



A national randomized controlled trial of the impact of public Montessori preschool at the end of kindergarten

Angeline S. Lillard^{a,1}, David Loeb^b, Juliette Berg^c, Maya Escueta^c, Karen Manship^c, Alison Hauser^c, and Emily D. Daggett^a

Affiliations are included on p. 10.

Edited by Greg Duncan, University of California, Irvine, CA; received March 18, 2025; accepted September 20, 2025

Although seminal studies from the early 1960s suggested quality preschool can have lasting positive effects, agreement is lacking on the efficacy of different preschool models. The Montessori model is longstanding but lacks rigorous impact studies; prior random lottery studies included just one or two schools, among other compromises. Here, we report on end-of-kindergarten (age 5 to 6) impacts from a national study of public Montessori preschool. We compared children offered a Montessori seat via competitive lottery admission processes at one of 24 public Montessori schools at age 3 ($n = 242$) to children not offered a seat ($n = 346$), estimating Montessori impacts with intention-to-treat and complier average causal effect models. Roughly half of the treatment sample still attended Montessori for kindergarten. Although there were no notable impacts at the end of PK3 or PK4, at the end of kindergarten, controlling for baseline scores and demographics, Montessori children had significantly higher reading, short-term memory, theory of mind, and executive function scores. Intention-to-treat effect sizes exceeded a fifth of a SD, considered large in field-based school research [M. A. Kraft, *Educ. Res.* 49, 241–253 (2020)]. This contrasts sharply with the more typical finding, where impacts of preschool are observed immediately following the program but disappear by the end of kindergarten. Further, a cost analysis suggested three years of public Montessori preschool costs less per child than traditional programs, largely due to Montessori having higher child:teacher ratios in PK3 and PK4. Although sensitivity and robustness analyses yielded similar results, important limitations of the study should be noted.

childhood education | school outcomes | randomized controlled trial

Public preschool is considered an effective public investment (1), but consensus on optimal preschool models is lacking (2). Sixty years ago, two intensive preschool programs led to markedly better adult outcomes (3, 4), but societal changes make replication unlikely today because today's "control group" is better-resourced (5). Studies of modern preschool programs often find immediate (end of the four-year-old school year, or PK4) benefits on academic skills like math, vocabulary, and particularly literacy, yet by the end of kindergarten, scores often converge with those of nonattenders, eliminating advantages associated with attending prekindergarten (pre-k) (6–8). Even more concerning, a Tennessee study found that although children randomized to public pre-k programs showed better academic skills at the beginning of kindergarten (9), by third and again in sixth grade children who had attended pre-k did worse academically and socially than children who lost the lottery to attend (10). This has focused discussion on which preschool programs implemented in public schools today are beneficial (5). A particular concern is how to instill academic learning when children naturally learn through play (11). Montessori is an alternative type of preschool program that bridges this dichotomy, offering academic material without whole class instruction, and with several features of play, including free choice, discovery, hands-on materials, and self-evaluation (12). Here, we report on a national study of the impact of public Montessori preschool on a wide range of end-of-kindergarten outcomes.

Montessori Education

Montessori is the longest-running (since 1907) and most common "alternative" pedagogy, estimated to exist in over 16,000 schools around the world (13). Working first with atypically developing and then with low-income children, Dr. Maria Montessori and her collaborators experimented to develop what she called "Scientific Pedagogy" and others called "The Montessori Method" (14). Montessori classrooms group children in

Significance

Although preschool is often considered positive, optimal preschool models are debated. Montessori is a longstanding model that has not been rigorously examined. We followed from age 3 through kindergarten 588 children entered in competitive lotteries at 24 public Montessori schools across the United States. We found that the experimental group, half of which still attended Montessori in kindergarten, had significantly better end-of-kindergarten outcomes for reading, short-term memory, executive function, and social understanding. We also found that three years of public Montessori from ages 3 to 6 cost districts \$13,127 less per child than traditional programs, largely due to higher child:teacher ratios in PK3-PK4. Given the impact and lower cost, Montessori might be considered by districts implementing preschool programs for 3-y-olds.

Author contributions: A.S.L., M.E., and K.M. designed research; A.S.L., M.E., K.M., A.H., and E.D.D. performed research; A.S.L., D.L., J.B., M.E., and K.M. analyzed data; and A.S.L., D.L., J.B., M.E., K.M., A.H., and E.D.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: lillard@virginia.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2506130122/-/DCSupplemental>.

Published October 20, 2025.

specific three-year age spans (e.g., 3 to 6), so older children can serve as role models and help younger ones (14). The teacher provides individual and small group lessons, but learning stems largely from freely-chosen interactions with a curated set of hands-on materials (14). Montessori teacher training conveys theory and a multitude of lessons on how to present the materials, and aims to instill in teachers a profound respect for the developmental process and the interconnectedness of all life. Although Dr. Montessori and her collaborators founded the Association Montessori Internationale (AMI) to carry on her work, “Montessori” is not trademarked so the name does not indicate authenticity (15). The over 600 public Montessori programs in the United States today include varied implementations. In addition, they include many Title I schools, and majority children of color (16). Despite its reach, evidence on Montessori’s efficacy is insufficient (17).

Research on the Outcomes of Montessori Education. In theory, Montessori education should show positive outcomes for students because the features of authentic Montessori just noted align well with the science of learning (18–20). For example, studies show that children learn and develop well when their movement and cognition are aligned, when they can freely choose activities that interest them, and when they can interact with peers (21). A recent meta-analysis in which all 32 analyzed studies had evidence of baseline equivalence found Montessori had better outcomes than traditional schooling on language and literacy (Hedges’ $g = 0.17$), math (Hedges’ $g = 0.22$), executive function (Hedges’ $g = 0.36$), and social skills (Hedges’ $g = 0.23$), among other outcomes (22). However, concerns about selection bias—e.g., parents who choose Montessori education differing in unmeasured ways that are the actual causes of effects—are optimally addressed with randomized trials, ideally where participants are compared with controls who had an equal chance of being admitted to the experimental group (1).

Four randomized trials of public Montessori preschool have been published (23–26), but all were limited to one or two schools, among other compromises (*SI Appendix, section 1*).

Current Study

We took advantage of random lotteries for admission to 24 oversubscribed public Montessori schools across the United States to examine the impact of being offered a seat (intention-to-treat or ITT design) at PK3 (3 y old) on end of kindergarten (5 to 6 y old) outcomes. The study had individual-level random assignment within lottery blocks in which children were clustered for analyses. For a more complete picture, we also examined the effect of attending Montessori for at least the PK3 year among those who complied with their initial lottery treatment assignment, known as a complier average causal effect (CACE) analysis. Eighty-eight percent of those offered a seat in the initial Montessori lottery accepted, and 19.4% of the control group ultimately attended a study Montessori school as well (cross-overs), making the sample compliance rate 68.7% (see *SI Appendix, Fig. S2* for details). Further, 49% of treatment and 11% of control children remained in a study Montessori school through kindergarten (*SI Appendix, Table S43*).

Children were tested at baseline (starting PK3) in the fall of 2021 and each subsequent spring through kindergarten (2024). All 24 public study Montessori schools met basic Montessori criteria (*SI Appendix, section 3A*), but implementation varied widely. In addition to examining a range of academic (reading, vocabulary, and math) and nonacademic (executive function,

memory, theory of mind, social problem-solving, and persistence) outcomes, we conducted an analysis of the cost of implementing Montessori vs. traditional preschool programs in public schools. Our hypothesis based on prior research was that Montessori education would have significantly better outcomes, although because we studied a range of Montessori implementations (meeting minimum standards), those outcomes might be tempered relative to findings from studies using schools with strong implementation of the model. We also expected the cost of implementing Montessori would exceed the cost of traditional preschool, given the cost of materials and high tuition often charged for private programs.

