Validity and Reliability of Early Childhood Normalization Instrument: First Steps

Tracey N. Sulak, Janet H. Bagby, Munube Yilmaz, & Crystal Collins
Baylor University

Literature Review

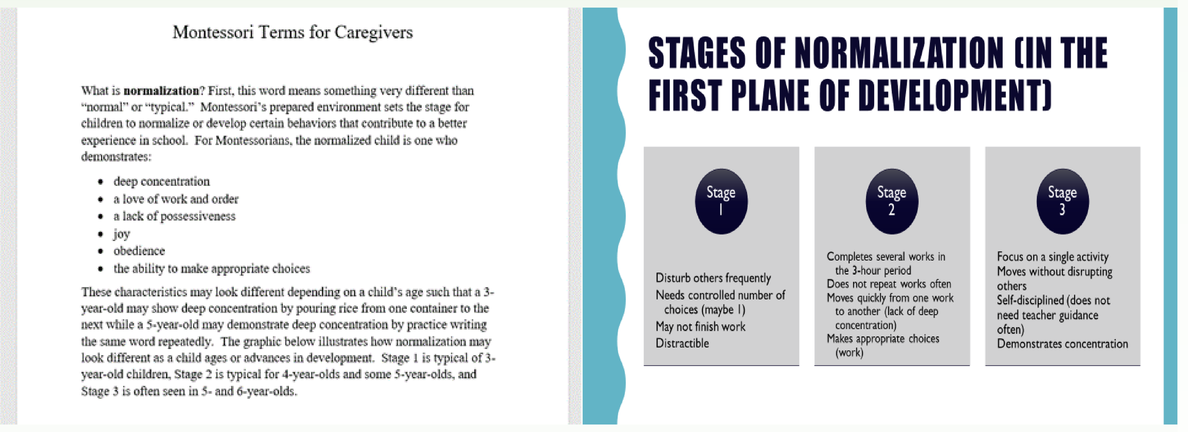
Montessori discussed the term "normalization" in many of her works. By her definition normalization occurs when a "normal child is one who is precociously intelligent, who has learned to overcome himself and to live in peace, and who prefers a disciplined task to futile idleness" (Montessori,1966, p.148.) Normalization is defined by terms such as willing compliance, independence, and self-discipline (Olaf, 2006). Montessorians know that a normalized child can concentrate and work constructively (Lillard, 2007). However, the terms used in describing normalization are not easily translated into observable, measurable behavior terms (Cooper, Heron, & Heward, 2007). While this is not an issue for seasoned Montessorians, it can be challenging to recognize and document for new teachers, parents, and researchers who are not trained in Montessori.

According to Montessori (2004) normalization of the child is the most important outcome of an educator's work, but discussing this phenomenon with the research world outside of Montessori requires a common language and a standardized measurement tool. The nuances of normalization in the first plane of development are easy to miss unless one is aware and educated about these small changes. For example, when a child moves from the first stage of normalization to the second stage, the change may be difficult to see because the child still has not developed true self-control or inner discipline (NAMC, 2008). Behaviors may not be consistent, but despite that inconsistency, the child may be progressing and becoming a productive member of the classroom. A psychometrically strong assessment that could capture these small differences would give educators and researchers a tool to document specific areas where a child is progressing.

Method and Procedure

The purpose of this study was to evaluate the reliability and validity of the Normalization Checklist. This Normalization Checklist was vetted by a group of Montessori experts for content validity and was utilized in a previous study but has not been studied thoroughly. This 7-item checklist is designed to serve as a progress monitoring tool for teachers and aides in early childhood classrooms.

Participants were 10 teachers from early childhood classrooms at a public Montessori school. In October the teachers participated in a training session on the definition and stages of normalization as well as the process for data collection. During the presentation, teachers were given the following graphic related to normalization for different stages of children. Teachers were also given a handout for families that explained normalization



Researchers created packets of paper copies of the Normalization Checklists for each new student in the classrooms and distributed them to the teachers. Beginning in October and every three weeks, the teachers completed the Normalization Checklist on each of the 7-12 children in their classrooms who were new to a Montessori environment. Each teacher assigned a code to each of his or her new children – the room number and the child's number in the alphabetical list. The code blinded the data such that only the teachers knew the ratings of a specific child. The completed forms were collected by the school office. Research assistants entered the data from the completed forms.

