Implementation of Montessori Education: Validity Evidence of Dutch Teacher Questionnaires for Montessori Practices

Jaap de Brouwer[ab], Lida T. Klaver[a], Patrick H. M. Sins[bc], and Symen van der Zee[a]
[a]School of Education, Saxion University of Applied Sciences
[b]Thomas More University of Applied Sciences
[c]Rotterdam University of Applied Sciences

Keywords: *progressive education, Montessori education, implementation measurement, validation, evaluation*

Abstract: Meta-analyses on the effectiveness of Montessori education report positive outcomes on both academic and social measures. However, Montessori schools vary considerably, and there is limited empirical literature examining how differences in implementation affect outcomes. To better understand the effectiveness of Montessori education, it is essential to examine the quality of its implementation and the outcomes that follow. Therefore, building upon instruments designed to measure Montessori implementation, this study strengthens the Dutch Teacher Questionnaires for Montessori Practices Revisited (D-TQMP-Rs) and examines validity evidence. In addition, this study explores how Montessori principles are implemented in Dutch Early Childhood and Elementary settings. Results provide initial validity evidence for both the Early Childhood and Elementary versions of the D-TQMP-R, with coherent factor structures and acceptable reliability, though limitations in discriminant and convergent validity remain. Montessori principles appeared to be implemented at a reasonably high level overall. The D-TQMP-Rs offer promising tools for assessing implementation of Montessori education, suitable for research and reflective practice, although still requiring further refinement.

Italian physician and educationalist Maria Montessori (1870–1952) gained international fame in the early 20th century through her Casa dei Bambini, the first Montessori school, which was introduced in 1907. Montessori theorized and realized a more individualized, child-centered approach to education, based on her close observations of young children in relatively free learning environments (Culclasure et al., 2019; Quarfood, 2022). From its origins, Montessori education has emphasized the holistic development of children, nurturing both their social and cognitive skills to foster autonomy and independent functioning. With almost 16,000 schools in 154 countries around the world, Montessori education is now by far the largest educational reform movement and the most researched alternative educational approach in the world (Debs et al., 2022; Guerrero et al., 2024).

A growing body of research is examining the effectiveness of Montessori education, with recent meta-analyses reporting positive academic and social outcomes for Montessori students (Demangeon et al., 2023; Randolph et al., 2023). However, these findings should be interpreted with caution as there are concerns regarding the adherence to Montessori practices in the studies included. For example, in the meta-analysis by Demangeon et al. (2023) only eight of the 33 included studies provided any information on the degree of Montessori implementation. Therefore, it remains unknown if and how specific principles (e.g., using Montessori materials, mixed-age classrooms) are related to particular student outcomes (Abry et al., 2015). Understanding how Montessori principles are applied in practice is needed for drawing more accurate conclusions about the effectiveness of Montessori education and for identifying which specific principles contribute to observed effects (Marshall, 2017; Nunes et al., 2024). In line with Marshall (2017), Demangeon et al. (2023) and Randolph et al. (2023) conclude that when studying Montessori education, there is a pressing need to examine how Montessori principles are applied in practice. Despite calls for measuring Montessori implementation, there is an absence of widely accepted instruments capable of assessing Montessori implementation in practice, as the field is still in an early stage of its development (Murray & Daoust, 2023; Murray et al., 2025).

To address this issue, the current study refines and improves the Dutch Teacher Questionnaires for Montessori Practices (D-TQMP). This instrument is designed to measure the degree to which Montessori principles are implemented in Early Childhood and Elementary classroom settings (De Brouwer et al., 2024; De Brouwer et al., 2024, 2025). Although earlier versions yielded promising results, revisions are needed to strengthen factor reliability, refine item wording, and include additional items to more accurately capture Montessori implementation. Because instrument validation is an ongoing process, the present study aims to refine the two D-TQMP versions—Early Childhood and Elementary—and examine their psychometric quality.

Background

Measurement of Montessori implementation remains an emerging field. Initial efforts can be traced to Murray et al. (2019), who developed the Teacher Questionnaire for Montessori Practices (TQMP) based on the Montessori logic model by Culclasure et al. (2019). The instrument includes separate versions for Early Childhood and Elementary education, reflecting Montessori's (1973) distinction between developmental phases (ages 0–6, 6–12, 12–18, 18–24). These phases emphasize different educational principles, necessitating differentiated measurement approaches.

The earlier version of TQMP comprised of three factors measuring Montessori education: *Classroom Structure*, *Children's Freedom*, and *Curriculum*. *Classroom Structure* was defined as the ways in which the group is organized, and how children are instructed to foster their independence and individual development (e.g., Montessori, 1937, 1949, 1997). *Children's Freedom* was defined as the extent to which children have the freedom to make their own choices (e.g., Montessori, 1935, 1937, 1997). Two distinct versions of the *Curriculum* construct were defined to reflect the developmental stages and educational focus of the respective age groups: Early Childhood and Elementary classrooms. The *Curriculum for Early Childhood* was defined as carefully constructed materials and activities designed to support the child's holistic development, encompassing areas from social skills to mathematics. In contrast, the *Curriculum for Elementary Classrooms* was defined as the structured sequence of lessons and materials that represent an integrated approach across multiple disciplines (e.g., Montessori, 1933, 2020a, 2020b). Although initial psychometric results were promising, Murray et al. (2019) emphasized the need for further refinement and validation evidence across contexts.

Montessori (2023) described her educational principles as universal, which makes assessing the external validity of the TQMP a logical next step. Considering that the TQMP is based on Montessori's universal principles, the three-factor structure Murray et al. (2019) identified

should also emerge in other populations and educational contexts, thereby demonstrating evidence of external validity. Replicating the Murray et al. (2019) study in different national Montessori settings therefore provides essential evidence about whether the instrument captures universal aspects of Montessori practice. If the expected structure does not replicate, this indicates that the TQMP reflects context-specific interpretations shaped by particular educational environments. In such cases, further refinement of the instrument offers a constructive way forward within those contexts as part of its continued development (Mook, 2016).

This need for external validation becomes evident when examining validity evidence from other national contexts. For example, Scippo (2023), who used the TQMP in the Italian context, did not found the original three-factor structure of the TQMP of Murray et al. (2019), meaning teachers' Montessori practices relate differently in Italy than in the United States, reflecting context-specific interpretations of Montessori practice. Instead, exploratory factor analysis (EFA) yielded an eight-factor model, from which two overarching factors—*Core Montessori Foundations* and *Non-Montessori Contaminations*—were identified as most informative for implementation fidelity.

Similarly, De Brouwer et al. (2024, 2025) examined the validity of the TQMP in the Dutch context. Confirmatory factor analyses (CFAs) failed to support the original three-factor structure Murray et al. reported (2019), necessitating EFA (Hair et al., 2020). In the Early Childhood version, this resulted in two factors original to the American TQMP—*Children's Freedom* and *Curriculum*—alongside a new factor: *Teacher Guidance*. In the Elementary version, a five-factor structure emerged: *Prepared Environment*, *Extended Environment*, *Teacher Guidance*, *Cosmic Education*, and *Self-Education*. The findings indicated that the Dutch TQMP (D-TQMP) requires further refinement through item revision, theoretical alignment, and additional data collection. These studies show that the original structure of the TQMP was also not valid within the Dutch context, making further instrument development a necessary next step.