The study was approved by the Institutional Review Board of the American Institutes for Research, Protocol 87236.

Results

The analysis plan for this study was registered prior to examining PK3 and later spring outcome data at the Registry of Efficacy and Effectiveness Studies in Education (REES) (#15183.1v4) and was followed closely; deviations are explained on the REES site and a timeline is in *SI Appendix, section 2*. Data, study instruments, and an updated analysis plan are posted at the Open Science Framework, <https://doi.org/10.17605/OSF.IO/VUC42>. Analytic code is at <https://github.com/david-loeb/montessori-rct> and is also accessible through the Open Science Framework page.

Baseline Equivalence. We first checked the equivalence of the treatment and control groups’ demographics and baseline (fall 2021) assessments through lottery fixed effects regressions of each variable onto the treatment indicator. Table 1 shows that a larger percentage of treatment group families reported incomes greater than \$75,000 (68% of treatment vs. 56% of control, Cox index = 0.324, $P < 0.01$) and caregivers with bachelor’s degrees (81% of treatment vs. 71% of control, Cox index = 0.306, $P < 0.10$). These imbalances likely stem from treatment families in ranked choice lotteries having higher incomes and education levels (*SI Appendix, Table S8*) and higher-income families being more likely to consent if assigned to treatment (*SI Appendix, Table S11*). Evidence suggests that results are robust to these sources of imbalance (*SI Appendix, section 9*), and we control for these variables in our analyses. Treatment group children were also younger by an average of 23 d (Hedges’ $g = -0.198$, $P < 0.10$). Age is factored into the Woodcock-Johnson Z score, acting as a built-in control. To maintain consistency in the covariate set, we do not control for age in our primary models, but results are highly similar when controlling for age (*SI Appendix, Table S31*). We control for all other collected child and family characteristics in our primary analyses.

Table 2 shows baseline equivalence on outcome variables. We used multiple imputation to handle missing outcome data; see *SI Appendix, section 5D* for details. After imputation, effect sizes were all within the What Works Clearinghouse (WWC) standard of <0.25 (27), and no differences were statistically significant. However, when examining only participants with observed baseline and kindergarten data, referred to as “complete cases,” the HTKS test of executive function (Hedges’ $g = -0.298$, $P < 0.05$) fell outside that range. Additionally, the Theory of Mind scale (Hedges’ $g = -0.183$, $P < 0.05$) significantly favored the control group at baseline, though it was within the WWC acceptable range. Because preschool executive function strongly predicts outcomes, we controlled for HTKS in the impact analyses, as well as for each outcome’s baseline score.

Table 1. Baseline sample characteristics and equivalence

	Unadjusted		Cluster-Adjusted			
	Ctrl Mean	Treat Mean	Ctrl Mean	Treat Mean	Diff	Eff. Size
<i>Child Characteristics</i>						
Female	0.500	0.479	0.498	0.482	−0.016 (0.051)	−0.039 (0.102)
Hispanic	0.224	0.269	0.236	0.252	0.016 (0.046)	0.054 (0.119)
Asian	0.064	0.071	0.082	0.046	−0.037 (0.030)	−0.380 (0.223)
Black	0.249	0.168	0.222	0.207	−0.015 (0.045)	−0.055 (0.125)
White	0.482	0.588	0.514	0.543	0.028 (0.059)	0.069 (0.103)
Multiracial	0.126	0.134	0.117	0.146	0.029 (0.024)	0.154 (0.151)
Other Race	0.079	0.038	0.064	0.059	−0.006 (0.031)	−0.059 (0.214)
Primary Lang English	0.842	0.828	0.824	0.854	0.029 (0.028)	0.132 (0.140)
Age	3.426 (0.310)	3.423 (0.323)	3.451	3.388	−0.063+ (0.035)	−0.198 (0.084)
<i>Family Characteristics</i>						
Fam Income 75k+	0.627	0.579	0.555	0.680	0.125** (0.042)	0.324 (0.106)
Caregiver Has B.A.	0.744	0.760	0.713	0.805	0.092+ (0.052)	0.306 (0.122)
Caregiver Married	0.744	0.783	0.767	0.751	−0.016 (0.049)	−0.054 (0.119)
Household Size	3.839 (1.141)	4.038 (1.153)	4.005	3.800	−0.205 (0.126)	−0.179 (0.084)

** $P < 0.01$, * $P < 0.05$, + $P < 0.1$. $N = 588$. Some variables have a small amount (< 2%) of missing data. See *SI Appendix, Table S17* for missing rates by variable. Unadjusted means are raw sample means. Numbers in parentheses are SD for raw means of continuous variables (age and household size) and SE for cluster-adjusted mean differences and effect sizes. Effect sizes are Hedges' g for continuous variables (all child variables except age, and all family variables except household size) and the Cox index for binary variables. Cluster-adjusted results were estimated with lottery fixed effects regressions of each variable onto the ITT treatment indicator. Control group means are the weighted means of the fixed effects, analogous to model intercepts. SE are clustered at the study school level.

Effects on Child Outcomes. The study's primary impact models were ITT analyses with multiply imputed data, and we also present results from CACE and complete case analyses. All models were estimated with lottery fixed effects regressions. There was only one statistically significant difference in outcomes at the end of PK3 (2022): Montessori children were more apt to choose the difficult puzzle in the unvalidated persistence task (*SI Appendix, Table S2*), a finding that reversed in PK4 (2023) when using complete case data (*SI Appendix, Table S5*).

At kindergarten, however, a pattern of differences emerged mainly favoring the experimental group. Table 3 shows the multiple imputation ITT and CACE model results, and Fig. 1 plots the ITT effect sizes across all four waves with 95% CI. At kindergarten end, children scored statistically significantly higher on Woodcock-Johnson (WJ) Letter-Word Identification, Head-Toes-Knees-Shoulders (HTKS), Forward Digit Span, and Theory of Mind in the ITT models, $P < 0.05$, with Hedges' g effect sizes ranging from 0.222 to 0.297. WJ Applied Problems and HTKS are near significance in the CACE models, $P < 0.10$. Patterns of statistical significance are otherwise similar in the CACE models, in which effect sizes among the five variables with some degree of statistical significance range from 0.383 (WJ Applied Problems) to 0.600 (Forward Digit). These CACE effect sizes are 77 to 123% larger than the ITT effect sizes. WJ

Picture Vocabulary and Backward Digit have positive estimated effects, but they are smaller than the aforementioned outcomes and not statistically significant. Social Problem-Solving and the unvalidated persistence task (Puzzle Choice) have negative but nonsignificant estimated effects.

Table 4 shows the complete case results. Patterns of statistical significance are the same as the multiple imputation models, other than WJ Applied Problems and HTKS being near-significant in the ITT model and Forward Digit being near-significant in the CACE model, $P < 0.10$. Effect size magnitudes are generally similar and slightly larger in the complete case models, though the WJ Applied Problems and Puzzle Choice effect sizes are slightly smaller, as is the HTKS ITT estimate.

Moderator analyses (*SI Appendix, Tables S46–S51*) find exploratory evidence suggesting that Montessori effects were stronger among lower-income children and boys. These analyses also indicate that in comparison to White children, effects were stronger among Asian children and weaker among Black children. Effects remained positive among all subgroups, however. We caution that these results are exploratory due to low statistical power.

