Developing a tool to evaluate early childhood education implementation fidelity: Measure of Montessori Implementation pilot study results

Angela K. Murray*, Amelia J. Murray, Carolyn J. Daoust and
Heather E. Gerker

Achievement and Assessment Institute, University of Kansas, Lawrence, KS, United States

Introduction: The Measure of Montessori Implementation-Early Childhood (MMI-EC) is a first of its kind classroom observation instrument designed to support research in Montessori classrooms. Fidelity measurement is an important aspect of quality educational research and evaluation and is even more important for research on Montessori education because the name is not legally protected. Any school can claim to be Montessori regardless of their practices. In the absence of a tool like the MMI-EC, researchers have utilized a variety of proxies to gauge Montessori fidelity with a wide range of rigor.

Methods: The MMI-EC involves a 45-min live classroom observation period conducted while children engage in learning activities as well as a separate documentation of classroom characteristics and equipment. The live observation consists of 5-min time samples of teacher activity and the materials students are engaged with. The classroom characteristics include both Montessori learning materials as well as other features supportive of high-quality Montessori implementation. We present results from a pilot study of 81 classroom observations using the MMI-EC conducted in both public and private Montessori early childhood classrooms in metropolitan areas in Houston, TX; Denver, CO; Kansas City, MO/Lawrence, KS; Cincinnati, OH; and Washington, DC.

Results: This pilot enabled us to evaluate the instrument itself along with the digital tool used to gather data and the process of training observers. These initial results suggest that the MMI-EC holds promise for reliably measuring Montessori practices.

Discussion: This study represents a first step in the development of an efficient, psychometrically sound assessment of Montessori fidelity for use in future research and evaluation. We offer recommendations for modifications and improvements for future iterations.

KEYWORDS

Montessori, early childhood education, fidelity measurement, classroom observation, instrument development

1 Introduction

Montessori education is an individualized pedagogical approach emphasizing long-term development. With an estimated 3,000 Montessori schools operating in the U.S.—of which almost 600 are publicly funded [National Center for Montessori in the Public Sector (NCMPS), 2024], Montessori has become the most prominent alternative educational approach in American public schools (Debs et al., 2022). Demand for Montessori

education has outpaced the availability of quality programs resulting in increasing pressure for deviation from its original design (Debs, 2019). Further, the term "Montessori" is no longer legally protected, so any school can use it in their name regardless of the degree to which they follow the principles of the Montessori philosophy (Murray and Daoust, 2023). In addition, no single governing body exists to deliver or monitor Montessori teacher preparation, so teachers' experiences can differ substantially before they enter a classroom leading to a wide range of specific practices (Murray and Daoust, 2023). Finally, the growth of programs in lower resourced locations also increases pressure for adapting Montessori by employing untrained teachers and substituting or locally making less expensive materials and equipment (Murray and Daoust, 2023).

Montessori herself focused significant attention on preserving the integrity of her method as her ideas became more popular globally and she became a public figure. As the success she achieved in the first Children's House which opened in 1907 in Rome gained momentum, demand for training teachers grew well beyond what she could effectively manage (Kramer, 1988). Schools were expanding in countries across Western Europe and the United States as well as in India, China, Mexico, Japan, Australia, New Zealand, and Argentina (Debs GLOBAL INTRO). Montessori believed that failure to implement her method exactly as she intended would result in "distortion and exploitation" (Kramer, 1988, p. 224). As a result, diplomas awarded in her early training programs included language that holders were permitted to open Montessori schools but not to train others (Standing, 1998). In addition, Montessori patented and licensed the didactic materials used to implement her method, and she conducted quality assurance checks herself to ensure precision and durability. Licensed American and British materials manufacturers stated that the "apparatus is not a set of separable toys... [and] should not be purchased by anyone who does not intend a careful, intelligence use according to the principles of the Montessori method" (Boyd, 1914, p. 14). In the United States, they further stated that "infringers and imitators will be vigorously prosecuted" (Boyd, 1914, p. 14).

Despite Montessori's early efforts to maintain control of her method, significant variation exists today in its implementation. Even so, the primary features of Montessori classrooms across all age levels which are generally agreed upon within the Montessori community and are listed below [Montessori Public Policy Initiative (MPPI), 2015]:

- Teachers as guides with specialized training who design a prepared environment
- Curriculum based on specially designed, hands-on materials
- Three-year age groupings and extended uninterrupted work time
- Emphasis on independent knowledge-building through internal development rather than extrinsic rewards
- Children learning at their own pace following individual interests
- Freedom for children to choose what to work on, where to work, for how long and with whom to work.

Montessori programs also typically de-emphasize whole-class teaching, grading, and standardized testing (Lillard and Else-Quest, 2006). While a significant proportion of Montessori institutions serve preschool-aged children, the approach is implemented across a broad age range, encompassing educational programs from infancy through high school (Lillard and Else-Quest, 2006).

As Montessori education has gained popularity, especially in the public sector, the need for research on its effectiveness has also grown. Research on the efficacy of educational approaches or interventions provides evidence about whether they produce desired outcomes for students. A great deal of focus is generally given to measuring the outcomes themselves (a study's dependent variables), but there is growing recognition of the importance of rigorous measurement of the intervention itself (the independent variable) through evaluating implementation fidelity. Implementation fidelity measurement has its roots the field of psychiatry in the 1960s (Bond et al., 2000). In simplest terms, fidelity measurement is the degree to which a program or intervention follows the original model which in this case is over 100 years old. Evaluating the fidelity of an intervention's implementation involves identifying key elements of the intervention, collecting data, and assessing critical measures in terms of their psychometric properties including reliability and validity (Mowbray et al., 2003). Gathering such information can require significant investment in resources and capacities beyond the scope of any single study, but doing so can also strengthen the research by supporting claims about a program's effectiveness as well as providing evidence to understand whether non-significance is the result of an ineffective or simply poorly implemented intervention (Allor and Stokes, 2016).

While program fidelity is a key concern in the evaluation of any educational intervention, the history and distinctive characteristics of Montessori education render the assessment of its authenticity particularly critical and methodologically complex. Since Montessori education is highly individualized with the content and pace of lessons customized for each student within a single multi-age classroom, fidelity assessment cannot rely on simply observing teachers delivering carefully crafted lessons across the curriculum to an entire class of students at the same time. Furthermore, since children in Montessori classrooms have free choice during an extended work period which is ideally close to 3 h long, every child in the classroom may be engaged in a completely different activity at any given time during their work cycle making the task of observing and documenting activity in a Montessori classroom particularly challenging. Complicating things further, researchers wishing to draw conclusions about Montessori education have no simple criteria or widespread, rigorous school accreditation to use in their studies (Murray and Daoust, 2023). A variety of tools have been used for this purpose, but there is no widely accepted instrument (Culclasure et al., 2018; Lillard, 2012; Lillard et al., 2017). In previous research, authors have either developed their own fidelity measures which require significant resources or have relied on less rigorous criteria such as teacher training or school accreditation (Murray and Daoust, 2023).