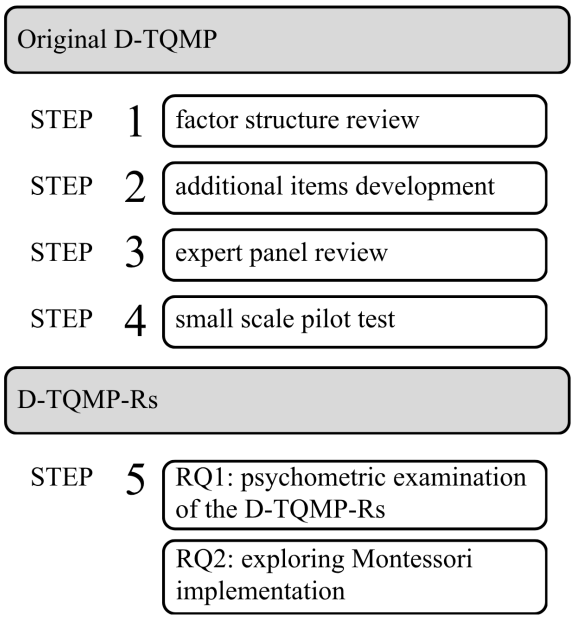
Why Context Matters in Measuring Montessori Implementation

The variation in factor structures between contexts may stem both from contextual differences in how Montessori education is enacted and from limitations within the instruments. Although grounded in a shared philosophy, Montessori implementation is shaped by cultural, institutional, and historical conditions (Debs, 2023). Implementation does not occur in a vacuum; the influence of context on educational implementation is widely acknowledged in implementation research (Century & Cassata, 2016; Ryan et al., 2024). For example, historically, there were attempts to adapt the Montessori Method to the American way of life, which conflicted with more universalist and orthodox interpretations of Montessori's approach (Quarfood, 2023).

The Dutch Montessori context differs substantially from the American context in terms of educational structure, accessibility, and regulation (De Brouwer et al., 2024). For example, all Dutch Montessori schools are publicly funded and therefore must comply with national educational regulations. One consequence is that Dutch law does not permit children to enter school before age 4, preventing the formation of the 3–6 mixed-age group that characterizes Montessori Early Childhood settings in many countries.

In addition, interpretation and translation issues complicate the operationalization of Montessori principles. Montessori's original writings are subject to variations in translation and interpretation (Beatty, 2024; Gynther, 2023). The American TQMP relies primarily on English-language sources, encompassing organizational standards and secondary publications on Montessori education predominantly from the United States, which may bias the instrument. For instance, several items in the American TQMP are derived from secondary sources rather than Montessori's original work, potentially limiting cross-context validity. This reliance on specific sources may have strengthened the questionnaire's alignment with the American interpretations of Montessori principles, but could potentially negatively influence the validity of the American TQMP in other countries. For example, the item "Classroom books feature realistic stories" in the American Early Childhood TQMP, and the item "Children repeat Montessori science experiments" in the American Elementary TQMP, are not derived from Montessori's original writings but from organizational standards and secondary literature. As a result, the American, Italian, and Dutch versions of the TQMP differ in item content, factor structure, and dimensionality. Such variation is consistent with broader psychometric literature, which shows instruments often function differently across populations (Flora & Flake, 2017). Translation further complicates construct validity and comparability (Flake et al., 2022; Gierl & Khaliq,

Figure 1
Development and Validation Process of the D-TQMP-Rs



2001), making measurement nonequivalence likely in cross-context applications.

Taken together, these findings indicate the need for an instrument that measures universal Montessori principles while remaining sensitive to the contextual factors that shape their practical enactment.

Present Study

This study begins with an examination of the revised Dutch TQMPs to determine whether they function adequately as measurement instruments. In this study, psychometric validation is understood as a multistage process, involving the examination of content validity, expert panel review, pilot testing, and statistical evaluation of its factorial structure, reliability, and convergent and discriminant validity (Lim, 2024). Therefore, the first research question (RQ1) is this: To what extent do the revised D-TQMPs demonstrate robust psychometric properties, in terms of their factorial structure, reliability, and convergent and discriminant validity?

These analyses provide the foundation for the second aim of the study in which we describe to what extent Montessori principles are implemented in Dutch Montessori education. Therefore, the second research question (RQ2) of this study is this: To what extent are Montessori principles implemented in the Montessori practice in Dutch Montessori schools? Together, these two aims allow us to refine the D-TQMPs while also providing a descriptive overview of Montessori implementation in the Netherlands.

To guide the validation process, we followed a five-step approach to systematically address each aspect of validity (Lim, 2024). First, we conducted a critical examination of the previously identified factor structures in the D-TQMPs, drawing on an expert discussion and relevant Montessori literature. Second, based on this review, we developed additional items to improve the conceptual coverage of the factors. Third, an expert panel evaluated the revised instruments to establish face validity. Fourth, a small-scale pilot was conducted with both questionnaires to assess clarity and usability. Finally, in the fifth step, the revised questionnaires (called D-TQMP-Rs) underwent psychometric examination in terms of their factorial structure, reliability, and convergent and discriminant validity.

To ensure coherence, the present study combines the psychometric examination of the D-TQMPs with subsequent analyses of overall implementation levels (RQ2). Together, these components provide a comprehensive account of both the validity and practical application of the D-TQMP-Rs. Figure 1 provides a visual summary of the five-step development and validation process and the research questions.

Method

Participants

The Dutch Montessori Association invited all Early Childhood and Elementary teachers of 162 Dutch Montessori schools to participate in this study by email. The email included a link to the online questionnaire, and schools were asked to have their teachers complete the questionnaire. The Early Childhood questionnaire was completed by 120 Early Childhood (ages 3–6) teachers, and the Elementary questionnaire was completed by 292 teachers. Dutch Montessori schools integrate Early Childhood and Elementary education, providing education for children ages 4 to 12 within the same school. The total sample of Early Childhood and Elementary teachers originated from 91 schools, representing 56% of all Montessori schools in the Netherlands.

Participants in both the Early Childhood and the Elementary samples were adults with comparable age distributions, teaching experience, and Montessori

Table 1
Descriptives of the Participants

	Early Childhood Teachers	Elementary Teachers
N	120	289
Age range	23–66	20–69
Age *M (SD)*	44.53 (12.44)	43.72 (11.76)
> 5 years Montessori teaching experience	65.0%	67.1%
Montessori teacher training completed / attending / none	74.2% / 17.5% / 8.3%	76.1% / 14.5% / 9.3%
Class size 15–20 / 21–25 / 26–30 / ≥ 31	30.8% /50.8% / 17.5% / 0.8%	9.0% / 39.4% / 42.6% / 9.0%

training backgrounds. Most Early Childhood teachers worked with class sizes common in Dutch primary education, whereas Elementary teachers generally taught larger groups. Table 1 provides a detailed overview of all participant characteristics.

Instruments

This study builds on previous work by De Brouwer et al. (2024, 2025), who developed the Early Childhood and Elementary versions of the Dutch Teacher Questionnaires for Montessori Practice (D-TQMPs). Both versions of the D-TQMP required further development and refinement, including clarification of the factors and its structure, and the addition of newly developed items to improve its validity and reliability. These refinements resulted in two revised instruments, which are referred to throughout this study as the D-TQMP-Rs.

Step 1: Review of Factor Structure, Original D-TQMP

Since previous studies on the Early Childhood version of the D-TQMP (De Brouwer et al., 2024) indicated that certain factors were conceptually ambiguous, we subsequently engaged a panel of four Montessori experts to reassess the factor structure and item content of the original D-TQMP, during two online meetings. Their discussion, supported by additional Montessori literature, showed that the Early Childhood factors *Teacher Guidance* and *Curriculum* were too broad. These were therefore reconceptualized into three clearer factors: *Prepared Environment*, *Teaching with Materials*, and *Practical Life Exercises*, resulting in a four-factor structure that also includes the original factor *Children's Freedom*. A definition of these factors is essential to establish content validity (Hardesty & Bearden, 2004). Therefore, *Children's Freedom* is defined as the extent to which children have the freedom to make their own choices (e.g., Montessori, 1935, 1937, 1997). The factor *Prepared Environment* is characterized by the dynamic relationship between the physical space and the social context, which in Montessori education are both intentionally designed to foster autonomy and support the child's development (Ahlquist, 2023; Lillard & McHugh, 2019; Montessori, 2016). The factor *Teaching with Materials* refers to how children and teachers interact with Montessori materials. Teachers adapt the materials to meet the individual needs of each child, provide guidance on their use, and promote independent learning and self-correction through these materials (Montessori, 1937, 2016). And the factor *Practical Life Exercises* pertains to encouraging children to engage in daily life activities that promote independence, develop motor skills, and foster self-care as well as responsibility for the care of their environment (Montessori, 1949, 2016, 2020b).