External and Internal Validity. Regarding external validity, in *SI Appendix, section 8*, we report data suggesting the schools in this study are not high-achieving among public Montessori

Table 2. Baseline differences in outcome measures before and after multiple imputation

	Complete Case				Multiple Imputation			
	Ctrl Mean	Treat Mean	Diff	Eff. Size	Ctrl Mean	Treat Mean	Diff	Eff. Size
WJ Letter Word	0.477	0.556	0.079 (0.136)	0.071 (0.116)	0.443	0.530	0.087 (0.138)	0.082 (0.102)
WJ Applied Prob	0.524	0.680	0.156 (0.119)	0.156 (0.116)	0.392	0.561	0.169 (0.116)	0.170 (0.100)
WJ Pic Vocab	0.601	0.617	0.016 (0.112)	0.019 (0.116)	0.553	0.542	-0.011 (0.118)	-0.012 (0.106)
HTKS	1.021	0.667	-0.355* (0.154)	-0.298 (0.118)	0.900	0.699	-0.201 (0.123)	-0.176 (0.100)
Forward Digit	4.220	4.111	-0.109 (0.394)	-0.050 (0.117)	3.509	3.599	0.090 (0.270)	0.036 (0.098)
Backward Digit	0.316	0.225	-0.091 (0.074)	-0.120 (0.138)	0.246	0.165	-0.081 (0.097)	-0.126 (0.139)
Theory of Mind	1.130	0.956	-0.174* (0.066)	-0.183 (0.117)	1.052	0.931	-0.121 (0.089)	-0.130 (0.108)
Social Prob-Solve	0.361	0.234	-0.128 (0.152)	-0.206 (0.119)	0.338	0.234	-0.104 (0.096)	-0.167 (0.103)
Puzzle Choice	0.176	0.137	-0.038 (0.055)	-0.177 (0.205)	0.182	0.133	-0.048 (0.045)	-0.223 (0.163)

** $P < 0.01$, * $P < 0.05$, + $P < 0.1$. $N = 588$ for multiple imputation results. N s vary by outcome for complete case results; see [SI Appendix, Table S16](#). Numbers in parentheses are SE. WJ = Woodcock-Johnson Z-scores, the age-normed version of the W score. HTKS = Head-Toes-Knees-Shoulders. Effect sizes are the Cox index for the dichotomous Puzzle Choice and Hedges' g for all other outcomes. Results are from lottery fixed effects regressions of each baseline outcome onto the ITT treatment indicator. Control group means are the weighted means of the fixed effects, analogous to model intercepts. SE are clustered at the study school level.

schools. On the other hand, school admission lottery applicants are more likely to be White and less likely to qualify for free and reduced-price lunch (29) than the general public, and applicants to Montessori schools might differ in these and other unmeasured ways that would make these results not generalizable to people who do not enter Montessori school admission lotteries. Further, the COVID-19 context might compromise the study's applica-

bility across time. For example, learning from older peers is important in Montessori classrooms, yet the pandemic made most older peers miss the prior year of classroom experiences when this study began, surely influencing their behavior as peer models.

Regarding internal validity, we conducted a series of analyses to assess the sensitivity of our results to three potential sources of bias, shown in [SI Appendix, section 9](#). We find that large positive

Table 3. ITT and CACE estimates at end of kindergarten, multiple imputation

	ITT				CACE			
	Ctrl Mean	Treat Mean	Diff	Eff. Size	Ctrl Mean	Treat Mean	Diff	Eff. Size
WJ Letter Word	1.029	1.396	0.367* (0.152)	0.297 (0.123)	0.832	1.510	0.678* (0.313)	0.548 (0.195)
WJ Applied Prob	0.669	0.823	0.155 (0.103)	0.172 (0.118)	0.567	0.908	0.341+ (0.180)	0.383 (0.141)
WJ Pic Vocab	0.188	0.263	0.075 (0.072)	0.107 (0.102)	0.130	0.312	0.183 (0.138)	0.265 (0.144)
HTKS	41.696	45.358	3.662* (1.696)	0.250 (0.112)	40.348	46.843	6.495+ (3.392)	0.443 (0.175)
Forward Digit	7.340	7.896	0.557* (0.228)	0.293 (0.109)	6.985	8.138	1.153* (0.462)	0.600 (0.137)
Backward Digit	2.711	2.940	0.229 (0.209)	0.160 (0.129)	2.616	3.031	0.415 (0.357)	0.289 (0.175)
Theory of Mind	2.960	3.182	0.222* (0.106)	0.222 (0.110)	2.849	3.264	0.414* (0.206)	0.422 (0.140)
Social Prob-Solve	1.486	1.416	-0.070 (0.202)	-0.055 (0.127)	1.574	1.376	-0.197 (0.420)	-0.158 (0.246)
Puzzle Choice	0.557	0.488	-0.068 (0.072)	-0.167 (0.117)	0.613	0.449	-0.164 (0.148)	-0.406 (0.293)

** $P < 0.01$, * $P < 0.05$, + $P < 0.1$. $N = 588$. Numbers in parentheses are SE. WJ = Woodcock-Johnson Z-scores, the age-normed version of the W score. HTKS = Head-Toes-Knees-Shoulders. Effect sizes are the Cox index for the dichotomous Puzzle Choice and Hedges' g for all other outcomes. Results are from lottery fixed effects regressions with full covariates; see [SI Appendix, section 5](#) for details. Control group means are the weighted means of the fixed effects, analogous to model intercepts; covariates are grand mean-centered. SEs are clustered at the study school level. The average effect size for the Cognition domain, which includes HTKS, Forward & Backward Digit, and ToM, is 0.231, $P = 0.003$. Benjamini-Hochberg-adjusted P -values are 0.053 for HTKS, Forward Digit & ToM (28).