Research Questions

We sought to answer four research questions with our analyses and will provide the type of analyses in parenthesis by each question. Our research questions are as follows:

1. Do patterns of responses on the normalization checklist differentiate between children who have and have not reached normalization (Descriptive analysis; SEM; Mixture modeling)?
2. What is the reliability of the indicators in the normalization checklist (CFA)?
3. Does the normalization checklist demonstrate construct validity (CFA; Mixture modeling)?
4. Do the results provide evidence of concurrent validity between the normalization checklist and the teachers' assessment of normalization (SEM; Mixture modeling)?

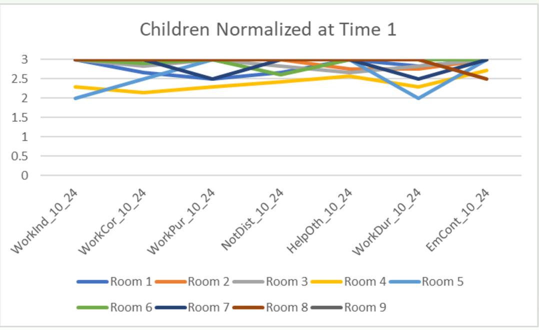
Analysis

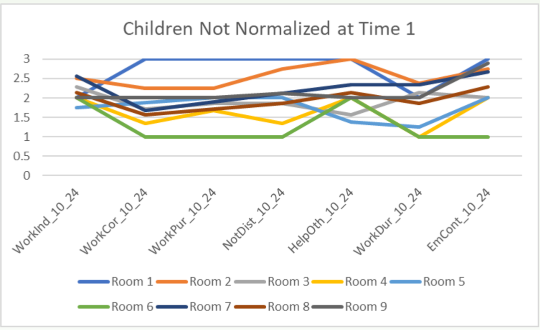
We completed three phases of data analyses: descriptive for all three timepoints, a confirmatory factor analysis (CFA) and structural equation modeling predicting normalization for time one data, and a mixture model for time one data. Descriptive analyses were conducted in Statistical Package for Social Sciences (*v.* 25) and Excel, while the CFA and mixture analyses were conduct in M*plus* (*v.*8.0). Longitudinal analyses were not possible due to missing data and multilevel modeling was not conducted due to the small sample size. Missing data were analyzed and determine to not be missing completely at random for time two and three.

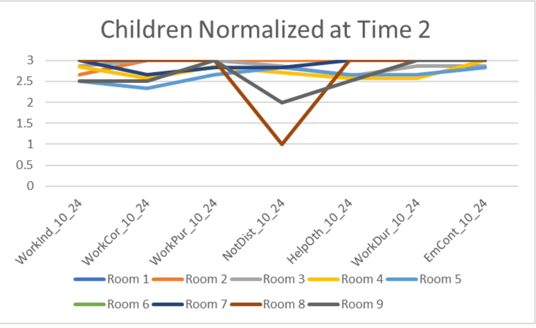
Results - Descriptive

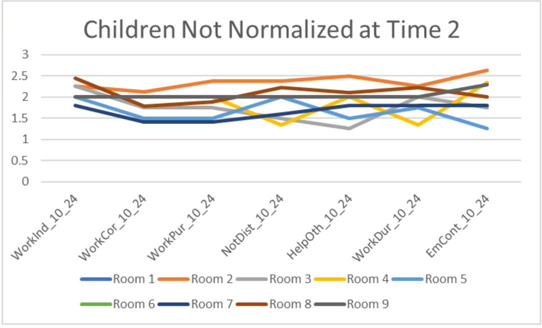
	Average age in classroom	Percent Normalized at Time 1	Percent Normalized at Time 2	Percent Normalized at Time 3
Room 1	3.4	0.9	n/a	n/a
Room 2	4.4	0.3	0.3	n/a
Room 3	4.2	0.5	0.7	0.67
Room 4	4.4	0.7	0.7	n/a
Room 5	4.6	0.2	0.6	n/a
Room 6	4.3	0.9	n/a	n/a
Room 7	4	0.2	0.6	0.6
Room 8	3.3	0.2	0.1	n/a
Room 9	4.4	0	0.2	0.56