The panel also discussed and reviewed the proposed five-factor solution of the Elementary version of the D-TQMP, comprising the factors *Prepared Environment*, *Extended Environment*, *Teacher Guidance*, *Cosmic Education*, and *Self-Education*, and decided to retain this structure. This decision was consistent with the psychometric findings of the earlier validation study (De Brouwer et al., 2025), which also supported the interpretability and coherence of this five-factor structure. The factor *Prepared Environment* is characterized by the interplay between the physical environment, the social environment, and an autonomy-supportive setting within the classroom (Ahlquist, 2023; Lillard & McHugh, 2019; Montessori, 2016). The factor *Extended Environment* is typically outside the classroom, where children explore the world and society, and learn to independently care for themselves and their environment (Montessori, 1973). This includes field trips, which Montessori (1973) referred to as "going out." The factor *Teacher Guidance* is defined as the way the teacher tailors their guidance based

on observations, provides lessons and demonstrations, and is intentionally in the background, allowing children to take an active role (Montessori, 2016). Additionally, the teacher ensures uninterrupted and focused work by refraining from interrupting the children (Montessori, 2020b). The factor *Cosmic Education* helps children understand the complexity of the world and their unique place within it. It involves the investigation and acquisition of knowledge about the world and the universe they live in (Grazzini, 2013). In addition, *Cosmic Education* invites children to develop an ethic of responsibility in their daily behavior and fosters a sense of community, interdependence, and social cohesion. It brings together ecological education, peace education, and education for the world (Raimondo, 2023). Finally, the factor *Self-Education* is defined as the way the children learn to make independent choices, work with self-discipline, and develop the ability to assess their own work and make corrections when necessary (Montessori, 1973; 2020b; 2022).

Step 2: Additional Item Development

After reviewing the factor structure and establishing factor definitions, items related to these definitions were reevaluated. Newly formulated items were reviewed for their clarity, alignment with the definitions, and sentence structure, and modified as necessary (Rubio et al., 2003). Original items were revised as needed, and additional items were developed based on expert input as well as Dutch and international Montessori literature. Reexamining the factor definitions in depth led to a substantial number of revisions, and resulted in the addition of 15 new items to the Early Childhood version and 22 new items to the Elementary version, ensuring more complete coverage of each construct (see Table 2 and Table 3).

Step 3: Expert Panel Review

To assess content validity of the revised questionnaires, an expert panel evaluated the instruments to obtain initial impressions and an intuitive understanding of whether the improved questionnaires adequately and appropriately covered the conceptual and operational factors with their corresponding items (Lim, 2024). The expert panel consisted of three content experts who each had more than five years of experience as a Montessori teacher trainer at a University of Applied Sciences, and three field experts who each had more than five years of experience as a teacher or principal in a Montessori school accredited by the Dutch Montessori Association. Experts first categorized each item individually according to the predefined factor definitions, and afterward the results were discussed collectively during a panel meeting. This procedure offered a preliminary indication of the factorial structure, which was particularly relevant given the revised and newly developed items (Rubio et al., 2003). A 60% agreement rule guided decisions to retain, revise, or delete items, or reassign to a different factor, as suggested by Hardesty and Bearden (2004).

In the Early Childhood version, four items failed to meet the 60% decision rule. After the expert panel discussed the items, three items were revised by the research team for clarity in wording and retained within the intended factor, and one item was deleted. Based on panelist feedback, the four-factor structure in the Early Childhood version could be maintained.

In the Elementary version, seven items failed to meet the 60% decision rule. After the experts discussed the items, three items were revised by the research team for clarity in wording and retained within the intended factor, and two items were deleted. Two items were adapted and reassigned to another factor, as the panel indicated these items would better represent this factor. Based on panelist feedback, the proposed five-factor structure in the Elementary version could be maintained.

Step 4: Small-Scale Pilot

After digitizing the improved questionnaires into data management software Qualtrics, five Montessori teachers—from a conveniently accessible school with which we had an established collaboration—volunteered to pilot both online versions to estimate the time required to complete them, assess the smoothness of the digitization, and evaluate the clarity of the items (Ahmed, 2024). The pilot indicated the wording of individual items was familiar and clear to the teachers, so no changes were necessary. The teachers completed the questionnaires in approximately 15 minutes, and the online version functioned correctly and required no further adjustments.

The improved D-TQMPs are called the Dutch Teacher Questionnaires for Montessori Practice Revised (D-TQMP-Rs). The Early Childhood D-TQMP-R consisted of 26 items, and the Elementary D-TQMP-R consisted of 36 items. Teacher responses were measured using a 4-point Likert-type scale, ranging from 1 (*never occurs in my classroom*) to 4 (*always occurs in my classroom*). Table 2 and Table 3 provide overviews of the previous Dutch version of the TQMPs and the D-TQMP-Rs after establishing content validity, the expert panel's feedback, and the pilot.

Table 2

Overview of the Original Items of the Early Childhood D-TQMP and the Improved Items after Step 4

Code	Original Items, D-TQMP (2024)	Improved and Additional Items, D-TQMP-R
Children's Freedom		
CF_01	Determine how long to work with an activity.	Children determine how long to work with an activity.
CF_02	Decide where they will work.	Children decide where they will work.
CF_03	May choose to work alone or with others.	Children may choose to work alone or with others.
CF_04	Choose their work/activities.	Children choose their own work/activities.
CF_05		Children who are concentrated are given the opportunity to work uninterrupted.
Prepared Environment		
PE_01	Children's activities are recorded each day.	The teacher keeps daily records of individual children's developmental milestones.
PE_02	A full set of Montessori materials is available.	Children have access to a full set of Montessori materials.
PE_03	Observation is used for daily lesson planning.	The teacher guides children based on observation.
PE_04		There is an uninterrupted work period of 2.5 to 3 hours.
PE_05		Children work in a mixed-age group consisting of three different age levels.
PE_06		Children help each other across groups.
PE_07		Only one copy of all Montessori materials is available in the environment.
PE_08		The environment is tailored to the developmental needs of the children.
Teaching with Materials		
TM_01		Mathematics is taught using Montessori materials.
TM_02		Biology and geography are taught using Montessori materials.
TM_03		Sensory development is stimulated using Montessori materials.
TM_04		Language development is stimulated by the use of Montessori materials.
TM_05		The teacher tailors his lessons to the child's individual needs.
TM_06		The teacher presents materials following a three-period lesson.
Practical Life Exercises		
PL_01	Children carry out Practical Life exercises during the work period.	Children carry out Practical Life exercises during the work period.
PL_02	Garden in a designated area.	Children maintain their own garden.
PL_03		Children practice taking care of themselves.
PL_04	Care for classroom plants. Children regularly prepare food.	Children care for classroom plants.
PL_05		Children practice grace and courtesy.
PL_06	Walk on the line carrying objects.	Children practice walking the line exercises.
PL_07		Children practice silence lessons.