A pilot study of a Teacher Questionnaire of Montessori Practices (TQMP) which relied on teacher self-report about

the practices in their classrooms provided some evidence of the measure's psychometric properties measuring Montessori implementation in early childhood and elementary classrooms (Murray et al., 2019). The results supported three constructs based on a Confirmatory Factor Analysis: classroom structure, curriculum, and children's freedom. The Structure construct included basic characteristics of an ideal Montessori classroom including age ranges, class size, recordkeeping practices, etc. Curriculum encompassed aspects of Montessori materials, lessons, classroom resources, etc. The Freedom construct was made up of items related to choices children make in the classroom. Research adapting the questionnaire developed by Murray et al. (2019) in contexts outside the U.S. was conducted in the Netherlands (de Brouwer et al., 2024) and in Italy (Scippo, 2023), but the field is early in understanding the implications of using tools in international contexts with the added complication of translations from English into Dutch and Italian, respectively.

Reliance on self-report offers an economical mechanism for examining fidelity; however, first-hand observation is significantly more rigorous and provides researchers with more confidence in relying on such measures particularly for larger grant-funded projects (McGrew et al., 2013). Thus, we set out to develop an efficient, psychometrically sound observation tool which can ascertain the degree of implementation of Montessori early childhood environments and can be employed across multiple projects. We call the tool being developed the Measure of Montessori Implementation – Early Childhood (MMI-EC).

Our primary research question in this pilot study is if the Measure of Montessori Implementation – Early Childhood (MMI-EC) observational tool demonstrates sufficient psychometric evidence to support its usefulness in reliably measuring the authenticity of Montessori environments for research purposes. The measures we will examine to answer this question include: (1) interrater reliability, (2) internal reliability consistency, and (3) correlation analysis. Results from this study lay the foundation for further development of this tool including refining items and enhancing observer training. Future research will add external measures to explore relationships between Montessori practices and other external measures.

2 Materials and methods

The MMI-EC pilot study involved classroom observations, conducted by trained observers, in five metropolitan areas with a significant Montessori presence: Kansas City, MO/Lawrence, KS ($N = 20$); Cincinnati, OH ($N = 28$); Denver, CO ($N = 17$); Richmond, VA ($N = 11$); and Houston, TX ($N = 5$). To address our research objectives, we conducted psychometric analysis of the MMI-EC tool which we will outline in the sections that follow.

2.1 Measure

The MMI-EC consists of four components with individual items and descriptive statistics for each listed in Tables 1–4:

(1) Teacher Activity which documents teacher actions while children are engaged in learning activities and includes both actions that are considered supportive of Montessori pedagogy as well as actions that would be considered contrary to authentic Montessori practices (Table 1).
(2) Engaged Materials which records which Montessori learning materials children are working with during the time available for them to exercise free choice (Table 2).
(3) Classroom Index which is a checklist of furnishings, equipment, and classroom organizational elements conducive to Montessori implementation (Table 3).
(4) Material Snapshot which includes a list of key hands-on Montessori learning materials that the observer counts and records as available in the classroom (Table 4).

The first two MMI-EC components together make up the Observation Checklist. The Observation Checklist includes nine 5 min time sampling observation periods which are rotated across the Engaged Materials (3 rotations) and Teacher Activity components for both the lead teacher (4 rotations) and an assistant teacher (2 rotations), if applicable, for a total of 45 min of observation. Within the Engaged Materials rotations, the observer notes which hands-on Montessori learning materials the children are using within each 5 min period. During the Teacher Activity rotations, observers record teacher behaviors, both supportive of and contrary to Montessori philosophy, as they interact with children and manage the classroom within each 5 min period. The results from each of the time sampling periods were aggregated to form a single measure for each item within each classroom.

The final two components of the MMI-EC make up the Room Inventory which is completed at a time when children are not involved in their free choice learning activities in the classroom, ideally when children were outside at recess or while children were eating lunch or having group instruction in circle time.

Observers participated in an 8 h training program developed by the researchers and completed on the Canvas Learning Management System. The training included videos describing Montessori education and the MMI-EC; specific modules on recognizing Montessori materials including interactive activities; details about using the digital data collection tool and how to score items; and the CITI Social & Behavioral Research human subjects protection tutorial. The training program concluded with a practice classroom observation visit with an experienced, trained Montessori practitioner we called a Montessori mentor. During the visit, the mentor and the trainee observer each completed the MMI-EC and then discussed the experience and compared differences in their observations so that the observer was prepared for their first classroom observation.

2.2 Content

In developing the items for the MMI-EC, we considered five criteria for measuring fidelity of implementation as outlined by O'Donnell (2008): adherence, duration, quality of delivery, participant responsiveness, and program differentiation which can broadly be categorized under the headings of structure (adherence, duration) and process (quality of delivery, program differentiation)

TABLE 1 Teacher activity descriptive statistics (sum across four time periods).

Item	Minimum	Maximum	Mean	Std. deviation
Presents a material	0	4	2.173	1.349
Tells instead of presenting	0	3	0.469	0.760
Focuses on lesson	0	4	2.395	1.366
Permits interruptions	0	4	0.988	1.066
Watches child use	0	4	2.938	1.176
Does not observe child use	0	1	0.074	0.264
Positions for max visibility	0	4	3.160	1.042
Positions w/limited visibility	0	4	1.062	1.298
Looks up/scans room	0	4	3.173	1.034
Escalating activity not noticed	0	1	0.099	0.300
Non-judgmental redirection	0	4	1.827	1.358
Unconstructive redirection	0	4	0.358	0.780
Talks eye to eye	2	4	3.765	0.507
Talks from a distance	0	4	0.679	1.105
Listens attentively	0	4	2.346	1.534
Mostly talks, not listening	0	2	0.383	0.681
Facilitates choice	0	4	1.938	1.307
Overrides choice	0	4	0.704	1.006
Touches respectfully	0	4	2.185	1.184
Unhelpful physical activity	0	1	0.012	0.111
Calm amid high emotions	0	4	0.593	1.046
Becomes agitated	0	2	0.074	0.307

Items in italics are negative indicators of effective Montessori practice.

with participant responsiveness having elements of both. The four components of the MMI-EC can be aligned with O'Donnell's (2008) criteria. While duration information is captured in a separate teacher profile survey which will be analyzed in a separate article, the other structural aspect of fidelity, adherence, is reflected in both the Classroom Index and Teacher Activity observation which assess the degree to which the equipment and types of teacher interactions align with the Montessori method as described in foundational literature. The process aspects of fidelity, made up of quality of delivery and program differentiation, are also reflected in the Teacher Activity observation in ascertaining the manner of teacher implementation of Montessori practices and in the Classroom Index in documenting that key features distinguishing Montessori from other types of educational settings are present. Participant responsiveness is evident in the Engaged Materials observation component which records the specific Montessori materials children are working with during the observation.

