Measuring Instructional Practice in Mathematics through Self-Report: A Construct Validation Study

Robert C. Schoen, Claire Riddell, Ahmet Guven, Gizem Solmaz-Ratzlaff, Qi Wang, and Lu Liu
Florida State University

Leave Empty

This space will be automatically filled with a QR code and number for easy sharing

Introduction

Studies of instructional practice require valid and reliable measures, and observation of instruction in large-scale studies is expensive. The Mathematics–Cognition, Language, Interaction, and Problem Solving (M-CLIPS; Riddell et al., 2021; Schoen et al., 2021) is an observation protocol designed to measure instructional practice in a large-scale, three-year randomized controlled trial of the effects of a professional development program called Cognitively Guided Instruction (CGI) program on teachers, teaching, and students. M-CLIPS Self-Report questionnaire was developed to measure instructional practice in a larger proportion of the sample than can be done through observation.

Purpose

The purpose of this study is to begin to assess the validity of the M-CLIPS Self-Report instrument as a measure of instructional practice.

Measures

The M-CLIPS Self-Report web-based questionnaire uses a retrospective pretest-posttest self-report design (Lam & Bengo, 2003; Little et al., 2020). There are 13 items describing instructional practices that are consistent with CGI and eight items that are not consistent with CGI but describe typical instructional practice in the U.S. (with the latter reverse coded). Each item asked about the daily frequency of the instructional practice using a response scale slider that ranged from 0 (none of the time) to 100 (all of the time) in 1-point increments.

Procedures

The questionnaire was administered to 1,277 K–5 Florida educators in the spring of 2021. Participants responded two times for each item: once corresponding to their current school year, and once corresponding to previous school year. This enabled separate evaluation of the response data for the current year and the previous year.

Psychometric Analyses and Related Results

Data analysis involved three main phases: (1) review of available data for missingness and errant values, (2) dimensionality and item analysis using used parallel analysis (PA) in R (R Core Team, 2017) 4.0.2 and exploratory factor analysis (EFA) using Mplus 8.0 (Muthén & Muthén, 1998 & 2017), and (3) reliability analysis.

Dimensionality and Item Analysis

Parallel analysis (PA) using the full available sample of 1,277 educators clearly suggested the presence of two factors.

The 2-factor EFA model had reasonable fit according to the traditional cut-offs suggested by Hu and Bentler (1999). The eight items that were flagged for reverse coding loaded onto one of the factors, while the remaining items loaded onto the other. PA indicated a single component for each of the two resulting scales. The two scales were subsequently named CGI and Traditional Instruction (TRAD).

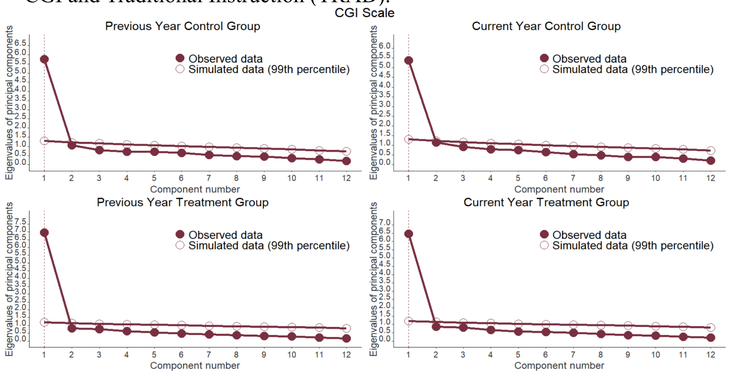

Figure 1. *Scree plots for the 12-item CGI scale, split by time point and subsample.*

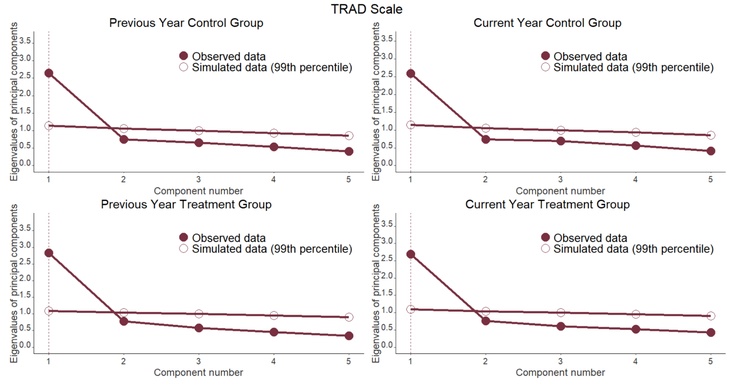

Figure 2. *Scree plots for the 5-item TRAD scale, split by time point and subsample.*

Further scale refinement then occurred. Items 1 and 2 were removed due to relatively low standardized factor loadings. Item 11 was removed due to a combination of relatively low factor loadings and concerns about whether the differences in grammar in the item might lead to a methods effect. PA still indicated a single component for each scale.

Reliability

Reliability was assessed using coefficients α and ω for the full sample as well as for the subsamples of educators in schools assigned to the CGI and comparison schools. (See Table 1.) Patterns suggest slightly higher reliability for the educators in the CGI condition than for those in the comparison condition.

Table 1. *Reliability Estimates for the CGI and Traditional Instruction Scales for the Intervention and Comparison Samples*

	CGI scale				TRAD scale		
Reliability coefficient	Comparison subsample	Intervention subsample	Full sample		Comparison subsample	Intervention subsample	Full sample
	Previous year						
α	.909	.941	.935		.772	.804	.796
ω	.910	.942	.935		.777	.808	.800
	Current year						
α	.894	.929	.921		.760	.779	.774
ω	.896	.930	.920		.764	.780	.777

Note. CGI = Cognitively Guided Instruction; TRAD = Traditional Instruction.

Limitations and Future Directions

M-CLIPS Self-Report was not administered before the intervention started. Data analysis split the sample by treatment condition, but a more formal study of measurement invariance is needed to rule out potential bias by treatment condition.

Conclusions

The M-CLIPS Self-Report appears to measure two distinct and recognizable scales. Initial field-testing and data analysis provide some evidence of structural validity (Flake, Pek, & Hehman, 2017) and reliability. More validation work is needed.

Acknowledgments

The research and development reported here was supported by the Institute of Education Sciences and U.S. Department of Education through grant award numbers R305A180429 and U423A180115 to Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute or the Department.