Table 3

Overview of the Original Items of the Elementary D-TQMP and the Improved Items after Step 4

Code	Original Items, D-TQMP (2025)	Improved and Additional Items, D-TQMP-R
Prepared Environment		
PE_01	Children have access to a full set of Montessori materials.	Children have access to a full set of Montessori materials.
PE_02	Lessons are given to individuals. Lessons are given with Montessori materials.	Children are given the opportunity to work at their own level.
PE_03	Children choose their own work or activities.	Children are given the opportunity to choose their own work or activities.
PE_04	Children are in a group consisting of three grade levels.	Children work in a mixed-age group consisting of three different age levels.
PE_05	Children determine how long to work with an activity.	The environment gives children the opportunity to work on an activity uninterrupted.
PE_06	Children decide where they will work.	Children are given the opportunity to choose where they want to work.
PE_07		Children are given the opportunity to work together in small, mixed-age groups.
PE_08		The environment is tailored to the developmental needs of the children.
PE_09		The environment provides children with the opportunity to work uninterrupted for 2.5 to 3 hours.
Extended Environment		
EE_01	Children take part in community service projects.	Children take part in community service projects.
EE_02	Children conduct field trips in small groups.	Children conduct "going out" excursions.
EE_03		Children are prepared for "going out" excursions.
EE_04		Children learn to maintain the school environment throughout the day.
EE_05		Children take "intellectual walks" through the school.
Teacher Guidance		
	Children decide when to take their own fruit break.	
TG_01	Every other week an individual child-teacher meeting takes place. There is a 3-hour uninterrupted work period.	The teacher engages in reflective conversations with individual children.
TG_02		The teacher does not interrupt children when they are working with concentration.
TG_03		The teacher offers Montessori materials, allowing children to work with them independently.
TG_04		The teacher is present in the background.
TG_05		The teacher decides to guide children only after observation.
TG_06		The teacher tailors his lessons to the child's individual needs.
TG_07		The teacher only guides children who are not working with concentration.

Code	Original Items, D-TQMP (2025)	Improved and Additional Items, D-TQMP-R
Cosmic Education		
CE_01	Four or more cosmic subjects are covered each year.	A minimum of four subjects are covered in cosmic education each year.
CE_02	Children learn about the interconnection between nature, culture, time, and space.	In cosmic education, children learn about the interconnection between nature, culture, time, and space.
CE_03		Children also work on cosmic education during the 3-hour work period.
CE_04		Children develop timelines, cycles, and models in cosmic education.
CE_05		In cosmic education stories are told.
CE_06		In cosmic education real-life materials are used.
CE_07		Telling stories in cosmic education leads to children conducting their own research.
CE_08		In cosmic education, attention is given to societal issues.
Self-Education		
SE_01	Children correct their own work.	Children learn to correct their own work.
SE_02	Children record activities in planning journals.	Children learn to maintain an overview of their own activities.
SE_03		Children learn to work with concentration.
SE_04		Children learn to complete their work independently.
SE_05		Children learn to decide which activities they want to work on.
SE_06		Children learn to take the initiative to start working.
SE_07		Children learn to determine how long they will work on an activity.

Step 5: Analyses of Psychometric Properties

The primary aim of this study was to examine the psychometric properties of the D-TQMP-Rs. Therefore, in this fifth step we examined the D-TQMP-Rs psychometric properties, in terms of their factorial structures, reliability, and convergent and discriminant validity (RQ1). After this psychometric evaluation, we explored implementation levels across Dutch Montessori schools (RQ2).

Procedure

The D-TQMP-Rs were administered digitally using Qualtrics, starting with a brief introduction about the study's aim. The questionnaires were administered between October 2023 and January 2024 after receiving active informed consent by the participants. This procedure complies with the standards set by the ethical commission of Saxion University of Applied Sciences (den Ouden et al., n.d.). Questions could not be skipped, to avoid omitted data.

RQ1: Analyses of Psychometric Examination

Because this study aims to evaluate the validity of the revised D-TQMPs, the fifth step involved examining their psychometric quality. This included assessing how well the proposed factor structure and the underlying items fit the observed data. Fit indices provide indicators of model adequacy, helping to determine whether the overall structure of the instruments is acceptable, needs refinement, or is inconsistent with the data.

Since the D-TQMPs were already subject of previous studies, a confirmatory factor analysis (CFA) was employed to evaluate how well our collected data fitted into the prespecified revised factor structures (Brown, 2015). All analyses were conducted in data analysis software R (version RStudio 2025.09.0+387). To assess the found factor structures, we examined the following: root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker–Lewis index (TLI), standardized root mean square residual (SRMR), and chi-square (χ^2) (Prudon, 2015). The data were analyzed using

diagonally weighted least squares (DWLS) estimation, given the ordinal 4-point Likert-type response format (Xia & Yang, 2019). To indicate a good fit, RMSEA should be below 0.08, CFI and TLI values should be close to 0.95 (Lee & Cham, 2024), and SRMR should be below .08 (Guenther et al., 2023). However, these guidelines are often relaxed (RMSEA ≤ 0.08, CFI ≥ 0.90, TLI ≥ 0.90) to indicate moderate fit (Nye, 2023). The χ^2 test should be nonsignificant ($p > .05$) (Brown, 2015). Xia and Yang (2019) demonstrated that using the DWLS estimator tends to produce inflated RMSEA, CFI, and TLI values, suggesting a better model fit as compared to other estimators, and potentially resulting in a misspecified model. Therefore, Shi and Maydeu-Olivares (2020) suggest using SRMR, which is generally robust to the choice of estimation methods to evaluate model fit. In addition, considering the relatively small sample sizes, CFI is comparatively robust to sample-size effects (Sun, 2005). Following these recommendations, SRMR and scaled CFI were used as the primary indicators of overall model fit, and scaled RMSEA, TLI, and χ^2 were used as diagnostic tools for model refinement. Standardized factor loadings, item content, and modification indices were examined to ascertain whether the CFA factor structures could be fine-tuned. Fine-tuning was conducted only when statistical indicators and theoretical considerations aligned (Nye, 2023).

Convergent and Discriminant Validity and Reliability

For both revised versions of the D-TQMP, convergent validity, discriminant validity, and reliability were examined. Convergent validity, which is the internal consistency of the items measuring the same construct, is assessed by the average variance extracted (AVE), calculating the composite reliability (CR), and examining the factor loadings. Following recommendations of Cheung et al. (2024), we examined not only the estimates of these indicators but also their 90% bootstrap confidence intervals (CIs) to account for sampling error. To demonstrate convergent validity, the upper bound of the 90% CI of AVE should be greater than or equal to .50, the upper bound of the 90% CI of CR should exceed .70, and the upper bound of the 90% CI of each standardized factor loading should be greater than or equal to .50. If the upper bound of the CI for any loading falls below .50, the construct is considered not to exhibit convergent validity.

Discriminant validity, which ensures a construct is empirically distinct, means each factor measures only its own concept and not others (Lim, 2024). For example, the factor *Prepared Environment* should not inadvertently measure aspects of the factor *Teacher Guidance*. Discriminant validity was assessed using the heterotrait–monotrait ratio of correlations (HTMT) (Henseler et al., 2015). Following recommendations by Franke and Sarstedt (2019), we evaluated HTMT using the upper bound of the 90% bootstrap CIs to avoid false positive conclusions. Discriminant validity is established when the upper bound of the 90% bootstrap CI falls below the 0.90 threshold (Guenther et al., 2023).

Reliability was evaluated by computing McDonald's ω for each scale, as it accounts for unequal factor loadings and therefore is more realistic than Cronbach's alpha for estimating reliability (Dunn et al., 2014). Values greater than .70 were considered as adequate, indicating satisfactory internal consistency (Hair et al., 2020).

RQ2: Exploring Montessori Implementation

Finally, we explored how Montessori principles are implemented in Dutch Montessori schools. Therefore, descriptive statistics regarding the implementation of Montessori education were analyzed. Mean scores and standard deviations were computed for each factor in the final structure, based on the sum of a factor's item scores divided by the number of items within that factor. Mean scores could range between 1 and 4. These descriptives provide a general picture of implementation levels in Dutch Montessori schools.