To develop individual items and scoring criteria for assessing fidelity using the MMI-EC, we relied on multiple sources including previous work by the authors. First, we referenced the Logic Model for Montessori Education proposed by Culclasure et al. (2019). Documentation supporting each individual item came from a number of specific sources including Montessori organizational standards, writings of Maria Montessori herself, and other peer

reviewed publications as illustrated in Appendix A. Once we developed the initial list of items to be included in the MMI-EC using the sources listed in the Appendix A, we had draft iterations of the tool evaluated while under development by Montessori scholars, Montessori practitioners, and a three-member Technical Advisory and Cultural Relevance Committee from outside of the field of Montessori education. This diverse group of scholars provided input on the MMI-EC as well as the training materials to identify any issues and to recommend revisions.

2.3 Analysis

Analyses were conducted using SPSS v. 29 and R v. 4.3.0 and included descriptive statistics, correlations, item-total correlation analysis, ANOVA comparison of means, and internal reliability consistency scale analysis using Cronbach's alpha (Cronbach, 1951). For the Observation Checklist components which involved multiple time sample rotations, data were combined across time periods. The total score for Teacher Activity was the sum of observations across time periods, and the total score for Engaged Materials was the maximum value observed across time periods for each item. We used the maximum value for Engaged Materials

TABLE 2 Engaged materials descriptive statistics (maximum number of students across three time periods).

Item	Minimum	Maximum	Mean	Std. deviation
Dressing frames	0	5	0.321	0.972
Room care	0	7	0.420	0.947
Washing	0	9	1.272	1.844
Polishing	0	7	0.296	0.980
Food preparation	0	12	2.667	3.428
Pouring	0	6	0.802	1.308
Flower arranging	0	2	0.099	0.339
Transfer	0	6	1.025	1.414
Pink tower	0	6	0.272	0.837
Red rods	0	3	0.111	0.447
Brown stair	0	6	0.383	1.146
Knobbed cylinders	0	4	0.296	0.715
Knobless cylinders	0	3	0.259	0.667
Geometric solids	0	2	0.037	0.247
Geometry cabinet	0	5	0.123	0.659
Geometric form cards	0	5	0.099	0.604
Constructive triangles	0	3	0.148	0.450
Sound cylinders	0	3	0.148	0.503
Montessori bells	0	4	0.173	0.587
Color tablets	0	7	0.309	1.158
Algebraic cubes	0	5	0.383	0.874
Blindfolds	0	2	0.049	0.269
Sandpaper letters	0	12	0.889	2.019
Sand tray	0	1	0.025	0.156
Moveable alphabet	0	13	1.778	2.574
Metal insets	0	6	0.370	0.993
Labeling	0	10	2.123	2.556
Miniature environment	0	4	0.259	0.787
Grammar symbol box	0	0	0.000	0.000
Sentence analysis	0	2	0.062	0.289
Chalkboard writing	0	3	0.235	0.694
Numbers rods	0	6	0.247	0.942
Spindle box	0	5	0.309	0.861
Numerals and counters	0	5	0.383	0.995
Teen/10 boards	0	9	0.593	1.634
Teen rack	0	4	0.210	0.627
100 board	0	9	1.049	1.680
Golden beads and cards	0	8	0.877	1.893
Bead chains	0	7	1.407	1.822
Fraction skittles	0	0	0.000	0.000
Fraction circles	0	2	0.025	0.222

(Continued)

TABLE 2 (Continued)

Item	Minimum	Maximum	Mean	Std. deviation
Stamp game	0	4	0.432	0.987
Bead frame	0	1	0.025	0.156
Montessori globes	0	2	0.086	0.360
Land and water forms	0	2	0.086	0.394
Puzzle maps	0	14	0.852	2.231
Botany cabinet	0	2	0.062	0.289
Botany leaf cards	0	1	0.012	0.111
Small country flags	0	3	0.099	0.464
Montessori clock	0	5	0.074	0.565
Biology puzzles	0	6	0.494	1.026
Easel painting	0	5	0.284	0.810
Workbooks	0	25	2.407	4.483
Markers	0	10	0.679	1.634

because children could have been engaged in the same work in multiple time rotations. Using the maximum value in any given rotation ensured that we did not confound continued work with materials with new students choosing the same work. The Teacher Activity scale items represented discrete actions, so the sum of these observations across time samples was used. The Material Snapshot items were a count of key Montessori materials available in the classroom, and Classroom Index items were dichotomous measures of present or not present.

When calculating the Cronbach's alpha for the Teacher Activity section, we reverse coded the negative items so that observed negative behaviors were recorded as 0s and absences of negative behaviors were recorded as 1s. Reverse coding allowed us to calculate the internal reliability of the negatively weighted items in the same scale as the positively scored ones. We also treated the negative items as a separate subscale from the positive items because we identified more problematic negative items and wished to examine them separately. After initial analysis, we refined the four scales of the MMI-EC and removed items with item-total correlations below 0.05.

In addition to internal reliability, we also conducted an interrater reliability analysis on a subset of 15 classrooms with paired observers. We calculated the exact percent agreement across each of the four scales. In evaluating the results, we considered percent agreement above the widely used cutoff of 75% to be acceptable for consensus (Graham et al., 2012). Alternative approaches for scoring such as dichotomizing scales were also utilized. In addition, Cohen's Kappa was calculated as a more rigorous measure of interrater reliability. Several interpretations of Cohen's Kappa were found in the literature. Although some suggest values greater than 0.50 are acceptable (Stemler and Tsai, 2008), we consider Landis and Koch's (1977) more detailed ranges for interpretation (<0 Poor, 0.01–0.20 Slight, 0.21–0.40 Fair, 0.41–0.60 Moderate, 0.61–0.80 Substantial, 0.81–1.00 Almost perfect).

After examining individual items of subscales, we calculated a subscale total scores for each component by taking the sum after

eliminating items that showed low internal consistency with other items. We calculated the intercorrelation between each subscale in order to determine how aligned each subscale was to an overarching measurement of fidelity. The sum of these subscale scores can represent a total fidelity score for each observation. We analyzed total fidelity score by whether the school was public or private using an ANOVA test. We also used each individual subscale score in separate ANOVA tests with public or private school as the independent variable.