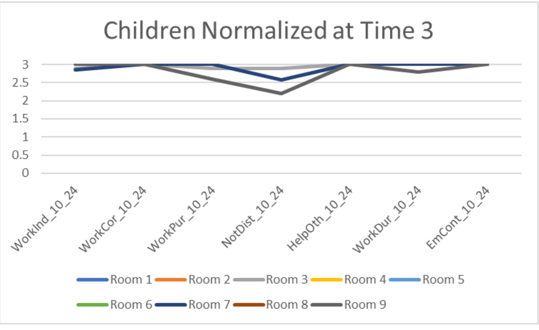
The average age in each classroom varied – 3.3 years to 4.6 years. There does not appear to be a relationship between average age in the classroom and the percentage normalized at time one. $r = -.161$, $p = .68$. The number of first year children in each classroom ranged from seven to twelve with classroom one having the fewest students and classroom two and three having the largest number of new students. Graphs of average ratings for children who were normalized and not normalized at each timepoint are below.

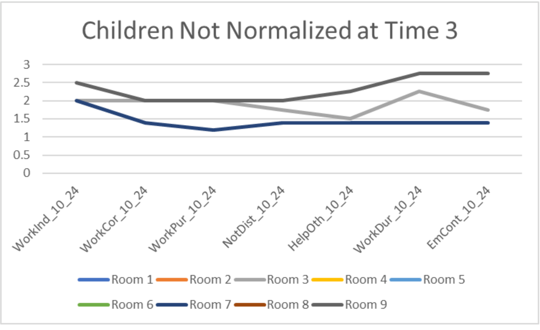












Results – CFA and SEM

To assess the strength of the reliability of the predictors as well as construct validity, a single-factor CFA was fit using the estimator Weighted Least Means Square Variance (WLSMV) . The model fit the data according to some indices – $\chi 2$ (14) = 39.7, $p = .0003$; Root Mean Square Error Of Approximation (RMSA) = .14; Comparative Fit Index (CFI) = .99; Tucker Lewis Index (TLI) = .98. The first indicator served set the scale for the latent variable (Brown, 2006). All indicator path values ranged between .80 and .97 – indicating high reliability for each item – and the model showed no discernable areas of strain. CFA standardized results are shown in Figure 1.

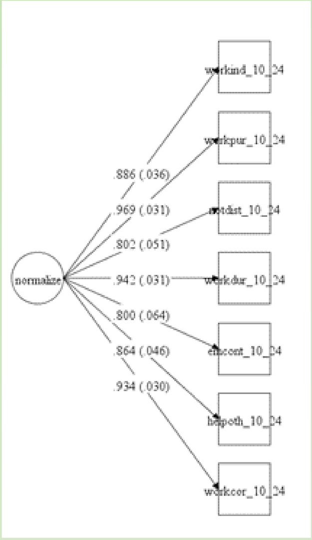


Figure 1. Standardized results of CFA

The CFA was used to predict normalization at time one. This timepoint was selected because it contained the least amount of missing data and would likely provide the most accurate estimate of the relationship between the measure of normalization and the teachers' ratings of normalization for first year students. As with the CFA, WLSMV was used as a estimator (Muthén & Muthén, 2012). The model fit the data with some caveats - $\chi 2$ (20) = 43.6, $p = .002$; RMSA = .11; CFI = .99; TLI = .99. It is likely the RMSEA value for the CFA and SEM is inflated due to the small sample size (Kenny & McCoach, 2003).