Results

The results are organized by research question. First, we present the results of the psychometric examination of the Early Childhood and Elementary D-TQMP-R, corresponding to the fifth step of the development and validation process (RQ1). Because both questionnaires were validated separately, results for the Early Childhood and Elementary D-TQMP-R are presented independently. Second, we summarize descriptive implementation levels to provide contextual insight into Montessori implementation in the Netherlands (RQ2).

Before starting the analysis, preliminary data checks were performed. Standard deviations and ranges of responses of all data were computed. Standard deviations for the Early Childhood data ranged between 0.36 and 1.14, and for 11 items not all response options were used at least once. Standard deviations for the Elementary data ranged between 0.54 and 1.12, and for six items not all response options were used at least once. There was no missing data.

RQ1: Results of Psychometric Properties, Early Childhood D-TQMP-R

We evaluated the factorial structure, reliability, and convergent and discriminant validity of the Early Childhood D-TQMP-R. Results from the CFA indicated a moderate to good model fit (CFI = 0.93; SRMR = .08), with additional fit indices also supporting this overall good fit (RMSEA = 0.03; TLI = 0.92; $\chi^2(293) = 321.95$, $p = 0.12$). Indeed, χ^2 is nonsignificant; however, it tends to be influenced by the DWLS estimator (Nye, 2023).

The standardized factor loadings of the 26 items of the five factors ranged from .11 to .69. Studying the factor loadings, modification indices, and the item content resulted in omitting the items PE_04 (see Table 3) and PL_4 (see Table 4). The final version of the Early Childhood D-TQMP-R consists of a four-factor structure, containing 24 items (CFI = 0.93; SRMR = .08; RMSEA = 0.03; TLI = 0.92; $\chi^2(246) = 274.71$, $p = .10$). These results indicate the Early Childhood D-TQMP-R demonstrates an acceptable overall model fit but that several items load only weakly on their intended factors.

The factor *Children's Freedom* consists of five items (McDonald's $\omega = .65$), the factor *Prepared Environment* contains seven items (McDonald's $\omega = .61$), the factor *Teaching with Materials* contains six items (McDonald's $\omega = .74$), and the factor *Practical Life Exercises* contains six items (McDonald's $\omega = .61$). Table 4 shows the factor loadings of the final version. Taken together, these reliability results suggest internal consistency is acceptable for some factors and weaker for others, indicating several item sets may benefit from further refinement.

To evaluate convergent validity, AVE, CR, and standardized factor loadings were examined. None of the factors exceeded the AVE threshold of .50, and two factors exceeded the CR threshold of .70. Examination of the standardized factor loadings showed that 21 of the 24 items had an upper 90% CI limit of .50 or higher, but three items did not meet this requirement. Taken together, these results indicate that although two factors showed acceptable composite reliability and most items demonstrated at least moderate loadings, the overall pattern did not meet the combined criteria required for convergent validity for all factors, meaning the items within each factor do not yet align strongly enough to be considered coherent measures of a single construct.

To examine discriminant validity, HTMT values were computed using 1,000 percentile-bootstrap replications. HTMT values ranged from 0.63 to 0.95, with the upper bounds of the 90% CIs ranging from 0.83 to 1.04. Using the 0.90 threshold, only two factor pairs (*Children's Freedom–Teaching with Materials*; *Children's Freedom–Practical Life Exercises*) met the criterion, whereas all other factor combinations exceeded this threshold. These results indicate that several factors overlap substantially in what they measure, meaning the early-childhood D-TQMP-R does not yet distinguish clearly between its factors. Based on these results, evidence for sufficient discriminant validity was not found. Table 5 displays the descriptive statistics, AVE, CR, ω, and HTMT for the final factor structure.

RQ1: Results of Psychometric Properties, Elementary D-TQMP-R

Results from the CFA indicated an acceptable model fit (CFI = 0.93; SRMR = .07). Additional indices also supported this conclusion (RMSEA = 0.05; TLI = 0.92; $\chi^2(584) = 942.37$, $p < .01$). Standardized factor loadings ranged from .30 to .86. Studying the factor loadings, modification indices, and item content resulted in omitting TG_07 (see Table 3). The final version of the Elementary D-TQMP-R consists of a five-factor structure, containing 35 items (CFI = 0.93; SRMR = 0.07; RMSEA = 0.05; TLI = 0.92; $\chi^2(550) = 907.89$, $p = .00$). These results indicate the Elementary D-TQMP-R demonstrates an acceptable overall model fit but that several items load only weakly on their intended factors.

The factor *Prepared Environment* consists of nine items (McDonald's $\omega = .81$), the factor *Extended Environment* consists of five items (McDonald's $\omega = .71$), the factor *Teacher Guidance* consists of six items (McDonald's $\omega = .78$), the factor *Cosmic Education* consists of eight items (McDonald's $\omega = .86$), and the factor *Self-education* consists of seven items (McDonald's $\omega = .85$). Table 6 presents the full item wording together with each item's final factor assignment and its corresponding standardized factor loading.

AVE, CR, and standardized factor loadings were examined to evaluate convergent validity. The factors *Cosmic Education* and *Self-Education* exceeded the AVE (90% CI upper bounds) threshold of .50, and all factors exceeded the CR threshold of .70. All standardized factor loadings were statistically significant, and 34 had an upper 90% CI limit of .50 or higher, although one item did not meet this requirement. Standardized factor loadings ranged from .35 to .86. Taken together, these findings indicate that the factors *Cosmic Education* and *Self-Education* show sufficient internal coherence

Table 4

CFA Solution for the Early Childhood D-TQMP-R with Standardized Factor Loadings with 90% CI per Item

Code	Item	CF	PE	TM	PL
CF_01	Children determine how long to work with an activity.	.40 (.18 – .60)			
CF_02	Children decide where they will work.	.58 (.42 – .72)			
CF_03	Children may choose to work alone or with others.	.43 (.19 – .58)			
CF_04	Children choose their own work/activities.	.44 (.26 – .60)			
CF_05	Children who are concentrated are given the opportunity to work uninterrupted.	.69 (.55 – .81)			
PE_01	The teacher keeps daily records of individual children's developmental milestones.		.37 (.21 – .53)		
PE_02	Children have access to a full set of Montessori materials.		.48 (.30 – .64)		
PE_03	The teacher guides children based on observation.		.60 (.48 – .72)		
PE_04	There is an uninterrupted work period of 2.5 to 3 hours.		0.22 (.02 – .40)		
PE_05	Children work in a mixed-age group consisting of three different age levels.		.47 (.30 – .61)		
PE_06	Children help each other across groups.		.46 (.32 – .58)		
PE_07	Only one copy of all Montessori materials is available in the environment.		.62 (.46 – .76)		
TM_01	Mathematics is taught using Montessori materials.			.43 (.22 - .58)	
TM_02	Biology and geography are taught using Montessori materials.			.66 (.52 – .77)	
TM_03	Sensory development is stimulated using Montessori materials.			.56 (.30 – .72)	
TM_04	Language development is stimulated by the use of Montessori materials.			.56 (.43 – .69)	
TM_05	The teacher tailors his lessons to the child's individual needs.			.56 (.39 – .73)	
TM_06	The teacher presents materials following a three-period lesson.			.59 (.44 – .73)	
PL_01	Children carry out practical life exercises during the work period.				.55 (.41 – .68)
PL_02	Children maintain their own garden.				.32 (.16 – .46)
PL_03	Children practice taking care of themselves.				.55 (.40 – .69)

Code	Item	CF	PE	TM	PL
PL_05	Children practice grace and courtesy.				.55 (.38 – .69)
PL_06	Children practice walking the line exercises.				.59 (.43 – .72)
PL_07	Children practice silence lessons.				.24 (.04 - .43)

Note. CF = *Children's Freedom*, PE = *Prepared Environment*, TM = *Teaching with Materials*, PL = *Practical Life Exercises*. Confirmatory factor analysis (CFA) was conducted with R 2025.09.0. Values represent standardized factor loadings. Ninety-percent CIs are bootstrap-based.