2.4 Data sources

The MMI-EC pilot study involved 11 observers who all had at least an undergraduate degree in education, psychology, social sciences, or a related field. They may have had some familiarity with Montessori education but not a Montessori teaching credential. To increase the credibility of the MMI-EC, we intentionally avoided using Montessori experts as observers and designed the instrument with low inference items to reduce subjectivity and avoid factors other than the items being measured being used to assess a classroom's Montessori authenticity. With support from an external funder, observers were provided a $100 stipend for the training and for each observation conducted. Observers participated in an 8 h online observer training course using the KU Canvas learning management system and received support from a Montessori Mentor who was an experienced and credentialed early childhood Montessori educator. The Canvas course included 11 modules, ranging from topics on understanding the MMI-EC tool to recognizing Montessori materials (Murray et al., 2024).

Additional data was collected for the purpose of correlating MMI-EC results to external comparison measures. First, teachers were invited to complete the TQMP (Murray et al., 2019), and we obtained 45 completed questionnaires from classrooms we

TABLE 3 Classroom index descriptive statistics.

Item	Minimum	Maximum	Mean	Std. deviation
All classroom areas are well lit	0	1	0.963	0.190
Hard surfaces make up 25% or more of flooring	0	1	0.914	0.283
There is floor workspace for 8+ children	0	1	0.988	0.111
Walls minimally decorated, somewhat bare	0	1	0.691	0.465
Commercial school décor three pieces or less	0	1	0.852	0.357
90% of wall displays are at the child's eye level.	0	1	0.864	0.345
Self-serve snack eaten at a table w/1–3 others	0	1	0.815	0.391
Surfaces are clean and dust free	0	1	0.889	0.316
There are no disorganized areas	0	1	0.728	0.448
Shelves do not appear crowded	0	1	0.914	0.283
Chairs are of 2 or more heights	0	1	0.852	0.357
Tables are of 3 or more heights	0	1	0.877	0.331
8+ tables seat 1–2 children	0	1	0.765	0.426
0 to 2 tables seat 3 or more children	0	1	0.753	0.434
A 1–2 walking line is 98% or more intact.	0	1	0.469	0.502
There is an easy to use sink	0	1	0.951	0.218
Furniture does not block view of children	0	1	0.753	0.434
Reading area has seating, 8+ age-appropriate books	0	1	0.938	0.242
Classroom books are reality based	0	1	0.580	0.497
There are 3+ fragile, easily breakable items	0	1	0.951	0.218
Classroom tools are real & child-sized	0	1	0.988	0.111
There is a set of cursive sandpaper letters	0	1	0.617	0.489
1 or more animal is accessible to children	0	1	0.469	0.502
There are 6+ plants w/different leaf shapes	0	1	0.815	0.391
90% of classroom plants are healthy	0	1	0.901	0.300
Materials are grouped by subject	0	1	0.728	0.448
5 selected rugs are plain, clean, in good repair	1	1	1	0

observed. We also invited Montessori mentors to complete an informal rubric assessing key components for a small subset of observed classrooms based on their extensive experience as practitioners. We received eight completed rubrics which were scored on a four-point scale on the following items with each scale point offering a description of what reflects a particular score:

(1) Lead teacher activity reflects high fidelity Montessori practice (instruction, supervision, and atmosphere)
(2) Assistant teacher activity reflects high fidelity Montessori practice (instruction, supervision, and atmosphere)
(3) Children primarily engage with Montessori materials
(4) Classroom features and furnishings reflect a high fidelity Montessori environment (Overall classroom, furnishings, and resources)
(5) Montessori materials available reflect a high fidelity Montessori environment (practical life, sensorial, math, language, and cultural)

2.5 Participating classrooms

The MMI-EC pilot study involved 81 early childhood classroom observations with 21 public school classrooms participating Further, a subset of observations were conducted by two observers visiting the same 15 classrooms for inter-rater reliability (Hallgren, 2012), leaving 66 unique classroom observations. Criteria for classroom participation included having at least 18 children enrolled across at least a 2-year age range. Classroom teachers were not required to be Montessori trained. The mean number of children in attendance during observations was 19.49 with a standard deviation of 3.245.

This project was approved by the KU Institutional Review Board, and all participant observers and mentors submitted signed consent forms. With support from the Brady Education Foundation, observer/participants and Montessori Mentors were provided a $100 stipend for the training and for each observation conducted.

TABLE 4 Material snapshot descriptive statistics.

Item	Minimum	Maximum	Mean	Std. deviation
Dressing frames	0	12	5.235	2.570
Broom_Duster_Mop	0	8	2.074	1.687
Washing	0	6	2.494	1.467
Polishing	0	8	1.716	1.543
Food preparation	0	7	1.519	1.370
Pouring	0	7	2.877	1.382
Flower arranging	0	5	0.741	0.863
Transfer	0	8	3.407	2.066
Pink tower	0	2	0.988	0.193
Red rods	0	1	0.963	0.190
Brown stair	0	1	0.988	0.111
Knobbed cylinders	1	4	3.383	1.146
Knobless cylinders	0	4	2.975	1.466
Geometric solids	0	2	0.864	0.379
Geometry cabinet	0	1	0.938	0.242
Geometric form cards	0	2	0.778	0.447
Constructive triangles	0	6	3.457	1.891
Sound cylinders	0	2	1.012	0.433
Montessori bells	0	2	0.580	0.521
Musical boards and notes	0	1	0.173	0.380
Color tablets	0	5	2.160	1.112
Algebraic cubes	0	4	1.840	0.993
Blindfolds	0	6	1.074	1.547
Sandpaper letters	0	4	1.049	0.444
Double sandpaper letters	0	2	0.901	0.339
Sand tray	0	2	0.420	0.521
Moveable alphabets	0	11	3.111	2.387
Metal insets	0	2	0.963	0.247
Object boxes	0	18	2.840	2.896
Miniature environment (farm)	0	2	0.568	0.523
Miniature environment figures (farm animals)	0	2	0.741	0.543
Grammar symbols	0	1	0.457	0.501
3-part card(s)	0	9	1.889	1.754
Booklet set(s)	0	16	1.963	3.080
Sentence analysis first chart	0	2	0.259	0.494
Sentence analysis first box	0	1	0.222	0.418
Sentence analysis advanced chart	0	1	0.210	0.410
Chalkboards	0	11	2.556	2.495
Number rods	0	2	1.000	0.158
Spindle box	0	2	1.012	0.250
Counters and numerals	0	5	0.864	0.754

(Continued)

TABLE 4 (Continued)

