



Supporting Online Material for

Evaluating Montessori Education

Angeline Lillard* and Nicole Else-Quest

*To whom correspondence should be addressed. E-mail: lillard@virginia.edu

Published 29 September, *Science* **313**, 1893 (2006)

DOI: 10.1126/science.1132362

This PDF file includes:

Materials and Methods
SOM Text
Tables S1 to S5
References

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/313/5795/1893/DC1)

Data files as zipped archives: Data file for 5-year-olds
Data file for 12-year-olds

Supplementary On-Line Materials

A. Lillard & N. Else-Quest

Contents

- i. Montessori Implementation
- ii. Methods
- iii. Reliability
- iv. Analyses
- v. Results
- vi. Scores Relative to Matriculation Age
- vii. Other Research Using Traditional Montessori Implementations
- viii. Supplementary Online References
- ix. Data

i. Montessori Implementation

The criteria for AMI/USA association include the following (see

<http://www.montessori-ami.org/> for more information):

- a teacher in each classroom having completed an AMI training course for the level being taught (normally 9 months of lectures interspersed with closely supervised teaching)
- a visit by a trained AMI consultant every 3 years

- a complete set of approved Montessori materials in each classroom
- a class size of 28 to 35 children balanced across the appropriate ages for the level (3 to 6, 6 to 9, and so on)
- an uninterrupted 3-hour work period each morning
- no more than one assistant in a 3- to 6-year-old class

ii. Methods

a. The Lottery. The lottery to get into this Montessori school at age 3 is conducted each February by the district, following a well-advertised 3-week application period. Milwaukee families submit applications naming their school of choice. All the children in the present study listed the Montessori school as their first choice. School principals notify the school district of the number of open slots in their school. A computer assigns each child a rank and admits children randomly in order of rank to fill the slots. Children whose rank is lower than the number of slots are placed on a wait list in order. The Montessori school adheres strictly to children's rankings with this wait list with the exception of admitting siblings of children who have already been admitted by the random lottery. Typically the school leaves 5 – 8 slots open (beyond the number designated as open for the lottery) for incoming siblings each year. Because the older siblings were themselves admitted by random selection, other characteristics of the families with siblings would still be expected to be roughly the same as that of control families, assuming that Montessori school applicants change little from one year to the next. Very few children are admitted to the Montessori school after age 3, and only if they have attended another Montessori school.

b. Initial Recruitment. Letters from the principal of the Montessori school and the Milwaukee Public School District accompanied the initial recruitment letters to all families, to legitimize the study. (Even control families were familiar with the Montessori school principal, as they had typically contacted him years earlier to inquire about admission off the waiting list.) Three mailings were sent to the control families (in September, November, and January), with each subsequent mailing targeting families who had not yet responded. By the third mailing, only 3 new participants were recruited. Two mailings were sent to the Montessori children (September and November).

c. Participation Rates. The participation rates are higher for the Montessori than for the control group. Early in the recruitment phase, the Montessori parents were not offered the financial incentive as it was thought to be unnecessary, and eight parents sent forms back explicitly refusing to participate. The school principal advised us that some had called and expressed that there was no benefit to them. To address this, we offered \$100 to the Montessori families as well. In addition, because the numbers of available Montessori children were small enough that very high participation rates at the Montessori school were required in order to have an adequate sample size, the principal also contacted families of all 12-year-olds to remind them of the study, and teachers of 5-year-olds mentioned the study in all of their fall parent-teacher conferences. Following these steps, participation rates from the Montessori increased to required levels. It was not possible to use direct contact with the control group because until parents returned the permission forms, we did not know what schools the children attended, and had no way to contact them except by letters (which we were not certain were received or read). Only 2 control families explicitly declined to participate, and one of those stated that the

reason was a recent change of schools. (This was not a reason to exclude a child from our vantage point, but we did not contact the family again.)