Table 5

Descriptives of the Final Factors of the Early Childhood D-TQMP-R, AVE, CR, ω, and HTMT

						HTMT (upper bound 90% CI)			
	M	*SD*	AVE (90% CI)	CR (90% CI)	*ω*	CF	PE	TM	PL
CF	3.59	.36	.30 (.21–.40)	.63 (.46–.74)	.65				
PE	3.28	.45	.19 (.15–.25)	.62 (.52–.69)	.61	0.85 (0.99)			
TM	3.49	.44	.34 (.27–.43)	.73 (.63–.78)	.74	0.72 (0.90)	0.95 (1.04)		
PL	2.87	.49	.22 (.16–.31)	.57 (.44–.68)	.61	0.64 (0.83)	0.84 (1.01)	0.83 (.0.98)	

Note. CF = *Children's Freedom*, PE = *Prepared Environment*, TM = *Teaching with Materials*, PL = *Practical Life Exercises*. Subscale means are based on the average of items in the final factor structure. Mean scores could range between 1 and 4.

to be considered convergently valid, but evidence for sufficient convergent validity is not found for the factors *Prepared Environment*, *Extended Environment*, and *Teacher Guidance*. These factors require further item-level refinement before they can be treated as reliable measures of their underlying factors.

Regarding discriminant validity, HTMT values were computed with 1,000 percentile-bootstrap replications. HTMT values ranged from 0.53 to 0.94, with the upper bounds of the 90% CIs ranging from 0.62 to 1.00. Using the 0.90 threshold, three factor pairs (*Teacher Guidance–Self-Education*, *Prepared Environment–Teacher Guidance*, *Prepared Environment–Self-Education*) exceeded the threshold, indicating insufficient discriminant validity, whereas all other factor pairs were below the 0.90 threshold. These results suggest that some conceptual overlap exists among the factors *Teacher Guidance*, *Prepared Environment*, and *Self-Education*. Table 7 displays the descriptive statistics, AVE, CR, ω, and HTMT for the final factor structure.

RQ2: Results Exploring Dutch Montessori Implementation

The means and standard deviations of the factors were computed to gain some insight into the extent Dutch teachers implement Montessori principles in their classrooms (see Table 5 and Table 7). For Early Childhood teachers, mean scores were relatively high (2.87–3.59), but the variability around these means indicates differences in how teachers implement Montessori principles. A similar pattern was found in Elementary classrooms, where means ranged from 2.74 to 3.26 and again showed clear variation between teachers.

Discussion

Recent meta-analyses on the effectiveness of Montessori education report positive outcomes on both academic and social measures (Demangeon et al., 2023; Randolph et al., 2023). However, Montessori schools vary considerably in their implementation of Montessori principles, and there is limited empirical literature examining how differences in implementation affect outcomes. To better understand Montessori education's effectiveness, there is a need to study its implementation. Instruments that demonstrate evidence of validity and reliability are needed to assess Montessori implementation; however, the field is still young in its understanding. This study contributes to a better understanding of how to assess and examine Montessori implementation. Therefore, the aim of this study was twofold: (1) to improve and refine both versions of the revised D-TQMPs by gathering new validity evidence, and (2) to describe the extent to which Montessori principles are implemented in Dutch Montessori schools.

The first research question of this study examined to what extent the revised D-TQMPs demonstrated psychometric properties, including their factorial

Table 6

CFA Solution for the Elementary D-TQMP-R with Standardized Factor Loadings with 90% CI per Item

Code	Item	PE	EE	TG	CE	SE
PE_01	Children have access to a full set of Montessori materials.	.57 (.51–.66)				
PE_02	Children are given the opportunity to work at their own level.	..68 (.60–.75)				
PE_03	Children are given the opportunity to choose their own work or activities.	.73 (.65–.80)				
PE_04	Children work in a mixed-age group consisting of three different age levels.	.35 (.23–.47)				
PE_05	The environment gives children the opportunity to work on an activity uninterrupted.	.57 (.47–.66)				
PE_06	Children are given the opportunity to choose where they want to work.	.48 (.39–.57)				
PE_07	Children are given the opportunity to work together in small, mixed-age groups.	.52 (.43–.62)				
PE_08	The environment is tailored to the developmental needs of the children.	.72 (.66–.79)				
PE_09	The environment provides children with the opportunity to work uninterrupted for 2.5 to 3 hours.	.40 (.30–.50)				
EE_01	Children take part in community service projects.		.53 (.40–.63)			
EE_02	Children conduct "going out" excursions.		..57 (.45–.69)			
EE_03	Children are prepared for "going out" excursions.		61 (.50–.71)			
EE_04	Children learn to maintain the school environment throughout the day.		.67 (.55–.77)			
EE_05	Children take "intellectual walks" through the school.		.48 (.37–.58)			
TG_01	The teacher engages in reflective conversations with individual children.			.53 (.43–.63)		
TG_02	The teacher does not interrupt children when they are working with concentration.			.50 (.40–.60)		
TG_03	The teacher offers Montessori materials, allowing children to work with them independently.			.71 (.65–.78)		
TG_04	The teacher is present in the background.			.59 (.50–.68)		
TG_05	The teacher decides to guide children only after observation.			.55 (.46–.64)		
TG_06	The teacher tailors his lessons to the child's individual needs.			.74 (.68–.81)		
CE_01	A minimum of four subjects are covered in cosmic education each year.				.59 (.47–.70)	

Code	Item	PE	EE	TG	CE	SE
CE_02	In cosmic education, children learn about the interconnection between nature, culture, time, and space.				.69 (.61–.77)	
CE_03	Children also work on cosmic education during the 3-hour work period.				.67 (.59–.76)	
CE_04	Children develop timelines, cycles, and models in cosmic education.				.70 (.61–.77)	
CE_05	In cosmic education stories are told.				.61 (.52–.70)	
CE_06	In cosmic education real-life materials are used.				.66 (.57–.74)	
CE_07	Telling stories in cosmic education leads to children conducting their own research.				.69 (.61–.77)	
CE_08	In cosmic education, attention is given to societal issues.				.67 (.57–.75)	
SE_01	Children learn to correct their own work.					.52 (.41–.62)
SE_02	Children learn to maintain an overview of their own activities.					.55 (.46–.65)
SE_03	Children learn to work with concentration.					.71 (.62–.79)
SE_04	Children learn to complete their work independently.					..61 (.52–.72)
SE_05	Children learn to decide which activities they want to work on.					.86 (.80–.91)
SE_06	Children learn to take the initiative to start working.					.76 (.69–.83)
SE_07	Children learn to determine how long they will work on an activity.					.68 (.61–.75)

Note. PE = *Prepared Environment*, EE = *Extended Environment*, TG = *Teacher Guidance*, CE = *Cosmic Education*, SE = *Self-Education*. Confirmatory factor analysis (CFA) was conducted with R 2025.09.0. Values represent standardized factor loadings. Ninety-percent CIs are bootstrap-based.