Item	Minimum	Maximum	Mean	Std. deviation
Teen/ten boards	0	2	1.086	0.505
Teen rack	0	3	0.728	0.725
100 board	0	5	1.185	0.792
Multiplication board	0	3	0.877	0.640
Golden beads 1s, 10s, and 100s	0	3	1.025	0.387
Golden bead 1,000 cubes	0	13	1.543	1.898
Number cards 1–9,000	0	6	1.704	1.289
Trays	0	6	2.160	1.512
Short bead chains	0	2	1.000	0.224
Long bead chains	0	9	1.025	0.935
Bead chain arrows	0	1	0.914	0.283
Fraction skittles	0	1	0.272	0.448
Fraction circles 1–5	0	1	0.556	0.500
Fraction circles 6–10	0	1	0.543	0.501
Stamp game	0	5	0.975	0.821
Small bead frame	0	4	0.630	0.679
Dot exercise	0	5	0.679	0.878
Land and water globe	0	2	0.963	0.293
Continent globe	0	2	0.926	0.307
Land and water forms	0	2	0.593	0.519
World puzzle map	0	7	1.025	0.707
Puzzle maps	0	10	4.272	2.907
Botany cabinet	0	2	0.704	0.486
Botany cabinet leaf cards	0	1	0.444	0.500
Garden	0	3	0.704	0.558
Flags and clock	0	6	0.716	1.015
Montessori clock	0	1	0.358	0.482
Biology puzzles	0	12	2.802	2.981
Easel painting	0	1	0.556	0.500
Workbooks	0	24	2.938	6.090
Marker sets	0	11	1.185	2.242

3 Results

In this psychometric analysis of the pilot data, we present descriptive statistics, internal reliability statistics including coefficient alpha and item-total correlations as well as interrater reliability for the four scales that make up the MMI-EC. Descriptive statistics for all items in each section of the instrument are provided in Tables 1–4. These descriptive statistics give us an initial indication of which items have little variability due to either the vast majority or very few classrooms showing evidence of a particular aspect of Montessori education. In Table 5, we report the minimum, maximum, and mean internal consistency measured by Cronbach's alpha for each section. It also shows the reduced scale number of items and resulting coefficient alpha after eliminating items with item total correlations below 0.05 as described in the Section 2.

3.1 Internal reliability consistency

As shown in Table 5, we found the strongest internal reliability consistency with the Material Snapshot section which is at an

TABLE 5 Internal consistency of MMI-EC scales.

Scale	Initial		Reduced	
	Number of items	Coefficient alpha	Number of items	Coefficient alpha
Observation checklist				
Teacher activity (sum)	22	0.745	19	0.784
Teacher activity-positive behaviors	11	0.772	11	0.772
Teacher activity-negative behaviors	11	0.429	8	0.558
Engaged materials (dichotomized)	54	0.644	38	0.692
Engaged materials (% of number of children)	54	0.459	36	0.617
Room inventory				
Classroom index	26	0.705	23	0.716
Material snapshot	72	0.611	54	0.811

acceptable level using the common criteria of a minimum of 0.70 (Tavakol and Dennick, 2011). The Material Snapshot scale is also the most extensive with 72 items initially. Since many items did not contribute to the internal consistency, with item total correlations below 0.05, we removed 18 items to achieve produced a reasonable degree of internal consistency. The Classroom Index scale began with 26 items and after removing 3 items with item total correlation below 0.05 gave us an internal consistency just over the 0.70 threshold. The positive Teacher Activity scale items began with 11 items and did not require removal of any of them for having an item total correlation of below 0.05. The internal reliability for the positive Teacher Activity scale items was acceptable at the 0.70 cutoff. The Engaged Materials scale (even if we dichotomize it) and the negative items on the Teacher Activity scale were less consistent and require additional scrutiny (Taber, 2018). Tables 6–9 show item-total correlations of individual items in the reduced scales for each of the four sections of the MMI-EC with the Teacher Activity scale separated between positive and negative behaviors.

We also examined the correlation among the MMI-EC subscales to better understand the similarities and differences in what each is measuring. The results in Table 10 demonstrate that the overlap for the subscales is rather limited with the highest correlations being between negative and positive Teacher Activity, between Engaged Materials and Material Snapshot, and between Engaged Materials and the Classroom Index. These correlations make intuitive sense because one would expect that positive Teacher Activity that is conducive to Montessori education would be related to an absence of negative Teacher Activity which would be detrimental to Montessori education. Similarly, it is logical to expect that classrooms with a larger number of Montessori materials available in the classroom would be related to classrooms in which more children are actively engaged in working with Montessori materials. And, finally, classrooms that are well equipped with Montessori materials would also be expected to have the resources to provide other important classroom features to support an effective Montessori environment. Correlations among other subscales are negligible. Table 10

TABLE 6 Teacher activity item-total correlation.

Item	Item-total correlation
Positive behaviors	
Facilitates choice	0.639
Looks up/scans room	0.599
Watches child use	0.566
Listens attentively	0.559
Focuses on lesson	0.452
Talks eye to eye	0.451
Positions for max visibility	0.388
Non-judgmental redirection	0.365
Presents a material	0.273
Touches respectfully	0.254
Calm amid high emotions	0.239
Negative behaviors	
Unconstructive redirection	0.373
Mostly talks, not listening	0.359
Overrides choice	0.344
Tells instead of presenting	0.330
Becomes agitated	0.278
Talks from a distance	0.265
Permits interruptions	0.191
Does not observe child use	0.183

also shows the correlation of each of the subscales with the total MMI-EC scores, showing the strongest relationships with a strong positive correlation between total MMI-EC and positive Teacher Activity. Negative Teacher Activity and Material Snapshot are both moderately correlated with the MMI-EC total score. Engaged Materials and Classroom Index are

TABLE 7 Engaged materials item-total correlation (dichotomized maximum across time periods).

Item	Corrected item-total correlation
Numerals and counters	0.360
Transfer	0.332
Easel painting	0.319
Flower arranging	0.307
Dressing frames	0.301
Polishing	0.298
Pink tower	0.292
Miniature environment	0.288
Labeling	0.287
Room care	0.285
Metal insets	0.257
Markers	0.248
Color tablets	0.220
Bead frame	0.219
Moveable alphabet	0.214
Washing	0.213
100 board	0.211
Stamp game	0.206
Montessori bells	0.198
Blindfolds	0.198
Food preparation	0.195
Botany leaf cards	0.195
Spindle box	0.191
Brown stair	0.186
Pouring	0.170
Biology puzzles	0.163
Sentence analysis	0.145
Chalkboard writing	0.143
Teen rack	0.131
Botany cabinet	0.131
Golden beads and cards	0.124
Sound cylinders	0.101
Sand tray	0.100
Algebraic cubes	0.098
Puzzle maps	0.089
Red rods	0.076
Workbooks	0.071
Fraction circles	0.056

TABLE 8 Classroom index item-total correlation.