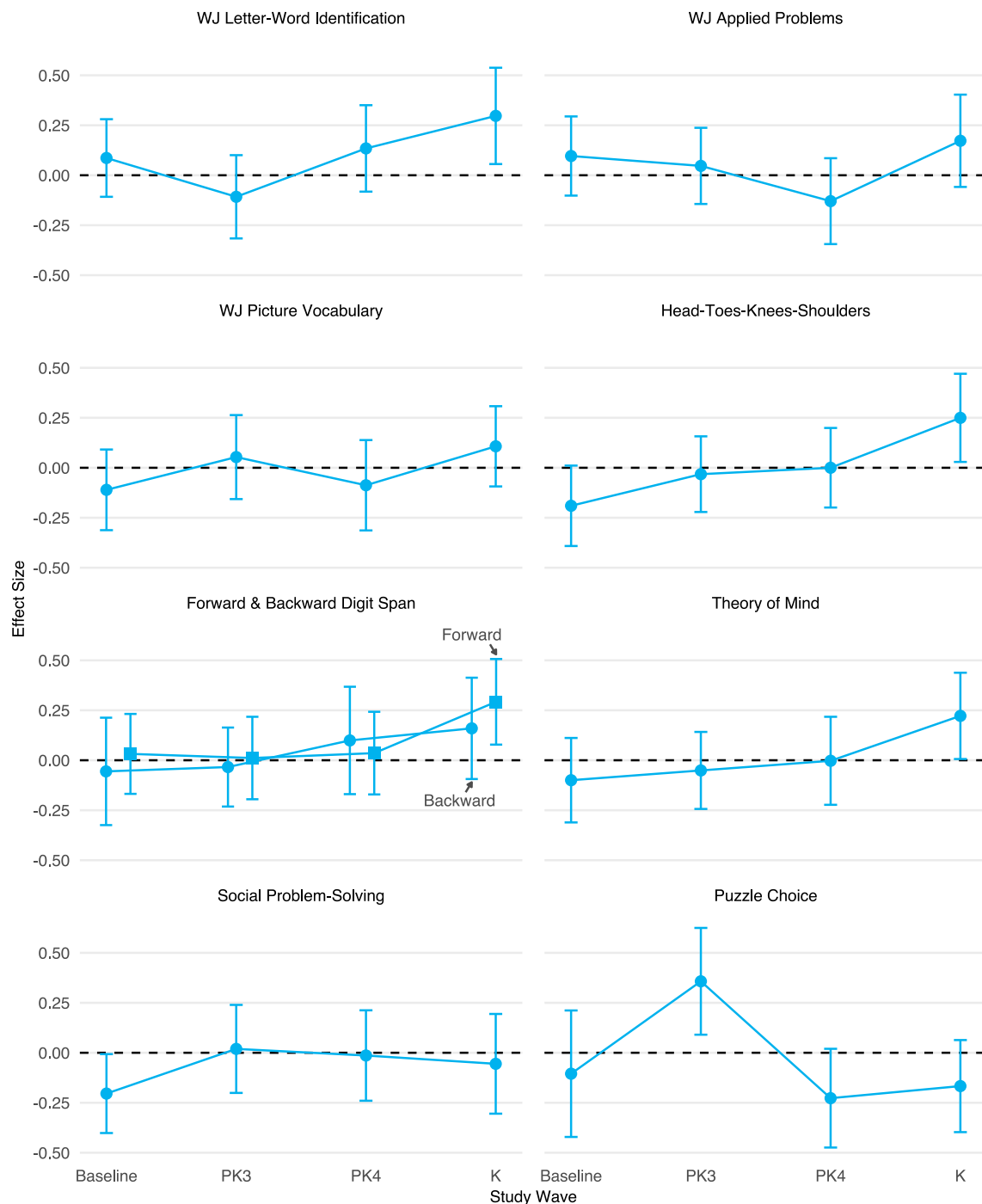


Fig. 1. ITT effect sizes across study waves. Each panel shows mean differences between the treatment and control groups across the four study waves for the outcome noted in the panel header, expressed in effect size units. Effect sizes greater than zero indicate the Montessori group scored higher than the control group. Effect sizes were computed using results of lottery fixed effects regressions of the outcomes onto the ITT treatment indicator and full covariates. Covariates were included in baseline effect size estimates, in contrast to effect sizes shown in Table 2 which were estimated without covariates, so the numbers differ slightly. Effect sizes are the Cox index for the dichotomous Puzzle Choice and Hedges' g for all other outcomes. Error bars are the effect size 95% CI. Results shown are from the multiple imputation models.

bias would be required from differential consent and differential missingness to produce the observed treatment effects if the true effects are zero. Also, results are robust to excluding the 13 schools where ranked choice lotteries create potentially unequal probabilities of assignment to treatment across applicants: effect sizes are similar and remain uniformly large among the outcomes for which statistically significant effects are found in the primary models (*SI Appendix, Table S9*).

We conducted additional analyses including alternative approaches to handling missing data (*SI Appendix, section 9C*), direct balancing of baseline covariates through propensity score weighting and matching (*SI Appendix, section 9D*), alternative covariate specifications including no covariates (*SI Appendix, section 10*), and an examination of whether school stability may explain some of the observed treatment effects (*SI Appendix, section 12B*). Results are highly robust to these specifications.

Table 4. ITT and CACE estimates at end of kindergarten, complete cases

	ITT				CACE			
	Ctrl Mean	Treat Mean	Diff	Eff. Size	Ctrl Mean	Treat Mean	Diff	Eff. Size
WJ Letter Word	1.000	1.366	0.365* (0.143)	0.300 (0.121)	0.744	1.524	0.780* (0.324)	0.640 (0.124)
WJ Applied Prob	0.698	0.846	0.148+ (0.074)	0.167 (0.121)	0.596	0.910	0.314+ (0.176)	0.355 (0.122)
WJ Pic Vocab	0.174	0.273	0.099 (0.068)	0.147 (0.121)	0.104	0.316	0.212 (0.149)	0.315 (0.121)
HTKS	42.452	45.573	3.120+ (1.663)	0.228 (0.121)	40.270	46.914	6.644+ (3.557)	0.485 (0.122)
Forward Digit	7.408	7.987	0.579* (0.276)	0.307 (0.122)	6.998	8.245	1.248+ (0.607)	0.661 (0.124)
Backward Digit	2.924	3.237	0.312 (0.288)	0.238 (0.141)	2.648	3.404	0.756 (0.679)	0.577 (0.144)
Theory of Mind	2.946	3.201	0.254* (0.101)	0.266 (0.122)	2.768	3.309	0.541* (0.231)	0.567 (0.123)
Social Prob-Solve	1.469	1.258	-0.210 (0.179)	-0.173 (0.123)	1.625	1.156	-0.469 (0.406)	-0.387 (0.124)
Puzzle Choice	0.561	0.521	-0.040 (0.094)	-0.097 (0.153)	0.589	0.504	-0.084 (0.200)	-0.207 (0.154)

** $P < 0.01$, * $P < 0.05$, + $P < 0.1$. Sample sizes vary by outcome due to skipped questions; see *SI Appendix, Table S16*. Numbers in parentheses are SE. WJ = Woodcock-Johnson Z-scores, the age-normed version of the W score. HTKS = Head-Toes-Knees-Shoulders. Effect sizes are the Cox index for the dichotomous Puzzle Choice and Hedges' g for all other outcomes. Results are from lottery fixed effects regressions with full covariates; see *SI Appendix, section 5* for details. Control group means are the weighted means of the fixed effects, analogous to model intercepts; covariates are grand mean-centered. SE are clustered at the study school level.

Cost of Montessori. A cost analysis estimated the overall and per-child costs of Montessori classrooms and traditional preschool classrooms for each year from PK3 through kindergarten, amortizing material and Montessori teacher training costs over their expected lifespans. Based on these calculations, we found that for these three preschool years combined, implementing Montessori programs in public schools costs an estimated \$13,127/child less than implementing a traditional program (see *Materials and Methods* and *SI Appendix, section 15* and *Tables S53–S55* for more details about these calculations). The cost difference is primarily attributable to the intentionally higher child-to-adult ratios in Montessori classrooms at PK3 and PK4, which more than compensated for amortized higher teacher training and material costs and relatively lower ratios in kindergarten.

Discussion

Although many agree that preschool programs benefit society, experts disagree about optimal programs, especially regarding play-based versus academic emphases (5)—a dichotomy bridged by Montessori (12). This nationwide randomized controlled trial (RCT) of public Montessori preschool, which included a range of Montessori programs and used multiple analytic approaches, consistently found positive impacts of Montessori preschool on a wide array of end-of-kindergarten outcomes. Furthermore, a cost analysis suggested that three years of Montessori from ages 3 to 6 costs districts less per child than traditional programs covering that full age range.

Across all models, a standardized test of early reading yielded significantly higher scores for the treatment group at the end of kindergarten. This finding has replicated across all four recent RCTs examining the impact of public Montessori preschool (*SI Appendix, section 1*): compared to children in other programs, children who entered Montessori at PK3 read at a higher level not in PK3 or PK4, but at the end of kindergarten. The effect size was almost a third of a SD. The methods

Montessori uses to teach reading—e.g., beginning with writing, emphasizing phonics—align with the science of reading, perhaps explaining this finding. American children's reading performance is declining (30), and because early reading skill predicts later reading skill, this finding may be of particular significance (31, 32).

Children offered a seat in a public Montessori preschool at age three also performed significantly better on a Forward Digit Span task, showing more developed short-term memory (33). Memory span is a component of working memory, which also involves manipulating information held in mind (34). On our Backward Digit Span task directly assessing working memory, no significant results were obtained, but results still favored Montessori. Working memory is an aspect of executive function, which is more fully tapped by the HTKS task on which Montessori children performed significantly better. Better executive function performance among Montessori students is often observed (22, 35), and higher executive function predicts more positive outcomes in school (36, 37) and beyond (38–40).