Table 7

Descriptives of the Final Factors of the Elementary D-TQMP-R, AVE, CR, ω, and HTMT

						HTMT (upper bound 90% CI)				
	M	*SD*	AVE (90% CI)	CR (90% CI)	*ω*	PE	EE	TG	CE	SE
PE	3.23	.44	.33 (.2–.38)	.81 (.77–.84)	.81					
EE	2.74	.51	.33 (.27–.40)	.71 (.64–.76)	.71	0.75 (0.85)				
TG	3.10	.45	.37 (.33–.43)	.78 (.74–.81)	.78	0.94 (1.00)	0.74 (0.85)			
CE	2.97	.53	.44 (.38–.50)	.86 (.83–.89)	.86	0.67 (0.77)	0.73 (0.83)	0.62 (0.74)		
SE	3.26	.44	.46 (.4–.51)	.85 (.83–.88)	.85	0.88 (0.94)	0.61 (0.74)	0.85 (0.92)	0.53 (0.62)	

Note. PE = *Prepared Environment*, EE = *Extended Environment*, TG = *Teacher Guidance*, CE = *Cosmic Education*, SE = *Self-Education*. Subscale means are based on the average of items in the final factor structure. Mean scores could range between 1 and 4.

structures, reliability, and convergent and discriminant validity. The final version of the Early Childhood D-TQMP-R demonstrated a good model fit with a four-factor structure—*Children's Freedom*, *Prepared Environment*, *Teaching with Materials*, and *Practical Life Exercises*—comprising 24 items. The Elementary version showed a good model fit with a five-factor structure—*Prepared Environment*, *Extended Environment*, *Teacher Guidance*, *Cosmic Education*, and *Self-Education*—consisting of 35 items. The Early Childhood version demonstrated lower reliability for three of the four factors, whereas the Elementary version showed good reliability of all the factors, as indicated by ω coefficients. However, further analysis revealed that neither version of the revised D-TQMP-R established convergent and discriminant validity for all factors. Although further refinement and additional validation remain necessary, both versions of the D-TQMP-R provide stronger validity and reliability evidence than the earlier TQMPs and D-TQMPs. This means that, at their current stage, the instruments are suitable for reflective use in practice, such as lesson observations, schoolwide discussions about Montessori implementation, and formative self-evaluation, but the instruments are not yet robust enough for research purposes or for high-stakes evaluative decisions.

Taken together, the present findings suggest that the D-TQMP-Rs represent an important improvement over earlier instruments as they demonstrate coherent factor structures, theoretically meaningful subscales, and acceptable internal consistency. At the same time, limitations in discriminant and convergent validity reveal the constructs are not yet empirically distinct and that some items do not load strongly enough on their intended factors. However, considering that Montessori education is a complex phenomenon, historically layered, and culturally interpreted, it is clear that Montessori education does not lend itself easily to generalization (Beatty, 2024). The combination of cultural adaptation and interpretive variation complicates the validity of developing instruments for assessing Montessori implementation. The D-TQMP-Rs should therefore be considered promising yet still under development, requiring further refinement before robust and generalizable evidence of reliability and validity can be established.

The second research question of this study was to explore to which extent Dutch teachers in Early Childhood and Elementary schools implement Montessori principles. The relatively high mean scores suggest these principles are implemented at generally high levels, although variability exists between teachers. However, the results of this second research question need to be interpreted with caution, as the D-TQMP-Rs still require further improvements in psychometric quality. Nevertheless, as suggested, the instruments can currently be used for reflective purposes in practice.

Why differences among teachers exist is subject to debate. Some argue that more experienced teachers and well-trained teachers implement Montessori principles with greater sensitivity and authority, leading to higher levels of fidelity (Randolph et al., 2023). This argument is consistent with research in other domains that suggests skillful practitioners implement interventions or programs on a higher level (Mowbray et al., 2003). Class sizes could be another potential moderator for implementation differences, as Montessori (2016) argued that large class sizes with up to 40 children are found to be optimal for implementing her pedagogy. For example, Chen (2021) found that lower class sizes in Montessori Early Childhood in China led to lower levels of implementation fidelity. Future studies employing a more robustly validated version of the D-TQMP-Rs will be required in order to ascertain whether the observed implementation differences represent true variation in practice and to clarify which teacher- and classroom-level factors contribute most to Montessori implementation.

Directions for Future Research

Although work has been done on the validation of the D-TQMP-Rs, validation is an ongoing process that cannot be established through a single study. Current analyses provide an important step, but additional cross-validation and evidence are needed to classify the D-TQMP-Rs as a fully validated instrument.

Moreover, several limitations and directions for future research remain. First, relying on self-report items limits the understanding of what is happening in classrooms. Data triangulation, such as incorporating classroom observations to validate self-reported questionnaire responses, could address this limitation. Second, a more diverse and larger sample would enhance the robustness of the analyses. Ideally, larger sample sizes are better for CFA, as they improve the accuracy of parameter estimates (Worthington & Whittaker, 2006). However, the Dutch Montessori landscape is relatively small, with only 162 Early Childhood and Elementary schools, which restricts the possibility of obtaining a larger sample. This calls for other researchers to refine and revalidate the D-TQMP-Rs in different contexts with

a larger sample size. Third, although the questionnaire items are grounded in Montessori's original writings and may carry universal relevance, contextual influences could affect how the constructs are understood and measured. To address this limitation, providing clear, theory-based descriptions of each factor can help teachers across different contexts interpret the constructs more consistently. Until such refinements are fully established, the long-term applicability and cross-contextual validity of the D-TQMP-Rs remain open questions, and the instrument should be viewed as a foundation for further validation efforts within different cultural and educational settings.

In addition, rather than solely focusing on fidelity of implementation, researchers could also explore different patterns of practice and catalog the types of implementation in order to study their differences (Century & Cassata, 2016). The characterization of implementation types could offer an overarching view of Dutch Montessori implementation and provide a better understanding of how differences in implementation relate to particular outcomes. Furthermore, it would be interesting to study how variance in Montessori implementation is related to class/teacher level or the school level, and whether implementation differences could be explained by class/teacher variables or school variables.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Microsoft Copilot, an AI-powered digital assistant, to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Disclosure of interest

The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper. De Brouwer is professionally involved in Montessori education and research; this affiliation did not influence the design, analysis, or interpretation of the findings. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

Material and data are available upon request by sending an e-mail to the first author.

Author Information

Jaap de Brouwer is an educational researcher at the Progressive Education Research Group at Saxion University of Applied Sciences and Thomas More University of Applied Sciences. She can be reached at j.debrouwer@saxion.nl. https://orcid.org/0000-0002-8641-8306

Lida T. Klaver is an educational researcher at the Progressive Education Research Group at Saxion University of Applied Sciences. https://orcid.org/0000-0002-2994-8634

Patrick H. M. Sins is professor of Progressive Education at Thomas More University of Applied Sciences and professor of Learning at Rotterdam University of Applied Sciences. https://orcid.org/0000-0003-0443-6074

Symen van der Zee is professor of Progressive Education at the Progressive Education Research Group at Saxion University of Applied Sciences. https://orcid.org/0000-0003-3615-4397

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, *36*(3), 320–338. https://doi.org/10.1177/1098214014557009

Ahlquist, E. (2023). Learning in the Montessori school environment. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *The Bloomsbury handbook of Montessori education* (pp. 117–127). Bloomsbury Academic.

Ahmed, S. K. (2024). How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. *Oral Oncology Reports*, *12*, 100662. https://doi.org/10.1016/j.oor.2024.100662

Beatty, B. (2024). Angela K. Murray, Eva-Maria Tebano Ahlquist, Maria K. McKenna, and Mira Debs, eds. *The Bloomsbury handbook of Montessori education* London: Bloomsbury Academic, 2023. 585 pp. *History of Education Quarterly*, *64*(4), 550–553. https://doi.org/10.1017/heq.2024.39

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what,

how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

Chen, A. (2021). Exploration of implementation practices of Montessori education in mainland China. *Humanities and Social Sciences Communications*, *8*(1), 250. https://doi.org/10.1057/s41599-021-00934-3

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*(2), 745–783. https://doi.org/10.1007/s10490-023-09871-y

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

De Brouwer, J., Morssink-Santing, V., & van der Zee, S. (2024). Validation of the Teacher Questionnaire of Montessori Practice for Early Childhood in the Dutch context. *Journal of Montessori Research*, *10*(1). https://doi.org/10.17161/jomr.v10i1.21543

De Brouwer, J., Klaver, L. T., Sins, P. H. M., & van der Zee, S. (2025). The validation of the Dutch Teacher Questionnaire of Montessori Practice. *Journal of Research in Childhood Education*, 1–17. https://doi.org/10.1080/02568543.2025.2455669

Debs, M. (2023). Introduction: Global Montessori education. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *The Bloomsbury handbook of Montessori education* (pp. 283–289). Bloomsbury Academic.