Item	Item-total correlation
Walls minimally decorated, somewhat bare	0.431
8+ tables seat 1–2 children	0.393
There are 6+ plants w/different leaf shapes	0.369
There are no disorganized areas	0.357
0 to 2 tables seat 3 or more children	0.354
A 1–2 walking line is 98% or more intact.	0.319
90% of classroom plants are healthy	0.308
Materials are grouped by subject	0.300
Chairs are of 2 or more heights	0.291
Classroom books are reality based	0.285
There is a set of cursive sandpaper letters	0.281
1 or more animal is accessible to children	0.276
Hard surfaces make up 25% + of flooring	0.264
There are 3+ fragile, easily breakable items	0.250
Surfaces are clean and dust free	0.246
Commercial school décor 3 pieces or less	0.245
Self-serve snack eaten at table w/1–3 others	0.231
90% of wall displays are at child's eye level.	0.213
Tables are of 3 or more heights	0.156
Classroom tools are real and child-sized	0.146
Furniture does not block view of children	0.144
There is floor workspace for 8+ children	0.111
Reading area has seating, 8+ age-appropriate books	0.052

3.2 Interrater reliability

We also conducted an analysis of the 15 classrooms with two observers gathering data simultaneously to gauge interrater reliability. We used percentage rater agreement based on how frequently the two raters assigned the exact same rating. Results for the percent agreement calculations for each scale are shown in Table 11. We also reevaluated the percent agreement with dichotomized items for scales that utilized time sample rotations (Teacher Activity and Material Snapshot). In other words, we examined only if the action was observed in any of the time samples rather than using the sum or maximum of each time sample. The Material Snapshot was also dichotomized which resulted in changing the items from counting the number of specific materials available in the classroom to a checklist representing whether or not each material was present in the classroom. For example, instead of the value for dressing frames being the number of individual frames available to being an indicator of whether or not any dressing frames were available in the classroom. The Classroom Index scale was already a dichotomous score, so we did not adjust the scale. With these updated dichotomous percent agreement scores, every subscale satisfied the 75% cutoff with the exception of the negative

significantly but weakly correlated with the overall total MMI-EC scale score.

TABLE 9 Material snapshot item-total correlation.

Item	Item-total correlation
Sentence analysis first chart	0.561
Fraction circles 6–10	0.535
Trays	0.534
Fraction circles 1–5	0.524
Polishing	0.522
Dot exercise	0.509
Food preparation	0.507
Sentence analysis first box	0.464
Land and water forms	0.448
Sentence analysis advanced chart	0.445
Blindfolds	0.435
Botany cabinet leaf cards	0.426
Flower arranging	0.421
Geometric form cards	0.420
Montessori clock	0.403
Washing	0.400
Montessori bells	0.392
Musical boards and notes	0.386
Miniature environment (farm)	0.377
Botany cabinet	0.367
Color tablets	0.362
Grammar symbols	0.356
Sound cylinders	0.353
Miniature environment figures (farm animals)	0.344
Golden bead 1,000 cubes	0.334
Flags and clock	0.327
Metal insets	0.318
3-part card(s)	0.316
Double sandpaper letters	0.291
Small bead frame	0.281
Constructive triangles	0.268
Easel painting	0.266
Bead chain arrows	0.265
Continent globe	0.258
Golden beads 1s, 10s, and 100s	0.255
Geometry cabinet	0.252
Chalkboards	0.234
Stamp game	0.223
Garden	0.220
Pouring	0.213
Multiplication board	0.212

(Continued)

TABLE 9 (Continued)

Item	Item-total correlation
Broom/Duster/Mop	0.207
Sandpaper letters	0.200
Land and water globe	0.173
Teen/10 boards	0.159
World puzzle map	0.146
Short bead chains	0.144
Long bead chains	0.133
Knobless cylinders	0.127
Number cards 1–9,000	0.107
Puzzle maps	0.102
Pink tower	0.086
Spindle box	0.068
Counters and numerals	0.054

Teacher Activity items. In fact, with the exception of the negative Teacher Activity scale which only had 36.36% of individual items achieving the cutoff of 75% agreement, more than 70% of all items within each of the other subscales met the 75% exact agreement cutoff, ranging from 72.73% for the positive Teacher Activity items to 90.74% for the Engaged Materials items. The highest and lowest percent agreement items are listed in Table 12.

Cohen's Kappa values were also calculated in order to further examine interrater reliability. Across the 15 classrooms, Teacher Activity (positive and negative items combined) (K = 0.482, SE = 0.176) and Engaged Materials (K = 0.573, SE = 0.128) fall into the Moderate range for interrater reliability while the Material Snapshot (K = 0.713, SE = 0.093) shows good interrater reliability. The Classroom Index (K = 0.370, SE = 0.182) had the lowest interrater reliability according to Cohen's Kappa and the average fell into the Fair range. When investigating which classrooms had lower interrater reliability, we found that there were several instances of straightlining where observers rated a scale with all 1s and no 0s. Since this reduces Cohen's Kappa values to 0 due to the likelihood of random chance causing agreement, Kappa statistics were also calculated excluding these straightlined observations. With this adjustment, the Classroom Index (K = 0.427, SE = 0.210) average Cohen's Kappa was in the Moderate range. Several observers also straightlined the Teacher Activity positive items, so Cohen's Kappa was also examined with those observations excluded. The average positive Teacher Activity Kappa (K = 0.337, SE = 0.270) after excluding straightlined observations was still in the Fair range.

3.3 MMI-EC comparison between public and private schools

We compared the total MMI-EC score and subscale scores between public and private schools and found significant differences only for Classroom Index and Material Snapshot when

TABLE 10   Correlation across MMI-EC subscales.

Subscale	Positive teacher activity	Negative teacher activity	Engaged materials	Classroom index	Material snapshot
Negative teacher activity	0.307**				
Engaged materials	0.190	−0.033			
Classroom index	0.145	0.055	0.198		
Material snapshot	0.209	0.048	0.396***	0.391**	
Total score	0.839***	0.576***	0.396***	0.385***	0.547***

**significant at 0.01 level; ***significant at 0.001 level.

TABLE 11   Exact agreement across items.

Subscale	Original			Dichotomous		
	Minimum	Maximum	Average	Minimum	Maximum	Average
Teacher 1 positive	13.3%	100.0%	33.3%	26.7%	100.0%	80.0%
Teacher 1 negative	13.3%	100.0%	62.4%	33.3%	100.0%	68.5%
Engaged materials	33.3%	100.0%	83.2%	46.7%	100.0%	89.0%
Material snapshot	20.0%	100.0%	73.1%	53.3%	100.0%	89.3%
Classroom index	33.3%	100.0%	83.0%	33.3%	100.0%	83.0%

using a Bonferroni correction to control for inflation of Type I error rates due to multiple comparisons (Abdi, 2007). Private school classrooms on average had significantly higher scores on the Classroom Index and Material Snapshot (see Table 13). None of the other components of the MMI-EC demonstrated differences between public and private schools. One could explain this result by considering the limited resources and competing requirements of public schools as compared to private Montessori schools with ostensibly greater latitude to arrange classrooms in ways that reflect effective Montessori practices and are likely to have the budgetary resources and supportive administration to allow for purchasing a larger number of Montessori materials.