Table S1. Ns In Each Sample at Each Stage of the Recruitment Process.

| 5-year-olds | <i>Available (lost/won lottery)</i> | <i>Address in MPS database</i> | <i>Returned Letter</i> | <i>Consented</i> | <i>Child Partic- ipated</i> |
|--------------|---|--|----------------------------|------------------|-------------------------------------|
| Control | 112 | 90 | 32 | 31 | 25* |
| Montessori | 54 | 54 | 42 | 39 | 30† |
| 12-year-olds | | | | | |
| Control | 99 | 86 | 33 | 32 | 28‡ |
| Montessori | 36 | 36 | 30 | 30 | 29§ |

* 5 had attended other Montessori schools and 1 was home-schooled

† Thirty was the target *n* for all samples.

‡ 2 had attended other Montessori schools, 1 was severely learning disabled, and 1 refused to participate.

§ One child was severely learning disabled.

d. Samples. Although quite a good control sample for a field study in education, the control group is not ideal. The control families who chose to participate in this evaluation might be different in particular ways from the Montessori families who chose to participate. There are several potential sources of difference. For one, the recruitment

methods were somewhat different across samples by virtue of the fact that one group was all at one, known school.

Second, different types of control families might have been willing to participate. Perhaps only control parents with high-achieving children, or only ones who were less watchful of their child (thus did not mind their being tested by a stranger), participated. The Montessori families had “community support” in that many children at their school were being tested.

Another potential problem stems from selective attrition from the Montessori pool. Attrition from the control pool was achieved only by moving out of the area or attending a Montessori or being home schooled, but Montessori students became ineligible when they moved to other schools. Mitigating this concern, an earlier study at a different inner-city public Milwaukee Montessori school showed that attrition from that Montessori was due mainly to families moving out of the area (*SI*), which would impact the control sample equally. However, if it were the case that children who responded poorly to Montessori left the school, that would be problematic, especially by age 12 since they had more years for poor outcomes to drive schooling decisions.

Another potential problem is that all but one of the Montessori children started school at age 3, whereas 13 of the control children did not attend school until age 4. However, as is shown after the main results (See Table S5), the means of control children who matriculated before age 4 were much more like the means of the whole control sample than they were like the means of the Montessori children.

Although the control group is not ideal, the problems are part of natural experiment research using a retrospective lottery loser design. Given the desirability of doing school

outcome evaluations, and that the design has many positive aspects, the control group used here is a good step forward.

e. Schools. Table S2 shows what types of schools were attended by the control children. Six of the 28 12-year-olds were at public schools described as being for “Gifted and Talented” children.

Table S2. Ns at Each Type of Control School at Each Age.

| | Inner-city Public | Suburban Public | Charter | Private Voucher | Private |
|--------------|----------------------|--------------------|---------|--------------------|---------|
| 5-year-olds | 18 | 2 | 1 | 3 | 1 |
| 12-year-olds | 22 | 2 | 2 | 2 | 0 |

f. Table S3. Ages, Age Ranges, and Gender Composition of Samples.

| | | Mean Age | Age Range | Gender |
|--------------|------------|----------|---------------------|------------|
| 5-year-olds | Montessori | 70 m | 64 – 77 m | 15 F; 15 M |
| | Control | 70 m | 64 – 77 m | 10 F; 15 M |
| 12-year-olds | Montessori | 12 y 1 m | 11 y 5 m – 12 y 7 m | 17 F; 12 M |
| | Control | 12 y 0 m | 11 y 5 m - 12 y 6 m | 10 F; 18 M |

g. Procedures. Children were tested in quiet locations at their schools by one of four experimenters, each of whom tested at least 5 children in each group. Two experimenters coded playground behavior.

5-year-olds. Children were administered the False Belief test first, followed by 2 WJ scales (Letter-Word and Word Attack), then the Social Problem Solving task. As a

break they drew a picture for 5 minutes, then received the Dimensional Change Card Sort test of executive function and 2 additional WJ scales (Understanding Directions and Applied Problems). On the next visit, usually 2 days later, they answered 12 questions concerning their feelings about school, received three additional WJ scales (Picture Vocabulary, Spatial Relations, and Concept Formation), and participated in a standard Delay of Gratification test of executive function. The Playground Observation occurred on a separate day. One of two (or two, for reliability samples) coders was present on the playground during recess and coded the child's predominant activity during each 1-minute time block. Recess coding lasted for an average of 12 minutes at both types of schools (with a range of 7 to 20 minutes at both types of school).