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

Demangeon, A., Claudel-Valentin, S., Aubry, A., & Tazouti, Y. (2023). A meta-analysis of the effects of Montessori education on five fields of development and learning in preschool and school-age children. *Contemporary Educational Psychology*, *73*, 102182. https://doi.org/10.1016/j.cedpsych.2023.102182

Den Ouden, M. et al. (n.d.). Saxion University of Applied Sciences Ethical Commission. Retrieved October 15, 2023. https://srs.saxion.nl/ethischecommissie/

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), https://doi.org/10.1111/bjop.12046

Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, *77*(4), 576–588. https://doi.org/10.1037/amp0001006

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science*, *49*(2), 78–88. https://doi.org/10.1037/cbs0000069

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research*, *29*(3), 430–447. https://doi.org/10.1108/IntR-12-2017-0515

Gierl, M. J., & Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, *38*(2), 164–187. https://doi.org/10.1111/j.1745-3984.2001.tb01121.x

Grazzini, C. (2013). Maria Montessori's cosmic vision, cosmic plan, and cosmic education. *The NAMTA Journal*, *38*(1), 107–116. https://eric.ed.gov/?id=EJ1078117

Guenther, P., Guenther, M., Ringle, C. M., Zaefarian, G., & Cartwright, S. (2023). Improving PLS-SEM use for business marketing research. *Industrial Marketing Management*, *111*, 127–142. https://doi.org/10.1016/j.indmarman.2023.03.010

Guerrero, S., Valenciano-Valcárcel, J., & Rodríguez, A. (2024). Unveiling alternative schools: A systematic review of cognitive and social-emotional development in different educational approaches. *Children and Youth Services Review*, *158*, 107480. https://doi.org/10.1016/j.childyouth.2024.107480

Gynther, P. (2023). Societal responsibility and the child: The child, society and the world: Unpublished speeches and writings. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *The Bloomsbury handbook of Montessori education*. Bloomsbury Academic.

Hair, J. F., Jr., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, *109*, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: Implications for

improving face validity of measures of unobservable constructs. *Journal of Business Research, 57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Lee, H., & Cham, H. (2024). Comparing accuracy of parallel analysis and fit statistics for estimating the number of factors with ordered categorical data in exploratory factor analysis. *Educational and Psychological Measurement, 84*(6), 1173–1202. https://doi.org/10.1177/00131644241240435

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

Lim, W. M. (2024). A typology of validity: Content, face, convergent, discriminant, nomological and predictive validity. *Journal of Trade Science, 12*(3), 155–179. https://doi.org/10.1108/JTS-03-2024-0016

Marshall, C. (2017). Montessori education: A review of the evidence base. *npj Science of Learning, 2*(1), 11. https://doi.org/10.1038/s41539-017-0012-7

Montessori, M. (1933). *Het kind in de kerk* [*The child in the church*]. Paul Brand's Uitgeversbedrijf.

Montessori, M. (1935). *De Methode der wetenschappelijke pedagogiek* [*The Montessori Method*]. Van Holkema & Warendorf N.V. (Original work published 1914)

Montessori, M. (1937). *Het geheim van het kinderleven* [*The secret of childhood*]. Van Holkeman & Warendorf N.V.

Montessori, M. (1949). *The absorbent mind.* Theosophical Press.

Montessori, M., (1973). *From childhood to adolescence. Second edition.* Schocken Books. (Original work published 1948)

Montessori, M. (1997). *The California lectures of Maria Montessori, 1915.* Clio Press.

Montessori, M. (2016). *The child, society and the world.* Montessori-Pierson Publishing Company. (Original work published 1975)

Montessori, M. (2020a). *Door het kind naar een nieuwe Wereld.*[*Through the child into a new world*]. Montessori-Pierson Publishing Company. (Original work published 1941)

Montessori, M. (2020b). *Creative development in the child.* Montessori-Pierson Publishing Company. (Original work published 1998)

Montessori, M. (2022). *Zelfopvoeding deel 1* [*The Advanced Montessori Method*]. Montessori-Pierson Publishing Company. (Original work published 1916)

Montessori, M. (2023). *Van kind tot mens* [*The formation of man*]. Montessori-Pierson Publishing Company. (Original work published 1949)

Mook, D. G. (2016). In defense of external invalidity. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (4th ed., pp. 87–98). American Psychological Association.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation.* https://doi.org/10.1177/109821400302400303

Murray, A. K., Daoust, C. J., & Chen, J. (2019). Developing instruments to measure Montessori instructional practices. *Journal of Montessori Research, 5*(1), 75–87. https://doi.org/10.17161/jomr.v5i1.9797

Murray, A. K., & Daoust, C. J. (2023). Fidelity issues in Montessori research. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *The Bloomsbury handbook of Montessori education*. Bloomsbury Academic.

Murray, A. K., Murray, A. J., Daoust, C. J., & Gerker, H. E. (2025). Developing a tool to evaluate early childhood education implementation fidelity: Measure of Montessori Implementation pilot study results. *Frontiers in Education, 10*, 1603908. https://doi.org/10.3389/feduc.2025.1603908

Nunes, T., Stylianides, G. J., Lea, R., & Matthews, L. (2024). Replication in educational interventions: Developing a tool to measure and promote fidelity. *International Journal of Research & Method in Education, 48*(4), 402–423. https://doi.org/10.1080/1743727X.2024.2420336

Nye, C. D. (2023). Reviewer resources: Confirmatory factor analysis. *Organizational Research Methods, 26*(4), 608–628. https://doi.org/10.1177/10944281221120541

Prudon, P. (2015). Confirmatory factor analysis as a tool in research using questionnaires: A critique. *Comprehensive Psychology, 4*. https://doi.org/10.2466/03.CP.4.10

Quarfood, C. (2022). *The Montessori movement in interwar Europe: New perspectives*. Springer International Publishing. https://doi.org/10.1007/978-3-031-14072-3

Quarfood, C. (2023). Maria Montessori: Life and historical context. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *The Bloomsbury handbook of Montessori education* (pp. 5–20). Bloomsbury Academic.

Raimondo, R. (2023). Cosmic Education: The Vital Center of the Montessori Perspective. In M. McKenna, A. Murray, E. Alquist, & M. Debs (Eds.), *Handbook of Montessori education* (pp. 29-36). Bloomsbury Publications

Randolph, J. J., Bryson, A., Menon, L., Henderson, D. K., Kureethara Manuel, A., Michaels, S., Rosenstein, D. L. W., McPherson, W., O'Grady, R., & Lillard, A. S. (2023). Montessori education's impact on academic and nonacademic outcomes: A systematic review. *Campbell Systematic Reviews*, *19*(3), e1330. https://doi.org/10.1002/cl2.1330

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Ryan, A., Prieto-Rodriguez, E., Miller, A., & Gore, J. (2024). What can Implementation Science tell us about scaling interventions in school settings? A scoping review. *Educational Research Review*, *44*, 100620. https://doi.org/10.1016/j.edurev.2024.100620

Scippo, S. (2023). Construction and validation of a tool to measure Montessori educational practices in the Italian primary school. *Journal of Educational, Cultural and Psychological Studies (ECPS)*, *28*, 117–135. https://doi.org/10.7358/ecps-2023-028-scis

Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, *80*(3), 421–445. https://doi.org/10.1177/0013164419885164

Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, *37*(4), 240–256. https://doi.org/10.1080/07481756.2005.11909764

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2