Montessori students also performed better on theory of mind, which predicts social competence (41, 42). Children offered a seat at Montessori also did somewhat better on a math test at the end of kindergarten. A math advantage associated with Montessori preschool is seen less consistently than a reading advantage (43), but was observed in the meta-analysis (22). Performance on vocabulary, social problem-solving, and an exploratory task intended to tap persistence were not significantly different across groups, although the latter two tests had somewhat worse results for the experimental group, which warrants consideration.

Regarding social problem-solving, children randomly assigned to Montessori performed slightly but not significantly worse. One possible explanation is that social-emotional learning programs, implemented in over 80% of U.S. elementary schools (44), teach specific strategies to get a peer to share a coveted object—the task used in our study. Such programs may be less emphasized in

Montessori classrooms, where social lessons are taught in response to evident classroom needs. Alternatively, it could also be that using just one social-problem story, as we did, is not sufficiently sensitive to bring out additional strategies.

The unvalidated persistence task (Puzzle Choice) was the only task on which there were significant findings in earlier study waves: Montessori PK3 students chose to keep working on a difficult puzzle, whereas at older ages control children made this choice more, with a moderate but nonsignificant kindergarten effect size (Cox index = -0.167). Something led Montessori children to become less persistent at trying to solve our difficult (actually impossible) puzzles. This is not easy to explain given Montessori's philosophical alignment with encouraging mastery orientation, which can manifest as persistence in the face of difficulty; further research will need to explore this.

Together, the social problem-solving and persistence task findings raise the caution that public Montessori preschool might promote some less optimal social outcomes while also clearly promoting other positive social and academic outcomes. Specifically, in this study, outcomes related to academics, cognition, and executive function more consistently favored Montessori, whereas, except for theory of mind, those related to social outcomes slightly favored control. This concern may be mitigated by the facts that one social test (persistence on the puzzle task) was unvalidated, and another was truncated relative to other studies. Existing meta-analyses favored Montessori for all these types of outcomes, but also included older children, private Montessori schools, and studies with ample potential for bias (22, 35).

Our main analyses were conservative ITT analyses, testing the effect of being offered a seat in the initial lottery at one of the 24 study Montessori schools. Given that the ITT analyses indicate that just being offered a seat in a Montessori school had an impact, analyses that account for treatment assignment compliance would obtain larger effects. Indeed, the CACE effect sizes are between three quarters to double the effect sizes obtained in the ITT analyses.

Yet the ITT effect sizes obtained in this study are still meaningful: in field studies using standardized tests, effects less than 0.05 are small, 0.05 to 0.20 are medium, and effects greater than 0.20 are large (45). Translating the ITT multiple imputation effects to percentile growth (46), we estimate that a child in a traditional program who performed at the median in reading or memory would have performed at the 62nd percentile had they matched with a Montessori study school; a child who performed at the median on math would be at the 57th percentile in math had they matched at Montessori; a child with median performance on executive function (HTKS) would be at the 60th percentile, and a child at the median for theory of mind would be at the 59th percentile had they matched. Looking only at compliers, the CACE effect sizes are about twice as large, so, for example, a child at the 50th percentile in reading would be at the 71st percentile had they won a seat and attended Montessori. Multiplied across hundreds of thousands of children, such impacts would be very meaningful.

The fact that 48.8% ($n = 118$) of the experimental group still attended a study Montessori school in kindergarten may be important. Continuity or the alignment of curriculum across levels predicted better outcomes in a study of Boston's public prekindergarten program (47). However, program continuity does not account for our results (*SI Appendix, section 12*). We were also not powered to detect possible effects of attending Montessori for PK3 through kindergarten, vs. just for PK3 or PK3 and PK4.

The pattern of findings here varies from that often reported in the preschool literature, where preschool is associated with better end-of-program outcomes, but participants' scores converge with those of nonattenders in the kindergarten year (6). Our pattern is consistent with other longitudinal Montessori RCT studies, however (24, 25). We discuss four possible explanations for the pattern.

First, some might speculate that the pattern derives largely from child:teacher ratios in Montessori vs. control classrooms. Drawing on live observations of study classrooms at PK3 and teacher surveys across all three study years, we found that among classrooms for which we had data to calculate ratios, child:teacher ratios at PK3 and PK4 were larger in Montessori than control classrooms, whereas at kindergarten, when we saw significant differences, they were smaller (*SI Appendix, Tables S53 and S54*). It is commonly believed that smaller ratios are better for learning and development. Importantly, evidence on the benefit of smaller ratios is equivocal. One meta-analysis found that the supposition was true for cognitive and achievement outcomes among 3- to 5-year-olds, yet only when ratios were very small (7.5:1 and less); ratios extending from 7.5:1 to 15:1 showed no differences (48). In our study, at all three ages, ratios were at or under this threshold, and were far below it in control classrooms at PK3 and PK4; yet Montessori children did not fare worse than controls in those years. A second meta-analysis on ratios included kindergarten children and found no clear effects of ratio (49). A third meta-analysis included children from kindergarten through grade 12, and found the effect of ratio on reading was .11; ratio had a small negative effect on math (50). Thus, even in conventional settings, within the bounds allowed by state licensing, smaller and larger ratios may be less important than is often assumed.

No study has examined the impact of ratio on child outcomes in Montessori settings, where ratios might operate differently. Based on her observations, Dr. Montessori advocated a minimum of 25 children per class (with one teacher), and suggested better outcomes are achieved with even larger numbers (51). Ratio could operate differently in Montessori classrooms because lessons are virtually all individual, with classrooms set up so children learn by repeating on their own what they learned in lessons, as well as by observing and getting lessons from peers. Larger class sizes mean more peer examples, tutors, and tutees are available. Because the longitudinal RCTs of public Montessori preschool all find better results at the kindergarten year and not earlier, perhaps the experience of tutoring younger children is a strong driver of our pattern of results. Many studies show that the act of teaching is highly effective for learning (52). Thus, although some might think smaller ratios at kindergarten are driving these results, there is not strong evidence of such an effect in traditional classrooms and it may not make sense in Montessori, although it has not been studied.

A second potential driver of our pattern is the unique experiences children have in their third year in a Montessori classroom, when they are the oldest children. Montessori teachers often claim that they see children blossom in their third year in the classroom, consolidating and taking on leadership roles reflecting the knowledge they gained in the classroom in the first two years. Although more research is needed, two studies that directly addressed whether completing a third year in the same Montessori classroom is particularly beneficial suggest that it is (53, 54). This would suggest that the subset of children who remained in Montessori classrooms for kindergarten carried the results; we were underpowered to determine whether this was the case.

The third potential reason for our pattern begins with the fact that virtually all study children benefitted from pre-k of some sort, rendering the groups equal at younger ages. In kindergarten, however, teachers focus on less “school-ready” students (55); hence our sample likely got less teacher attention in the kindergarten year, which could lead to less advancement in school models where the teacher drives learning. However, in a Montessori setting, where learning can be largely self-guided, children with pre-k experience are (at least in theory) not held back by less teacher attention or by teachers aiming whole class lessons at less prepared children. Children drive their own learning in Montessori through interaction with peers and the materials, the most advanced of which are designed to challenge even 6-y-olds. As noted, Montessori prefers higher student:teacher ratios, in part because it theoretically encourages more peer interaction (15). By these lights, were conventional teachers more focused on children who entered the school year ahead (due to preschool experience), control children’s outcomes might converge with those of the Montessori sample.