12-year-olds. After a questionnaire regarding their feelings about school, 12-year-olds received 3 WJ scales (Letter-Word, Word Attack, and Understanding Directions) followed by the Narrative Composition. Next they read and responded to the six Social Skills Stories, followed by the other 4 WJ scales. Two control 12-year-olds refused to do one or more of the tasks; one more refused to participate at all.

iii. Reliability

All subjective scoring was submitted to commonly-used interrater reliability procedures, with a second blind coder coding at least 20% of the sample. The one place where this was not possible was the playground observation; the second coder had no knowledge of Montessori and no particular interest in the study, but by necessity did know when he was coding at the Montessori school.

Table S4. Reliability of Subjective Measures Across Two Raters.

| Age. | Measure | Reliability |
|------|---|-------------|
| 5 | Social Problem Solving: Appeals to Fairness | $r = .98$ |
| 5 | Playground: Positive Peer Play | $r = .98$ |
| 5 | Playground: Ambiguous Rough & Tumble | $r = .99$ |
| 12 | Narrative Creativity | $r = .73$ |
| 12 | Narrative Sentence Sophistication | $r = .70$ |

iv. Analyses

Most of the data were analyzed using 2-tailed t -tests. Effect sizes are increasingly viewed as more important than p values in psychology research, in part because they are insensitive to sample size. Large samples can yield impressive p values even when mean differences are actually quite small. The widely-used Cohen's d (2) gives the proportion of the pooled standard deviation by which two means are different, and is reported here for comparisons where the primary statistical tests were significant at the $p < .05$ level. Effect sizes of .2 to .4 are considered small but meaningful in social sciences research, .4 to .6 is considered medium, and higher than .6 is considered a large effect size.

v. Results

Table S5. Table of Significant Results.

| Test | Montessori M and <i>SD</i> | Control M and <i>SD</i> | Statistic | <i>p</i> value | Cohen's <i>d</i> |
|--|-------------------------------|----------------------------|--|----------------|---------------------|
| <i>5-Year-Olds</i> | | | | | |
| False Belief | .80 .41 | .52 .51 | Binomial (Montessori > chance) | .025 | .61 |
| Letter-Word | 13.3 7.00 | 10.00 7.83 | Mann-Whitney <i>U</i> <i>Z</i> = 2.24 | .025 | .44 |
| Word Attack | 7.30 4.26 | 5.00 3.19 | <i>t</i> (53) = 2.23 | .030 | .63 |
| Applied Problems | 19.00 3.11 | 17.00 4.19 | <i>t</i> (53) = 2.03 | .047 | .55 |
| Dimensional Card Sort | 22.70 1.58 | 21.00 4.06 | <i>t</i> (53) = 2.11 | .039 | .61 |
| Social Problem Solving: Fairness | 3.90 3.14 | 1.60 2.08 | <i>t</i> (53) = 3.13 | .002 | .89 |
| Positive Peer Play | 13.98 5.27 | 10.61 6.30 | <i>t</i> (53) = 2.16 | .035 | .58 |
| Ambiguous Rough/Tumble | 0.36 0.97 | 2.40 3.83 | <i>t</i> (53) = - 2.81 | .015 | .72 |

| <i>12-Year-Olds</i> | Montessori M and <i>SD</i> | Control M and <i>SD</i> | Statistic | <i>p</i> value | Cohen's <i>d</i> |
|---|-------------------------------|----------------------------|----------------|----------------|---------------------|
| Creativity of Narrative | 2.72 <i>1.07</i> | 1.96 <i>1.09</i> | $t(54) = 2.64$ | .011 | .71 |
| Sentence Sophistication: Narrative | 2.62 <i>0.96</i> | 2.06 <i>0.94</i> | $t(54) = 2.22$ | .031 | .59 |
| Social Problem Solving: Positive, Direct Strategy | 2.89 <i>1.47</i> | 1.81 <i>1.50</i> | $t(53) = 2.69$ | .009 | .73 |
| Positive School Feelings | 4.72 <i>5.69</i> | 1.21 <i>7.32</i> | $t(55) = 2.03$ | .048 | .54 |

vi. Scores Relative to Matriculation Age

Table S5. Means of Whole Control Sample and of Young Matriculators in Control and Montessori Groups.