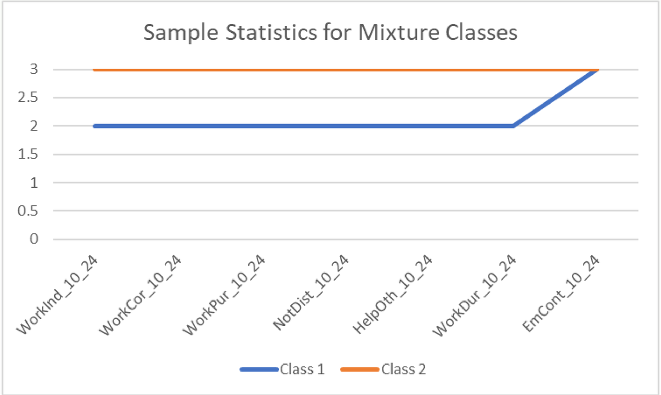
All standardized path values in the measurement model ranged between .80 and .96, indicating a high degree of reliability for all indicators. The standardized path value from the latent variable, normalize, to the teachers' rating of normalization was .91, $p = .001$. This value may be interpreted as a Pearson's r and means that about 82% of the variance in teachers' ratings is shared with the latent variable, normalization. In lay terms, there is a high level of agreement between these two measures at time one.

Results – Mixture Model

Since the data are nested within children within classrooms, a three-level model would honor the structure of the data, but missing data at time two and three led to convergence issues. Instead, we chose to honor the structure through mixture modeling for unobserved populations. Mixture modeling sorts observations by response patterns and is appropriate when heterogeneity is expected (Collins & Lanza, 2010). Three models were fit – one, two, and three class – to the CFA for normalize and the results were used to evaluate the classification of children who were normalized versus those who were not. Results from model testing are below.

	χ^2	p value	BIC	BIC_a	Entropy	LMR-LRT	BLRT
1 class	1023.6	1.00	1377	1310	n/a	n/a	n/a
2 class	916	1.00	1070	994	.95	.001	.001
3 class	540	1.00	1013	927	.91	1.00	.001

Models with more than three classes did not converge. Based on the interpretability of the results and indices, a two-class model was selected as the best fit. Sample statistics for the model are below.



Results – Mixture Model Continued

It is clear from the graph that the majority of children in class 1 had ratings of two for most of the items, but children in class two had ratings of three. Class 1 had 1.2% of children mis-classified according to the posterior probabilities. Class 2 had 1.6% misclassified. That means one child was mistakenly classified in class 1 and should have been in class 2. The reverse is true as well. Class 1 could be considered the class for children who were not normalized – 94% of these children were in class 1. Seventy-five percent of the normalized children were classified as class 2. All nine classrooms were evenly distributed between class 1 and class 2.

Discussion and Implications

To help the reader, we will discuss our findings and implications of the results by research question.

1. Patterns of responses on the instrument differentiate between children who are considered normalized and those who are not.

   Implications: Since all items are behavioral, teachers are observing and recording behavior data that translates into what we consider normalization. If the items differentiate, an outside research may be able to use this instrument to understand normalization in Montessori.

2. All of the items show strong reliability. This means changes in the rating on the items is connected to change in a latent variables – we call the latent variable normalization.

   Implications: With higher loading, it is assumed the instrument is performing in a patterned manner; if a student receives a 3 for using works correctly, he or she probably received a 3 for works independently.

3. The results of the CFA support construct validity. Low loadings on a CFA mean the items share less variance with the overall construct. In this case, anywhere from 64% to 85% of the variance in the items is explained by the construct. Content validity was addressed previously through a review by Early Childhood Montessorians. Establishing validity is an on-going process and we will continue studying the validity of the instruments results.

   Implications: All evidence collected thus far supports the use of these results for differentiating between students who have normalized and those who have not.

4. The teachers' assessment of normalization – a yes or no items – is predicted by the latent variable from the normalization checklist. The relationship between the two is strong indicating teacher assessment is often in agreement with the results of the instrument.

   Implications: These results support the concurrently validity of results from the instrument. Since a normalization measure does not exist, we used expert opinions to demonstrate this relationship. The relationship holds even when data are analyzed through person-centered procedures.

Future research will include repeating this study with other samples, continuing to collect data from the current participants, and conducting interrater reliability analysis.

Selected References

Lillard, A. S. (2007). Montessori: The science behind the genius. New York, NY: Oxford University Press.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling, 10, 333-3511.

Montessori, M. (1966). The secret of childhood. New York, NY: Ballantine Books.

Montessori, M. (2004). The discovery of the child. Delhi, India: Aakar Books.