A fourth possible reason for the pattern we observed is that it might simply take time for the Montessori model to have an effect on children. Developmental processes can be slow and occur “underground.” Perhaps features of Montessori like free choice and embodied learning take time to fruit.

Finally, the pattern could stem from a combination of these factors. About half of the children in the experimental group benefitted from a third year of self-initiated learning, and from the experience of being the eldest in the classroom and tutoring others; others in the experimental group (the half that had left Montessori by kindergarten) may have experienced delayed manifestation of some benefits derived from the Montessori preschool model (like the phonetic approach to reading). On the flip side, many control children may have experienced convergence in kindergarten due to teachers focusing less on children with pre-k experience, which might be especially disadvantageous in conventional school models that rely more on whole class teaching.

Limitations

RCTs are a gold standard in research. School studies often use random lotteries to approximate this standard, and validity is a key concern (29). There are several points where external and internal validity might be compromised in this study; we review some of the most important limitations here.

The Study Montessori Schools Were Not Randomly Selected. All 24 Montessori schools were oversubscribed at PK3; Montessori schools that are not oversubscribed at PK3 might have weaker outcomes. However, a check on this involving 11 of our schools found they did not perform better on 3rd grade standardized tests on average than 184 other public Montessori schools (*SI Appendix, section 8A*). Another concern regarding generalization to other Montessori schools concerns implementation. We applied minimal criteria for Montessori schools to enter the study, but here as elsewhere, Montessori implementation varies (56), and implementation predicts outcomes (57, 58). (Quality of course also matters for control programs.) In sum, our results may not apply to all public Montessori schools.

Montessori Lottery Applicants Are Not Representative of the General Population. The study population includes only children who apply to Montessori PK3 programs. Findings might

not apply to those who do not apply to Montessori or school lotteries in general.

Roughly Half of the Lotteries Were Districtwide Lotteries.

Thirteen study schools used districtwide lotteries in which parents rank choices, and chances of getting in depend in part on the order of rankings and priority factors like geographic proximity, using what are called deferred acceptance algorithms. Yet a sensitivity check limited to the lotteries that were entirely random had similar impact estimates, suggesting that the inclusion of ranked choice lotteries, while improving power, did not substantially bias the impact estimates (*SI Appendix, section 9A*).

Low and Differential Consent Rates. Just 20% of lottery entrants consented to participate in the research. External validity is thus compromised if those who consented differ meaningfully from those who do not consent. The racial composition of the consented sample was similar to that of the study Montessori schools, but our sample was less likely to qualify for free or reduced-price lunch (*SI Appendix, Table S7*). Because lower-income children in our sample were more impacted by Montessori, it is possible that our results might be stronger in the full population of applicants. However, our sample may differ in other important ways from applicants who did not consent, so we cannot know how our results would generalize to the broader population of applicants.

Additionally, 31% of the treatment group consented compared to just 17% of the control group. The differential consent may bias impact estimates if those who consented in the treatment group systematically differ from control group consenters. We explore this possibility with sensitivity analyses in *SI Appendix, section 9B*. While treatment group applicants of all income levels are more likely to consent than control applicants, the treatment-control consent gap is larger among higher- than lower-income families. We find that the differential consent bias would need to be large to spuriously produce the observed treatment effects, with a strong relationship to the probability of consenting and the outcomes, independent of the baseline covariates which include baseline outcome differences. Still, potential bias from differential consent is an important limitation of the study.

Missing Data. We were unable to assess a substantial portion of the full sample at each study wave, with 32% of treatment and 42% of control participants missing kindergarten assessments (*SI Appendix, Tables S15–S18*). If participants with missing data differ from those with observed data, and particularly if those in the treatment group differ systematically from those in the control group, our effect estimates may be biased. We use multiple imputation (MI) to help limit this potential bias, but MI requires the “missing at random” assumption, which holds that all variables related to missingness are included in the model, to eliminate bias. *SI Appendix, section 9C* reports four sensitivity analyses that suggest results are robust to potential bias from missingness. Nonetheless, we cannot rule out the possibility that differential missingness biases our estimates, and this remains a study limitation.

Another limitation related to missing data is reduced power. *SI Appendix, Table S52* shows that the minimum detectable effect sizes (MDES) for the complete case models range from 0.26 to 0.36 depending on the outcome. Because the multiple imputation models use partial data from participants with incomplete data, they have lower MDES than the complete case models (59). Still, the multiple imputation MDES are likely

higher than the 0.20 targeted by the initial study design. See *SI Appendix, section 14* for further discussion.

Results Are End-of-Kindergarten. We do not know whether the study children will continue to do well in their later school years. In addition to two Montessori meta-analyses that included studies of older children (22, 35), a very tightly controlled year-over-year study of public Montessori in South Carolina saw significantly better outcomes in later years (53), but following the specific children in this study is necessary to determine whether impacts occur over longer time scales.

Conclusions

Although our findings must be taken with caution because of study limitations, additional analyses suggest that the results are robust to most of these limitations. The findings are especially notable because this is a field study of a business-as-usual intervention that was not designed by the researchers: Montessori preschool in public U.S. schools. A typical finding in the intervention literature is positive impacts when researchers implement their own interventions, and no impact in field trials (60). Furthermore, studies comparing interventions in normal school practice with business-as-usual programs rarely yield positive results; a recent analysis of 77 such RCTs showed that 91% yielded null results (61).

The cost analysis suggested Montessori was less expensive due to using larger child-teacher ratios at younger ages. The CLASS measure (62) often used to evaluate preschool quality indicated that in Montessori preschool classrooms in this study, the higher the child-adult ratio in their classroom when study children were 3 (up to 13:1 in this study), the higher their CLASS quality scores (63). This and our impact results suggest larger ratios do not compromise classroom quality or learning in Montessori settings.

The positive results obtained here at least warrant further study of Montessori education, and suggest that expanding access to Montessori PK3 through kindergarten programs may be a cost-effective way to sustain early learning gains at least through the end of kindergarten, while also positively impacting skills like executive function and social understanding.

Materials and Methods

SI Appendix, sections 4 and 5 provide more detail on the study methods.

Study Sample. Families of 2,919 children who had applied to competitive random lotteries for a 2021-22 PK3 seat at 25 public Montessori schools across the United States were contacted about participating in the study. One school (with 12 viable lottery participants) withdrew prior to testing in the fall of 2021, leaving 24 schools. Lottery data for three schools were obtained in spring 2022, preventing baseline testing. Among the included 24 schools, 588 children were consented, 242 who won a seat and 346 who did not. *SI Appendix, Figs. S1–S3* detail school and sample recruitment. The overall consent rate was 20.9%. Parents provided demographic information with the consent forms. Children averaged 3.4 y of age when data collection began in the fall of 2021, and were followed through kindergarten (ages 5 to 6). Table 1 shows they were racially and economically diverse, and about half were boys. The consented sample's overall racial demographics were similar to the mean demographics across study schools, except fewer study families qualified for free/reduced lunch (*SI Appendix, Table S7*).

The vast majority ($n = 213$ of 242) of children randomized to the treatment group attended a study Montessori school at PK3 (88.0%); at PK4

and kindergarten these numbers were 65.7% and 48.8%, respectively. Sixty-seven of the 346 children in the control sample (19.4%) attended one of the study Montessori schools in the PK3 year (cross-overs). Nearly half, 44.8%, of the control group attended other public programs (e.g., district public schools) in PK3. The remainder attended private preschool programs (13.6%) or stayed home (6.1%) at PK3; for 56 children (16.2%) PK3 enrollment was unknown. See *SI Appendix, section 11* for more details on counterfactual school settings.

