consequences. *Neuron*, *96*, 56–71. https://doi.org/10.1016/j.neuron.2017.08.034

Fredembach, B., de Boisferon, A. H., & Gentaz, E. (2009). Learning of arbitrary association between visual and auditory novel stimuli in adults: The "bond effect" of haptic exploration. *PLoS One*, *4*. https://doi.org/10.1371/journal.pone.0004844

Gentaz, E., Sprenger-Charolles, L., Theurel, A., & Colé, P. (2013). Reading comprehension in a large cohort of French first graders from low socio-economic status families: A 7-month longitudinal study. *PLoS One*, *8*. https://doi.org/10.1371/journal.pone.0078608

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*, 114–128. https://doi.org/10.1016/j.jpubeco.2009.11.001

Huppert, E., Cowell, J. M., Cheng, Y., Contreras-Ibáñez, C., Gomez-Sicard, N., Gonzalez-Gadea, M. L., . . . Decety, J. (2018). The development of children's preferences for equality and equity across 13 individualistic and collectivist cultures. *Developmental Science*, *22*. https://doi.org/10.1111/desc.12729

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*, 801–825. https://doi.org/10.1146/annurev-economics-080213-040845

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press, Clarendon Press.

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review*, *65*, 107–125. https://doi.org/10.1016/j.econedurev.2018.05.001

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867. https://doi.org/10.1037/a0014939

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews*, *15*, 60–68. https://doi.org/10.1002/ddrr.46

Khomsi, A. (2001). *ELO: Assessment of oral language*. Paris: ECPA.

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science*, *4*, 731–739. https://doi.org/10.1111/j.1756-8765.2012.01221.x

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

Laski, E. V., Vasilyeva, M., & Schiffman, J. (2016). Longitudinal comparison of place-value and arithmetic knowledge in montessori and non-montessori students. *Journal of Montessori Research*, *2*, 1–15. https://doi.org/10.17161/jomr.v2i1.5677

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

Lillard, A. S. (2019). Shunned and admired: Montessori, self-determination, and a case for radical school reform.

Educational Psychology Review, *31*, 939–965. https://doi.org/10.1007/s10648-019-09483-3

Lillard, A. S., & Else-Quest, N. (2006). Evaluating Montessori education. *Science*, *313*, 1893–1894. https://doi.org/10.1126/science.1132362

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. *Frontiers in Psychology*, *8*, https://doi.org/10.3389/fpsyg.2017.01783

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

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

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, *76*, 978–988. https://doi.org/10.1111/j.1467-8624.2005.00891.x

Marshall, C. (2017). Montessori education: A review of the evidence base. *NPJ Science of Learning*, *2*. https://doi.org/10.1038/s41539-017-0012-7

McClellan, D. E., & Kinsey, S. (1997, April 3–6). Children's social behavior in relationship to participation in mixed-age or same-age classrooms. Biennial Meeting of the Society for Research in Child Development, Washington, DC.

Megherbi, H., Seigneuric, A., & Ehrlich, M.-F. (2006). Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. *European Journal of Psychology of Education*, *21*, 135–147. https://doi.org/10.1007/BF03173573

Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20. https://doi.org/10.1016/j.cobeha.2016.04.006

Ministère de l'Education Nationale, de la Jeunesse et des Sports. (2015). *Preschool and kindergarten curriculum* (Bulletin officiel special n°2 du 26 Mars 2015). Retrieved from https://www.education.gouv.fr/bo/15/Special2/MENE1504759A.htm

Ministère de l'Education Nationale, de la Jeunesse et des Sports. (2020). *Preparing the acquisition of reading and writing in preschool and kindergarten*. Retrieved from https://cache.media.eduscol.education.fr/file/maternelle/41/4/Guide_phonologie_1172414.pdf

Mix, K. S., Smith, L. B., Stockton, J. D., Cheng, Y.-L., & Barterian, J. A. (2017). Grounding the symbols for place value: Evidence from training and long-term exposure to base-10 models. *Journal of Cognition and Development*, *18*, 129–151. https://doi.org/10.1080/15248372.2016.1180296

Mounoud, P., Duscherer, K., Moy, G., & Perraudin, S. (2007). The influence of action perception on object recognition: A developmental study. *Developmental Science*, *10*, 836–852. https://doi.org/10.1111/j.1467-7687.2007.00624.x

OECD. (2016). *Low-performing students: Why they fall behind and how to help them succeed*. Author. https://doi.org/10.1787/9789264250246-en

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. https://doi.org/10.1037/a0015365

Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*, 51–72. https://doi.org/10.1007/s10648-014-9255-5

Puzio, K., & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*, 339–360. https://doi.org/10.1080/19345747.2013.775683

Rubin, K. H. (1988). *The social problem-solving test-revised*. Waterloo, ON: University of Waterloo.

Rubin, K. H., & Clarke, M. L. (1983). Preschool teachers' ratings of behavioral problems: Observational, sociometric, and social-cognitive correlates. *Journal of Abnormal Child Psychology*, *11*, 273–286. https://doi.org/10.1007/BF00912091

Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating Pre-K programs predict children's learning? *Science*, *341*, 845–846. https://doi.org/10.1126/science.1233517

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*, 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190. https://doi.org/10.1080/15475441.2014.927328

The Jamovi Project. (2019). *Jamovi (version 0.9)*. Retrieved from https://www.jamovi.org

Thomas, M. S. C., & Knowland, V. C. P. (2009). Sensitive periods in brain development: Implications for education policy. *European Psychiatric Review*, *2*, 17–20.

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, *34*, 208–238. https://doi.org/10.1080/02671522.2017.1420816

Valeski, T. N., & Stipek, D. J. (2001). Young children's feelings about school. *Child Development*, *72*, 1198–1213. https://doi.org/10.1111/1467-8624.00342

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian* *evaluation of informative hypotheses* (pp. 181–207). New York, NY: Springer. https://doi.org/10.1007/978-0-387-09612-4_9

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Woodcock, R. W., Mather, N., McGrew, K. S., & Wendling, B. J. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside.

Yen, S.-C., & Ispa, J. M. (2000). Children's temperament and behavior in Montessori and constructivist early childhood programs. *Early Education and Development*, *11*, 171–186. https://doi.org/10.1207/s15566935eed1102_3

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Previously Reported Effects Comparing Montessori and Conventional Preschool Education as a Function of the Domain

Table S1. Experiment 1: Sample Demographics

Table S2. Experiment 1: Percent fidelity of Montessori Implementation in Different Classrooms

Table S3. Experiment 1: Average Performance by Test and Group

Table S4. Experiment 2: Sample Demographics at Year #1

Table S5. Experiment 2: Average Performance by Test, Time and Group

Table S6. Main Effect of Time for Each Test in the Frequentist and Bayesian ANOVAs (Experiment 2)

Table S7. Main Effect of Group for Each Test in the Frequentist and Bayesian ANOVAs (Experiment 2)

Appendix S1. List of Studies Comparing Montessori and Conventional Preschool Education Between 2000 and 2020

Appendix S2. Parental Questionnaire

Appendix S3. Characteristics of Montessori Classrooms in the Public and Private Preschool