| | Control Whole Sample | Control Matriculated before 4 (<i>n</i> = 10) | Montessori Matriculated before 4 (<i>n</i> = 29) |
|------------------|-------------------------|--|---|
| False Belief | .52 | .40 | .79 |
| Letter-Word | 10.00 | 10.20 | 13.52 |
| Word Attack | 5.00 | 5.40 | 7.45 |
| Applied Problems | 17.00 | 16.80 | 19.00 |

| | | | |
|---------------------------------|-------|-------|-------|
| DCCS | 21.00 | 20.80 | 22.72 |
| Appeals to Justice | 1.60 | 2.30 | 3.86 |
| Positive Peer Play | 10.61 | 9.24 | 13.77 |
| Ambiguous Rough and Tumble Play | 2.40 | 2.34 | 0.37 |

vii. Other Research Using Traditional Montessori Implementations

Other studies of traditional Montessori programs have also recently showed positive effects. Students attending Montessori middle schools, relative to matched controls, were significantly more likely to report 1) feeling energized and engaged while doing schoolwork; 2) spending more time doing schoolwork and less time socializing and watching media during school; 3) that their friends and classmates are one and the same people; and 4) that their classrooms are orderly, that their teachers are supportive, and that they feel emotionally safe at school (S3, S4). In another study, children who attended other public Montessoris in Milwaukee from ages 3 to 11 scored significantly higher several years later on standardized tests of math and science, relative to their high school classmates (matched on ethnicity, gender, and free lunch status) (S5). Infants who were randomly assigned to a Montessori Early Head Start program had superior language and cognitive skills relative to other infants at 4 time points from 14 to 36 months of age (S6). In addition, the parents of these children were significantly more sensitive to their infants while at home. Of 29 school reform movements analyzed in a recent meta-analysis, Montessori obtained one of the largest effects on achievement ($d = .27$) despite the Montessori schools averaging only 3 years of implementation (S7). Because the researcher (Maria Montessori) died over 50 years ago, this would not be considered an

effect of superrealization (where the researcher is overseeing the implementation, see main article), and for reasons also noted in the main article, it might result in less optimal student outcomes.

Two studies of Head Start Programs begun around 1970 included Montessori as one of several programs to which 4-year-olds were randomly assigned (*S8-10*).

Unfortunately the implementation of the Montessori program was rather poor in both cases (e.g., only one age group in the class, short duration of program both daily and for the study, first year of classroom), and although the results were intriguing, the sample sizes were miniscule by the long-term follow up.

There are other studies of Montessori education, but either the control group leaves much to be desired, or the Montessori implementation was clearly poor or unspecified, limited to one classroom hence could simply reflect a teacher's influence, and so on. Many but not all show positive Montessori outcomes, yet all have significant scientific weaknesses.

viii. Supplementary Online References

- S1. Duax, *Journal of Research & Development in Education* **25**, 173 (1992)
- S2. J. Cohen, *Statistical Power for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ, 1988).
- S3. K. Rathunde, M. Csikszentmihalyi, *Elementary School Journal* **106**, 59-79 (2005).
- S4. K. R. Rathunde, M. Csikszentmihalyi, *American Journal of Education* **111**, 341-371 (2005).
- S5. K. R. Dohrmann, Paper Presented at the American Educational Research Association Annual Meeting, San Francisco 2006, April.
- S6. R. N. Emde, J. Robinson, J. Korfmacher, P. Spicer, N. Watt. (University of Colorado Health Sciences Center, 2001).
- S7. G. D. Borman, G. M. Hewes, L. T. Overman, S. Brown, *Review of Educational Research* **73**, 125-230 (2003).
- S8. L. B. Miller, R. P. Bizzell, *Child Development* **55**, 1570-1587 (1984).

- S9. L. B. Miller, J. L. Dyer, *Monographs of the Society for Research in Child Development* **40** (1975).
- S10. M. Karnes, A. Shewedel, M. Williams, in *As the twig is bent: lasting effects of preschool programs* C. f. L. Studies, Ed. (Lawrence Elbaum, Hillsdale, NJ, 1983) pp. 133-171.

ix. Data

Several variables were removed from the data file in the interest of protecting participant confidentiality.