We were unable to assess a substantial share of children in each wave, including 32% of treatment and 42% of control participants in kindergarten, due both to school absences and families consenting to participate but then not responding to later communication attempts. Our primary models use multiple imputation to address this missing data, and we also present results from models using only complete cases, meaning the subset of children who had no missing data on model variables. For details on multiple imputation, see *SI Appendix, section 5D*, and for sensitivity analyses related to missing data, see *SI Appendix, section 9C*.

Outcome Measures. Children were individually tested by professional data collectors in the fall of 2021 and for the three subsequent springs (2022, 2023, 2024) on reading, vocabulary, math, executive function (HTKS), memory (forward and backward digit span), persistence, social problem-solving, and social understanding (Theory of Mind Scale); details on the measures are in *SI Appendix, section 4E*. Testers were blind to this being a study of Montessori education.

Cost Analysis. We estimated total costs following the ingredients approach (64) with a resource cost model framework that models the structure and ingredients of services provided under a particular program. We first identified all resources (or ingredients) used to implement each classroom environment, such as the required teachers, their training and certification, and classroom materials. For each identified resource (personnel and nonpersonnel), information about its quantity (e.g., number of staff per classroom, number of hours spent on a certification program) was gathered, and information about quality (e.g., staff credentials, caliber of educational materials) was used to identify a market price. We recovered quantity and price information from publicly available sources (e.g., the cost of Montessori materials and teacher training), interviews with district administrators in two states (e.g., information about space used for both Montessori and traditional classrooms and teacher salary schedules), and teacher respondents ($N = 256$) to surveys sent to all teachers in treatment and control classrooms enrolling a study child each spring during the study (e.g., number of children enrolled and teacher training received). Quantities and prices for each ingredient were combined to estimate costs. The upfront costs of teacher training and Montessori materials were amortized over their expected lifespan of 25 y. Total costs for Montessori and traditional programs were divided by the average number of children in each type of classroom at the PK3, PK4, and kindergarten years, then summed to yield the cost difference of Montessori and traditional programming over three years of public preschool. More detail is provided in *SI Appendix, section 15*.

Data, Materials, and Software Availability. Anonymized (.csv file) data have been deposited in Open Science Framework (<https://doi.org/10.17605/OSF.IO/VUC42>) (65).

ACKNOWLEDGMENTS. Data collection through the PK4 year and the cost analysis was supported by Grant R305A180181 from the Institute of Education Sciences, U.S. Department of Education, to the American Institutes for Research (AIR). The kindergarten year data collection and analyses were supported by Grant 23-10339 from Arnold Ventures to the University of Virginia, and a Brady Education grant to AIR. David Loeb was supported by Grant R305B200035 from the Institute of Education Sciences to the University of Pennsylvania and Emily Daggett received some support from the Institute of Education Sciences, U.S. Department of Education Training Program Grant R305B200005 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Assistance with Research: Ann-Marie Faria (original co-PI, AIR),

Mark Lachowicz (original Data Analyst, AIR), Deeza Mae Smith (deputy project manager, AIR), Alejandra Martin (classroom observations task lead, AIR), Isabel Sorenson (Recruitment, AIR), Linda Choi (Recruitment, AIR), Dong Hoon Lee (Data Analyst, AIR), Adrian Duran (Data Analyst, AIR), and at UVA research assistants Abigail Krissinger, Samuel Diener, Justin Gregory, Paige Crawford, Leena Heilizer, Lucy Cooper, Hajer Anjum, and Allyson Snyder, graduate researchers Lee LeBoeuf and Abha Basargekar, and researcher Corey Borgman all assisted this project. School Readiness Consulting managed all child and classroom data collection, including Laura Hawkinson, BreAnna Davis Tribble, Dori Morman, Eugenia McRae, Aisha Pittman Fields, Sherylls Valladares Kahn,

Maya Manning, and their team. We also gratefully acknowledge the support of Technical Working Group members Christina Weiland, Stephanie Jones, Virginia McHugh, David Ayer, and the late Jacqueline Cossentino, and especially to the districts, schools, teachers, families, and children who made this work possible.

Author affiliations: ^aDepartment of Psychology, University of Virginia, Charlottesville, VA 22904; ^bDivision of Policy, Organizations, Leadership, and Systems, Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104; and ^cAmerican Institutes for Research, Chapel Hill, NC 27517