3.4 Relationships with external measures

We analyzed correlations between subscales of the MMI-EC and the TQMP scale for the 45 observed classrooms where teachers completed the questionnaire. Overall, the Pearson correlation between the total MMI-EC scale and two of the three TQMP constructs, Structure and Curriculum, are significant but weak. The relationship of the total MMI-EC scale with the third construct, Freedom, is not significant. These results suggest that while there are some similarities in the aspects of Montessori quality measured by each of these instruments, they do not represent the same constructs. Table 14 illustrates the Pearson correlation for the individual MMI-EC subscales and the TQMP constructs showing that a moderate relationship exists between the Classroom Index and the TQMP Structure and Curriculum constructs. Similar results are evident for Material Snapshot having a moderate degree of correlation with the TQMP constructs of Structure and Curriculum. A small but significant relationship also exists between Engaged Materials and the TQMP Structure construct.

Relationships among the other TQMP constructs and MMI-EC subscales are negligible and non-significant.

Finally, we analyzed the small subgroup of eight observed classrooms where a Montessori Mentor also completed a quality rubric. The extremely small number of completed rubrics as well as the relatively narrow range of scores on the rubric scales limits the value of this analysis. However, even at a high level the results showed that the two classrooms that received the highest scores from mentors (19 out of a total of 20 possible points) also scored above the median on the MMI-EC, and the two classrooms that received the lowest scores from mentors (12 and 15 points out of a total of 20 possible) also scored below the median on the MMI-EC. Clearly, these results are preliminary but do point to some support for the MMI-EC's ability to measure Montessori quality.

4 Discussion

The primary research question that guided this pilot study was whether the MMI-EC observational tool demonstrated adequate psychometric properties to support its validity as a reliable instrument for assessing the authenticity of Montessori environments in research contexts. To address this question, the study examined (1) interrater reliability, (2) internal consistency reliability, and (3) correlation analyses. The goal was to introduce a measure that could be useful across multiple research projects. Results from this initial pilot suggest that the MMI-EC holds promise for reliably measuring Montessori practices.

Although program fidelity is a central concern in the evaluation of any educational intervention, the historical development and distinctive pedagogical features of Montessori education make the assessment of its authenticity especially important and methodologically challenging (Mowbray et al., 2003; Murray and Daoust, 2023). These challenges stem from the individualized

TABLE 12 Highest and lowest percent agreement items by subscale (dichotomized).

Subscale	Item	Percent agreement
Highest		
Teacher 1 positive	5- Watches child use, 9- Looks up/scans room, 13- Talks eye to eye	100.00
Teacher 1 negative	20- Unhelpful physical activity	100.00
Engaged materials	14 items out of 36	100.00
Material snapshot	26 items out of 54	100.00
Classroom index	7 items out of 23	100.00
Lowest		
Teacher 1 positive	11- Non-judgmental redirection	26.67
Teacher 1 negative	4- Permits interruptions	33.33
Engaged materials	53-Workbooks	46.67
Material snapshot	72- Marker Sets	53.33
Classroom index	15- 1–2 walking line is 98% or more intact.	33.33

TABLE 13 MMI-EC subscale comparisons of private and public schools.

Subscale	Measure	Mean	SD	$F_{(1,79)}$	p
Total scale	Private	8.29	1.10	1.969	0.164
	Public	7.93	1.00		
Teacher 1 positive	Private	26.43	7.77	0.014	0.905
	Public	26.64	6.25		
Teacher 1 negative	Private	28.54	3.21	1.254	0.266
	Public	27.68	3.11		
Engaged materials	Private	16.25	10.21	2.594	0.111
	Public	20.4	11.79		
Classroom index	Private	19.04	2.87	13.88	<0.001
	Public	16.32	3.36		
Material snapshot	Private	66.50	16.58	7.498	0.008
	Public	55.88	15.02		

Total scale is reported as a mean while subscales are reported as totals so that the scales with more items do not have an inflated impact on the total.

TABLE 14 MMI-EC subscale correlation with TQMP constructs.

Subscale	TQMP constructs		
	Structure	Curriculum	Freedom
Positive teacher activity	0.153	−0.007	−0.081
Negative teacher activity	0.028	0.138	−0.199
Engaged materials	0.344*	0.229	0.030
Classroom index	0.479**	0.530**	0.093
Material snapshot	0.622**	0.673**	0.121

*significant at 0.05 level; **significant at 0.01 level.

structure of the Montessori curriculum, the extended work cycle and the simultaneous engagement of children in diverse tasks within mixed-age classrooms, as well as the absence of a centralized governing authority (Murray and Daoust, 2023). One way the MMI-EC addressed these challenges is by relying on the distinctive hands-on learning materials upon which much of the Montessori curriculum is based to offer visual evidence for understanding Montessori implementation in individual classrooms. We found that a reduced list of available Montessori materials (Material Snapshot) shows strong internal consistency, high interrater reliability, and a moderate correlation to an external measure of Montessori quality. A checklist of features and equipment in Montessori classrooms (Classroom Index) also shows solid internal consistency, high interrater reliability, and overlap with the TQMP constructs identified in previous research. A measure of the Montessori materials children are working with during their free choice time (Engaged Materials) approached the cutoff of reasonable internal consistency and had strong interrater reliability but no relationship with TQMP constructs.

We also attempted to leverage Montessori teacher practices as a way to overcome the challenges of ascertaining implementation fidelity. Observations of teacher behaviors (Teacher Activity-positive) theoretically supported as conducive to Montessori pedagogy also have solid internal consistency and interrater reliability. However, teacher behavior (either positive or negative) does not support the constructs identified in previous research investigating the teacher report measure TQMP. Finally, observations of teacher behaviors that are theoretically supported as being detrimental to high quality Montessori implementation (Teacher Activity-negative) have relatively weaker internal consistency and lower interrater reliability, and these negative behaviors do not correspond to constructs of the TQMP.

Our results suggest a number of next steps for the development of the MMI-EC and ultimately an elementary version of the instrument. First, classroom resources can be a valuable component for assessing Montessori fidelity. Additional components related to teacher behavior can add important information, but additional refinement of the scales is necessary particularly for contrary indicators. Children's engagement is an important aspect of Montessori pedagogy, so further exploration is necessary for ensuring a manageable process for observers documenting children's activities as well as for considering if additional indicators of engagement beyond focusing exclusively on Montessori materials may be useful. Finally, freedom is a key component of Montessori education, and current components of the MMI-EC do not seem to capture it. It is worth considering the possibility that other teacher behaviors beyond those documented in the MMI-EC may be necessary to fully represent Montessori fidelity.