- H. Yoshikawa *et al.*, "Investing in our future: The evidence base on preschool education" (Tech. Rep., Society for Research in Child Development, 2013).
- D. Bassok, M. Engel, Early childhood education at scale: Lessons from research for policy and practice. *AERA Open* **5**, 232858419828690 (2019).
- A. Baulos, J. L. Garcia, J. J. Heckman, Perry Preschool at 50: What lessons should be drawn and which criticisms ignored? <https://doi.org/10.3386/w32972> (Accessed 3 October 2025).
- F. A. Campbell, C. T. Ramey, E. Pungello, J. Sparling, S. Miller-Johnson, Early childhood education: Young adult outcomes from the Abecedarian Project. *Appl. Dev. Sci.* **6**, 42–57 (2002).
- M. Burchinal *et al.*, Unsettled science on longer-run effects of early education. *Science* **384**, 506–508 (2024).
- D. H. Bailey, G. J. Duncan, F. Cunha, B. R. Foorman, D. S. Yeager, Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychol. Sci. Public Interest* **21**, 55–97 (2020).
- A. Ansari, R. C. Pianta, J. V. Whittaker, V. E. Vitiello, E. A. Ruzek, Persistence and convergence: The end of kindergarten outcomes of pre-K graduates and their nonattending peers. *Dev. Psychol.* **56**, 2027–2039 (2020).
- D. A. Phillips *et al.*, "Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects" (Tech. Rep., Brookings, 2017).
- M. W. Lipsey, D. C. Farran, K. Durkin, Effects of the Tennessee Prekindergarten program on children's achievement and behavior through third grade. *Early Child. Res.* **Q.** **45**, 155–176 (2018).
- K. Durkin, M. W. Lipsey, D. C. Farran, S. E. Wiesen, Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Dev. Psychol.* **58**, 470–484 (2022).
- C. Weiland, M. McCormick, S. Mattera, M. Maier, P. Morris, Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open* **4**, 232858418757735 (2018).
- A. S. Lillard, Montessori as an alternative early childhood education. *Early Child Dev. Care* **191**, 1196–1206 (2021).
- M. Debs *et al.*, Global diffusion of Montessori schools: A report from the 2022 Global Montessori Census. *J. Montessori Res.* **8**, 1–15 (2022).
- M. Montessori, *The Discovery of the Child* (Montessori-Pierson Publishing, Amsterdam, The Netherlands, 2017), vol. 2.
- A. S. Lillard, V. McHugh, Authentic Montessori: The Dottressa's view at the end of her life part I: The environment. *J. Montessori Res.* **5**, 1–18 (2019).
- M. C. Debs, Racial and economic diversity in U.S. public Montessori schools. *J. Montessori Res.* **2**, 15–34 (2016).
- D. J. Ackerman, The Montessori preschool landscape in the United States: History, programmatic inputs, availability, and effects. *ETS Res. Rep. Ser.* **2019**, 1–20 (2019).
- L. Darling-Hammond, L. Flook, C. Cook-Harvey, B. Barron, D. Osher, Implications for educational practice of the science of learning and development. *Appl. Dev. Sci.* **24**, 97–140 (2020).
- P. Cantor, D. Osher, J. Berg, L. Steyer, T. Rose, Malleability, plasticity, and individuality: How children learn and develop in context. *Appl. Dev. Sci.* **23**, 307–337 (2019).
- B. T. Culclasure, C. J. Daoust, S. M. Cote, S. Zoll, Designing a logic model to inform Montessori research. *J. Montessori Res.* **5**, 35–49 (2019).
- A. S. Lillard, *Montessori: The Science Behind the Genius* (Oxford University Press, Oxford, UK/New York, NY, ed. 3, 2017).
- J. J. Randolph *et al.*, Montessori education's impact on academic and nonacademic outcomes: A systematic review. *Campbell Syst. Rev.* **19**, e1330 (2023).
- A. S. Lillard, N. Else-Quest, Evaluating Montessori education. *Science* **313**, 1893–1894 (2006).
- A. S. Lillard *et al.*, Montessori preschool elevates and equalizes child outcomes: A longitudinal study. *Front. Psychol.* **8**, 1783 (2017).
- P. Courtier *et al.*, Effects of Montessori education on the academic, cognitive, and social development of disadvantaged preschoolers: A randomized controlled study in the French public-school system. *Child Dev.* **92**, 2069–2088 (2021).
- L. B. Miller, J. L. Dyer, H. Stevenson, S. H. White, Four preschool programs: Their dimensions and effects. *Monogr. Soc. for Res. Child Dev.* **40**, 1–170 (1975).
- W. W. Clearinghouse, "What Works Clearinghouse procedures and standards handbook, version 5.0" (Tech. Rep., U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE), 2022).
- Y. Benjamini, R. Heller, D. Yekutieli, Selective inference in complex research. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **367**, 4255–4271 (2009).
- C. Weiland *et al.*, The effects of enrolling in oversubscribed prekindergarten programs through third grade. *Child Dev.* **91**, 1401–1422 (2020).
- S. Schwartz, Reading scores fall to new low on NAEP, fueled by declines for struggling students. *Education Week*, (29 January 2025).
- A. E. Cunningham, K. E. Stanovich, Early reading acquisition and its relation to reading experience and ability 10 years later. *Dev. Psychol.* **33**, 934–945 (1997).
- R. L. Sparks, J. Patton, A. Murdoch, Early reading success and its relationship to reading achievement and reading volume: Replication of '10 years later'. *Read. Writing* **27**, 189–211 (2014).
- A. Baddeley, Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* **63**, 1–29 (2012).
- S. E. Gathercole, S. J. Pickering, B. Ambridge, H. Wearing, The structure of working memory from 4 to 15 years of age. *Dev. Psychol.* **40**, 177–190 (2004).
- A. Demangeon, S. Claudel-Valentin, A. Aubry, Y. Tazouti, A meta-analysis of the effects of Montessori education on five fields of development and learning in preschool and school-age children. *Contemp. Educ. Psychol.* **73**, 1–22 (2023).
- R. Bull, G. Scerif, Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Dev. Neuropsychol.* **19**, 273–293 (2001).
- C. Blair, R. P. Razza, Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* **78**, 647–663 (2007).
- A. E. Koepf *et al.*, Attention and behavior problems in childhood predict adult financial status, health, and criminal activity: A conceptual replication and extension of Moffitt *et al.* (2011) using cohorts from the United States and the United Kingdom. *Dev. Psychol.* **59**, 1389–1406 (2023).
- T. Moffitt, R. Poulton, A. Caspi, Lifelong impact of early self-control. *Am. Sci.* **101**, 352–359 (2013).
- D. A. Robson, M. S. Allen, S. J. Howard, Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychol. Bull.* **146**, 324–354 (2020).
- K. Imuta, J. D. Henry, V. Slaughter, B. Selcuk, T. Ruffman, Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Dev. Psychol.* **52**, 1192–1205 (2016).
- V. Slaughter, K. Imuta, C. C. Peterson, J. D. Henry, Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Dev.* **86**, 1159–1174 (2015).
- A. S. Lillard, *The Montessori Difference: Evaluating the Evidence* (Oxford University Press, in press).
- A. Skoog-Hoffman *et al.*, "Social and emotional learning in U.S. schools: Findings from Case's nationwide policy scan and the American Teacher Panel and American School Leader Panel surveys" (Tech. Rep., RAND Corporation, 2024).
- M. A. Kraft, Interpreting effect sizes of education interventions. *Educ. Res.* **49**, 241–253 (2020).
- M. D. Baird, J. F. Pane, Translating standardized effects of education programs into more interpretable metrics. *Educ. Res.* **48**, 217–228 (2019).
- M. P. McCormick *et al.*, Instructional alignment is associated with PreK persistence: Evidence from the Boston Public Schools. *Early Child. Res. Q.* **67**, 89–100 (2024).
- J. B. Bowne, K. A. Magnuson, H. S. Schindler, G. J. Duncan, H. Yoshikawa, A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educ. Eval. Policy Analysis* **39**, 407–428 (2017).
- M. Perlman *et al.*, Child-staff ratios in early childhood education and care settings and child outcomes: A systematic review and meta-analysis. *PLoS ONE* **12**, e0170256 (2017).
- T. Filges, C. S. Sonne-Schmidt, B. C. V. Nielsen, Small class sizes for improving student achievement in primary and secondary schools: A systematic review. *Campbell Syst. Rev.* **14**, 1–107 (2018).
- M. Montessori, *The Child, Society and the World: Unpublished Speeches and Writings* (CLIO, Oxford, UK, 1989).
- A. L. Brown, M. J. Kane, Preschool children can learn to transfer: Learning to learn and learning from example. *Cogn. Psychol.* **20**, 493–523 (1988).
- D. J. Fleming, B. Culclasure, Exploring public Montessori education: Equity and achievement in South Carolina. *J. Res. Child. Educ.* **38**, 459–484 (2024).
- A. S. Lillard, R. H. Jiang, X. Tong, Perfect timing: sensitive periods for Montessori education and long-term wellbeing. *Front. Dev. Psychol.* **3**, 1546451 (2025).
- M. Little, Exploring the theory of preschool skill convergence through a national survey of kindergarten teachers. *Early Educ. Dev.* **36**, 625–639 (2024).
- A. K. Murray, C. Daoust, "Fidelity issues in Montessori research" in *The Bloomsbury Handbook of Montessori Education*, A. K. Murray, E. T. Ahlquist, M. K. McKenna, M. Debs, Eds. (Bloomsbury Academic, London, UK, 2023), pp. 199–208.
- A. S. Lillard, Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs. *J. Sch. Psychol.* **50**, 379–401 (2012).
- A. S. Lillard, M. J. Heise, An intervention study: Removing supplemented materials from Montessori classrooms associated with better child outcomes. *J. Montessori Res.* **2**, 16–26 (2016).
- R. Zha, O. Harel, Power calculation in multiply imputed data. *Stat. Pap.* **62**, 533–559 (2021).
- M. W. Lipsey, *Design Sensitivity: Statistical Power for Experimental Research* (Sage Publications Inc., Thousand Oaks, CA, 1990).
- R. T. Jacob, F. Doolittle, J. Kemple, M. A. Somers, A framework for learning from null results. *Educ. Res.* **48**, 580–589 (2019).
- R. C. Pianta, K. M. LaParo, B. K. Hamre, *Classroom Assessment Scoring System™: Manual K-3* (Paul H. Brookes Publishing Co., Baltimore, MD, 2008).
- A. S. Lillard *et al.*, When bigger looks better: CLASS results in public Montessori preschool classrooms. *Early Child. Res. Q.* **70**, 199–210 (2025).
- H. M. Levin, P. J. McEwan, C. Belfield, A. B. Bowden, R. Shand, *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis* (SAGE Publications, Inc., 2018).
- A. S. Lillard *et al.*, Dataset for a national randomized controlled trial of the impact of public Montessori preschool on children's kindergarten outcomes. Open Science Framework. <https://osf.io/vuc42/>. Deposited 28 August 2025.