We acknowledge a number of limitations encountered in the existing study. First, while this is the largest study of its kind, 66 unique classrooms in five sites may not fully reflect the breadth and variety of Montessori practices in the U.S. or globally. Second, the limited opportunity to see classroom activity during the time sampling rotations for the Teacher Activity and Engaged Materials sections may prove to be insufficient for truly reflecting the essence of Montessori education. Third, additional resources to supervise

and train observers more thoroughly with additional opportunities for practicing and refining their observation skills would likely improve the results. Finally, we had less control of the devices being used by the observers since this study was conducted by such a large and geographically dispersed group of observers. The differences in devices and the possible additional cognitive load on observers to navigate less than ideal devices and to manage a separate timer during time sampling periods may have introduced additional measurement error that could be improved with more seamless tools. Future work to more fully validate the MMI-EC should consider significantly more resources for observer management and support.

Fidelity measurement is an important aspect of quality educational research and evaluation. While observation tools can require a significant investment in development resources and data collection effort, the investment can be outweighed by the benefits of enhanced rigor and stronger validity arguments particularly for large-scale, grant-funded projects. Significant growth in Montessori in the public sector leads to increasing demand for this type of quality Montessori research [National Center for Montessori in the Public Sector (NCMPS), 2024]. This multi-year, multi-state study is the first of its kind to focus primarily on developing a psychometrically sound instrument for the purposes of assessing fidelity of Montessori practices for research purposes. The development of the MMI-EC offers a first step in the process of establishing a robust tool which will provide a consistent definition of fidelity for research purposes. Findings from this investigation provide a foundational basis for the continued development of the MMI-EC, including the refinement of individual items and improvements to observer training protocols. While work remains to optimize and calibrate the instrument, this initial study supports the potential of the MMI-EC to offer a psychometrically sound instrument for researchers to use in their research in Montessori environments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University of Kansas Human Research Protection Program Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2025.1603908/full#supplementary-material

References

Abdi, H. (2007). "Bonferroni test," in *Encyclopedia of Measurement and Statistics*, ed. N. J. Salkind (Thousand Oaks, CA: Sage), 175–180.

Allor, J. H., and Stokes, L. (2016). "Measuring treatment fidelity with reliability and validity across a program of intervention research: practical and theoretical

considerations," in *Treatment Fidelity in Studies of Educational Intervention* (London: Routledge), 138–155.

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., and Kim, H.-W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Ment. Health Serv. Res.* 2, 75–87. doi: 10.1023/A:1010153020697

Boyd, W. (1914). *From Locke to Montessori: A Critical Account of the Montessori Point of View*. London: G.G. Harrap.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Culclasure, B., Fleming, D. J., Riga, G., and Sprogis, A. (2018). *An evaluation of Montessori education in South Carolina's public schools*. The Riley Institute, Furman University, Greenville, SC, United States.

Culclasure, B. T., Daoust, C. J., Cote, S. M., and Zoll, S. (2019). Designing a logic model to inform Montessori research. *JMR* 5, 35–49. doi: 10.17161/jomr.v5i1.9788

de Brouwer, J., Morssink-Santing, V., and van der Zee, S. (2024). Validation of the teacher questionnaire of montessori practice for early childhood in the Dutch context. *JMR* 10, 12–24. doi: 10.17161/jomr.v10i1.21543

Debs, M. (2019). *Diverse Families, Desirable Schools: Public Montessori in the Era of School Choice*. Cambridge: Harvard Education Press.

Debs, M., de Brouwer, J., Murray, A. K., Lawrence, L., Tyne, M., and von der Wehl, C. (2022). Global diffusion of montessori schools: a report from the 2022 Global Montessori Census. *JMR* 8, 1–15. doi: 10.17161/jomr.v8i2.18675

Graham, M., Milanowski, A., and Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings (ED532068)*. ERIC: Westat Center for Educator Compensation Reform. Available online at: https://files.eric.ed.gov/fulltext/ED532068.pdf (Accessed September 17, 2025).

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor. Quant. Methods Psychol.* 8:23. doi: 10.20982/tqmp.08.1.p023

Kramer, R. (1988). *Maria Montessori: A Biography*. New York: Hachette Books.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

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

Lillard, A. S. (2012). Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs. *J. Sch. Psychol.* 50, 379–401. doi: 10.1016/j.jsp.2012.01.001

Lillard, A. S., Heise, M. J., Richey, E. M., Tong, X., Hart, A., and Bray, P. M. (2017). Montessori preschool elevates and equalizes child outcomes: a longitudinal study, *Front. Psychol.* 8:1783. doi: 10.3389/fpsyg.2017.01783

McGrew, J. H., White, L. M., Stull, L. G., and Wright-Berryman, J. (2013). A comparison of self-reported and phone-administered methods of ACT fidelity assessment: a pilot study in Indiana. *Psychiatr. Serv.* 64, 272–276. doi: 10.1176/appi.ps.001252012

Montessori Public Policy Initiative (MPPI) (2015). *Montessori essentials*. Available online at: https://montessoriadvocacy.org/wp-content/uploads/2019/07/MontessoriEssentials.pdf (Accessed September 17, 2025).

Mowbray, C. T., Holter, M. C., Teague, G. B., and Bybee, D. (2003). Fidelity criteria: development, measurement, and validation. *Am. J. Eval.* 24, 315–340. doi: 10.1177/109821400302400303

Murray, A. K., and Daoust, C. J. (2023). "*Fidelity issues in Montessori research,*" in *The Bloomsbury Handbook of Montessori Education,* eds. A. Murray, E.-M. Tebano Ahlquist, M. McKenna, and M. Debs (London: Bloomsbury Publishing), 199–208.

Murray, A. K., Daoust, C. J., and Chen, J. (2019). Developing instruments to measure Montessori instructional practices, *JMR* 5, 50–74. doi: 10.17161/jomr.v5i1.9797

Murray, A. K., Daoust, C. J., Murray, A. J., and Gerker, H. E. (2024). "Developing an observational measure of montessori implementation [conference presentation]," in *American Educational Research Association* (Philadelphia: AERA Online Paper Repository).

National Center for Montessori in the Public Sector (NCMPS) (2024). *About Montessori*. Available online at: https://www.public-montessori.org/montessori/ (Accessed September 17, 2025).

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K−12 curriculum intervention research, *Rev. Educ. Res.* 78, 33–84. doi: 10.3102/0034654307313793

Scippo, S. (2023). Costruzione e validazione di uno strumento per misurare le pratiche educative Montessori nella scuola primaria italiana. *J. Educ. Cult. Psychol. Stud.* 1, 117–135. doi: 10.7358/ecps-2023-028-scis

Standing, E. M. (1998). *Maria Montessori: Her Life and Work*. Palo Alto, CA: Plume.

Stemler, S. E., and Tsai, J. (2008). Best practices in interrater reliability three common approaches. *Best Pract. Quant. Methods* 29–49. doi: 10.4135/9781412995627.d5

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd