A VALIDATION ARGUMENT FOR THE PRIORITIES FOR MATHEMATICS INSTRUCTION (PMI) SURVEY

Michele Carney
Boise State University
michelecarney@boisestate.edu

Joe Champion
Boise State University
joechampion@boisestate.edu

Angela Crawford
Boise State University
Angelacrawford1@boisestate.edu

Mathematics education needs measures that can be used to research and/or evaluate the impact of professional development for constructs that are broadly relevant to the field. To address this need we developed the Priorities for Mathematics Instruction (PMI) survey consisting of two scales focused on the constructs of Explicit Attention to Concepts (EAC) and Student Opportunities to Struggle (SOS) – which have been linked to increased student understanding and achievement. We identified the most critical assumptions that underlie the proposed interpretation and use of the scale scores and then examined the related validity evidence. We found the evidence for each assumption supports the proposed interpretation and use of the scale scores.

Keywords: Instructional Activities and Practices, Teacher Beliefs, Measurement

Teacher beliefs are important predictors of classroom practice (Stipek, Givvin, Salmon, & MacGyvers, 2001). The field of mathematics education needs measures of teacher beliefs that are broadly applicable and useful across multiple research studies (e.g., for comparisons), and linked to student learning outcomes of value to the field (e.g., student achievement). In many cases, this has led to development of surveys to assess the degree to which teachers hold beliefs aligned with preferred approaches to mathematics instruction. However, teachers' beliefs are just one aspect of a complex system affecting teachers' instructional practices (Leatham, 2006), and though survey scores may be associated with implementation, the competing priorities of teachers instructional practice have important effects on classroom practice. There is a need for a survey about mathematics instruction that describes teachers' beliefs while foregrounding the competing priorities teachers must consider when making instructional decisions.

Our interest in developing a survey stems for our involvement in multiple K-12 teacher professional development (PD) projects with a goal to influence teachers' beliefs about particular instructional strategies. We value our collaborations with teachers and the competing priorities they weigh while making instructional decisions (e.g., limited time vs. a desire for building both conceptual and procedural fluency). Therefore, we wanted a survey that does not devalue the knowledge teachers have about their contexts, and that gives us the ability to understand and use a broader perspective to support use of effective instructional practices. In particular, our survey is aimed to be applicable and useful for examining the impact of PD on teachers' beliefs and implementation across our PD projects, and with scales that recognize teachers' priorities without explicitly privileging particular instructional strategies.

Perspectives

Our instrument development work is framed through two perspectives. We first describe the theoretical framework for effective mathematics instruction from which our survey scales are based. We then draw from modern validity theory, explaining our choice to use an argument-based approaches to validation.

Theoretical Framework for EAC and SOS

Our perspective on effective mathematics instruction centers on Explicit Attention to Concepts (EAC) and Student Opportunity to Struggle (SOS), which come from Heibert and Grouws' (2007) synthesis of literature regarding classroom practices connected to increases in student conceptual understanding and mathematics achievement. EAC refers to instructional practices involving public noting of connections among mathematical facts, procedures, and ideas, while SOS occurs when students expend effort to make sense of mathematics or figure something out that is not immediately apparent. Recently, Stein, Correnti, Moore, Russell, and Kelly (2017) investigated the relationship between EAC, SOS, and student achievement across a large group of teachers. They found students in classrooms with high EAC and SOS performed higher on mathematics achievement assessments of both conceptual understanding and skills efficiency. Based on the extensive literature base, the connections to student achievement, and the likelihood for broad applicability, we used the constructs of EAC and SOS as the starting place to develop our survey scales.

Our goal was to identify and situate the EAC and SOS constructs in contrast to common competing priorities for instructional focus. Studies of traditional mathematics instruction highlight beliefs among teachers that emphasize ways in which beliefs about learning and context factors relate to teachers' choices to prioritize mastery of procedural skills (Philipp, 2007) and identify a need to 'funnel' tasks to reduce cognitive demand (Peterson, Fennema, Carpenter, & Loef, 1989). We label this set of priorities as Single Methodological Focus (SMF) and Highly Scaffolded Content (HSC), respectively, and situate them as contrasting priorities to EAC and SOS.

Argument-Based Validation

Modern validity theory has been articulating and promoting the idea of instrument validation through the lens of argumentation for many years (Cronbach, 1988, Kane, 1992, Messick, 1995), culminating in recommendations for argument-based validation in *The Standards for Educational and Psychological Testing* (AREA, APA, NCME, 1999, 2014). While a variety of approaches to argument-based validation have been articulated, there is not one generally accepted approach (Carney, Crawford, Siebert, Osguthorpe, Thiede, 2019). Therefore, we use the recommendations from *The Standards* (AREA, APA, NCME, 1999, 2014) and Kane (1992, 2001, 2016) to guide our work.

Validity involves the degree to which the score interpretation for proposed uses is supported by theory and evidence[1], and validation involves constructing and evaluating arguments related to the score interpretation for proposed uses (AERA, APA, NCME, 2014). Therefore, the articulation of the score interpretation for proposed uses must be the first step in validation (Kane, 2001, 2016). The argument is further developed by articulating the assumptions that underlie the score interpretation and use (AERA, APA, NCME, 2014). Once the assumptions have been articulated, it is incumbent upon the instrument developers to gather evidence to investigate the most critical or suspect assumptions first (Kane, 2001, 2016).

The goals of this paper are to (a) articulate the score interpretations for proposed uses for two survey scales we have developed, (b) articulate the most critical or suspect assumptions that underlie the score interpretations for proposed uses, and (c) examine evidence in relation to those assumptions. We see this work as an initial step in the iterative cycle of instrument development and validation, with the goal of others using the scales and continuing to gather evidence in support of, or to refute, the assumptions that underlie the score interpretation for proposed uses.

Methods

Instrumentation

Structure of the scales. As mentioned in the section Theoretical Framework for EAC and SOS, we wanted to structure our survey scales to recognize the most likely competing priorities for these constructs. We set up these competing priorities along a continuum for each construct. One continuum contrasts Explicit Attention to Concepts (EAC) with Single Methodological Focus (SMF), which prioritizes a compartmentalized approach to mathematics instruction that focuses on teaching one important mathematical idea and/or procedure at a time, often in an attempt to reduce student confusion between different approaches to solving problems. This approach is often manifested in classroom practice by asking students to correctly apply a particular procedure to a set of problems. A second continuum contrasts Student Opportunity to Struggle (SOS) with Highly Scaffolded Content (HSC), which prioritizes a gradual increase in complexity of mathematics, with scaffolding for students to move from relatively easy to more challenging ideas and procedures. This approach is often manifested in classroom practice by teachers breaking down students' work into progressively more challenging tasks, with the teacher providing explanations as needed, so students can gradually build fluency.

Using the two continuums as underlying constructs, the Priorities for Mathematics Instruction (PMI) survey has two scales focused on teachers' prioritization of beliefs - *PMI: SMF-EAC beliefs* and *PMI: HSC-SOS beliefs*. Each survey item starts with a common stem and presents instructional practices representative of the two ends of the target continuum. Respondents select one of six positions to describe the relative priority they place on the competing statements. See Figure 1 for the directions at the start of the survey and an example item highlighting the continuum:

Survey Directions: Each item asks you to choose between two statements. Both have value, and you may believe both are important to your teaching. Nonetheless, we ask you to choose one over the other. That is, please position the slider to indicate which statement you believe has greater priority in terms of your perspective on teaching mathematics. The further you move the slider to one side or the other, the greater priority you give that statement.

Survey Item: Students are more likely to succeed in math when they can:

Apply a particular method to solving similar math problems	•—•—•—•—•—•	Make connections among math topics
	SMF … *EAC*	
	-2.5 … *2.5*	

**The information in italics is provided for reference and is not part of the item.*

Figure 1. Example of directions and an item for the PMI: SMF-EAC beliefs scale.

Interpretation and Use. The *PMI: SMF-EAC beliefs* and *PMI: HSC-SOS beliefs* scale scores (calculated as an average of the responses within the scale) can be interpreted in the following way. A score above 0 indicates beliefs more closely aligned with EAC or SOS practices, respectively. The closer the score gets to 2.5, the more closely the beliefs align with EAC or SOS. A score below 0 indicates the beliefs more closely align with SMF or HSC practices, respectively. The closer the score gets to -2.5, the more closely the teacher's beliefs align with SMF or HSC. A score near 0 indicates the teacher tries to balance the competing beliefs in their instructional priorities. The *PMI: SMF-EAC beliefs* and *PMI: HSC-SOS beliefs* scale scores can be broadly used by professional developers to examine beliefs relative to these constructs, inform professional development activities, and evaluate the effectiveness of PD activities in regards to their impact on teachers' beliefs related to EAC and SOS.

Critical Assumptions. Once the interpretation and use are clearly stated for an instrument, it is incumbent upon the developer to investigate the underlying assumptions (AERA, APA, NCME, 2014). The initial focus should be on the assumptions that are the most critical to demonstrate or the most likely to fail (i.e., are most suspect) (Kane, 2001). We have identified the following assumptions as particularly critical in our initial investigation of the interpretation and use of the *PMI: SMF-EAC beliefs* and *PMI: HSC-SOS beliefs* scale scores. For all instruments, there is an assumption that the operationalization aligns with the construct(s) theorized structure (assumption 1). For instruments such as the PMI survey where use is proposed (a) across a variety of professional development projects, the assumption is that the construct is broadly relevant to a mathematics education audience (assumption 2), and (b) related to measuring growth, the assumption is the instrument is sensitive enough to detect growth in an individual or group (assumption 3). Lastly, for instruments such as the PMI survey where social desirability of the response is a potential unintended factor, the assumption is social desirability is not impacting the scores (assumption 4).

Instrument Administration

Data Collection. The survey was administered to teachers participating in programs offered by a single K-12 math PD center in the Pacific Northwest. The programs are diverse in format, content, and duration, ranging from content-focused workshops to multi-year collaborative projects. There are clear differences in approach across the three PD groups [Blinded for Review]: Program 1, Program 2, and Other. Program 1 is a state-mandated 3-credit course in which K-12 educators build mathematical knowledge for teaching with a special emphasis on increased awareness of EAC and SOS, Program 2 is a federally-funded teacher-researcher alliance of Grades 6-8 teachers with an emphasis on adapting EAC and SOS strategies for their classroom practice, and the Other programs incorporate EAC and SOS ideas in their design, but not as the primary emphasis. Surveys were administered online via email invitation just before participating in the PD (pre, N = 645) and again (depending on program timing) 2 to 8 months later (post, n=321). Data collection spanned July 2019 to February 2021, with paired post/pre-response rates differing by PD group (Program 1: 48/107 (45%), Program 2: 78/106 (74%), Other: 195/432 (45%)).

Analysis

Statistical analyses of the survey response data was conducted in the statistical software package *R* (R Core Team, 2020), following recommendations for scale development by Jackson, Gillaspy, and Purc-Stephenson (2009). This included inspection of item response distributions, estimation of the bivariate correlational structure, and confirmatory factor analysis (CFA) using the *lavaan* software package (Rosseel, 2012). Missingness assumptions were evaluated under Little and Rubin's recommendations (1989), with iterative multiple imputation (van Buuren & Groothuis-Oudshoorn, 2011) used to augment incomplete responses (6.7%) without introducing bias into the fitted factor model. Evaluation and reporting of CFA model fit and parameter estimates followed guidelines by Cabrera-Nguyen (2010), with emphasis on indications of construct validity given the space restrictions of this report. Potential differences in pre-post PMI beliefs across subsamples were assessed using standard inferential statistical procedures (e.g., descriptive summaries, plots, ANOVA).

Results

Operationalization Aligns with Theory (Assumption 1)

The internal structure of the pre-responses were analyzed via a two factor CFA model using maximum likelihood estimation, with the eight EAC items loaded onto a latent "eac" factor, and the seven SOS items loaded onto a latent "sos" factor. The two factors were standardized (mean 0, standard deviation 1) and assumed to be correlated. The estimated model converged in 19 iterations with 31 free parameters, with indicators suggesting good fit between the theoretical model and the observed structure (model $X^2(89) = 304$, null $X^2(105) = 3228$, AIC $= 30140$, BIC $= 30279$, RMSEA $= .06$, CFI $= .93$, TFI = .92, SRMR $= .04$) with no areas of local strain and statistically significant factor loadings (all z > 10, $p < .0001$). Similarly, the model exhibited strong convergent and discriminant validity with standardized factor loadings strictly between 0.4 and 0.8 (see Figure 2). Follow-up principle component analysis identified no indications of cross-loadings (suggesting strong convergent validity), and the correlation between eac and sos beliefs (0.71) was below 0.80, suggesting strong discriminant validity. The evidence of model fit provides support for the unique operationalization of SOS to HSC and EAC to SMF as a continuum.

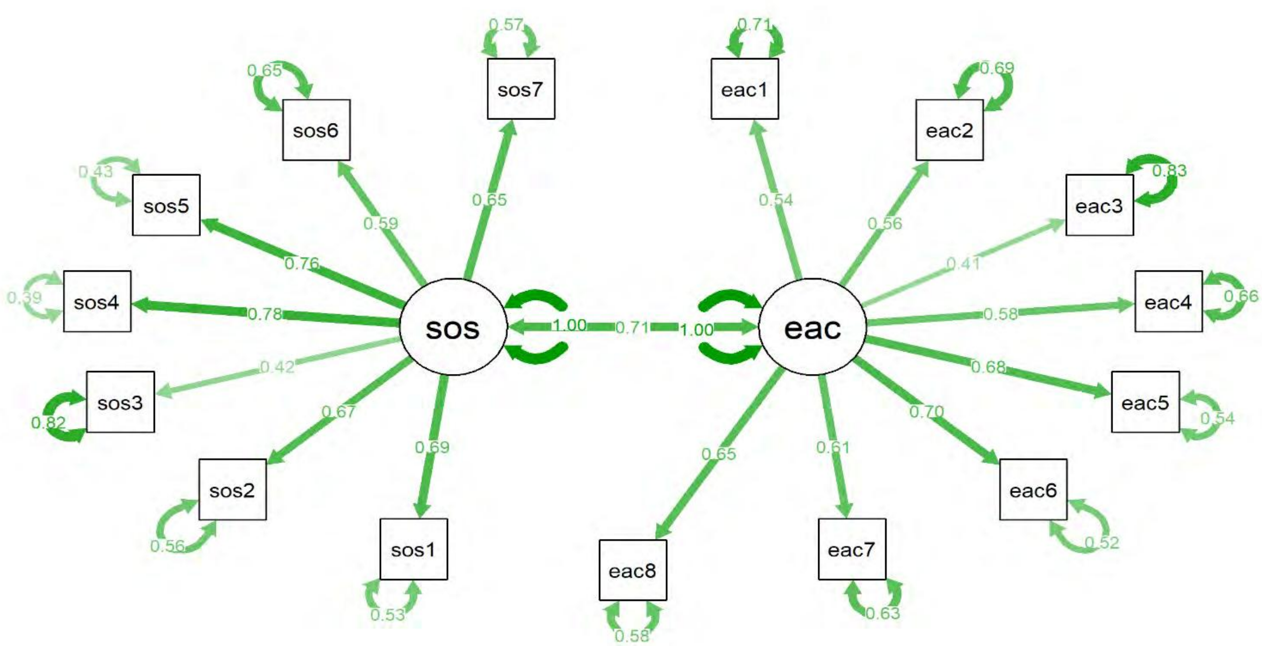

Figure 2. Standardized estimates for two-factor CFA model of EAC and SOS beliefs.

Broad Relevance (Assumption 2)

In addition to the theoretical argument establishing broad relevance and applicability of EAC and SOS across mathematics education settings (see section Theoretical Framework for EAC and SOS), the pre-distributions of PMI scale scores across the PD groups supports Assumption 2. Though each group differed in contextual variables, they had similar initial distributions of EAC and SOS belief scores (Table 1), this indicates the scales are likely to be broadly useful across different PD groups and settings.

Table 1. Distributions of EAC, SOS, and PMI Quadrants by PD Group

		EAC		*SOS*		*PMI Quadrants*			
Group	*n*	M	SD	M	SD	EAC&SOS	SMF&SOS	SMF&HSC	EAC&HSC
Program 1	107	0.1	0.8	0.0	0.9	41%	9%	36%	13%
Program 2	106	-0.2	0.8	-0.1	0.7	32%	16%	48%	4%
Other	432	0.0	1.0	0.0	1.0	43%	7%	40%	10%

Sensitivity to Group and Individual Changes (Assumption 3)

Figure 3 supports the potential for the PMI survey to detect change in teachers EAC and SOS beliefs. The chart illustrates how teachers in each PMI quadrant shifted in the post assessment, including a general pattern of small changes among teachers who began in the EAC&SOS quadrant, while teachers in the other quadrants showing increased variability in their post scores while generally shifting toward EAC&SOS. The ability to detect differential growth based on pre-PD scale scores indicates utility of the survey for detecting group and individual changes.

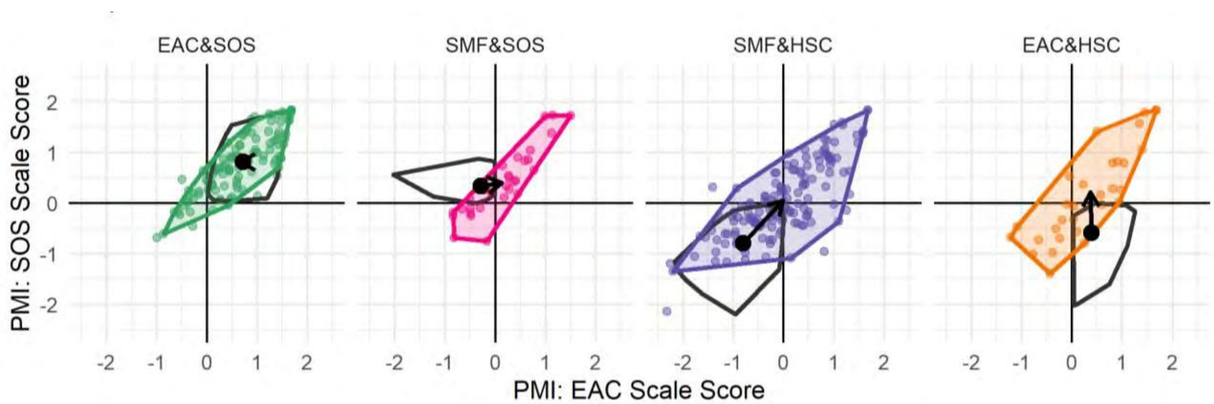

Figure 3. Post EAC and SOS scores, split by pre PMI Quadrant.
(Polygons capture the middle 90% of points by group, arrows indicate mean change.)

Figure 4 illustrates pre-post changes across the PD groups in EAC and SOS. The chart demonstrates substantial shifts toward prioritizing EAC and SOS among teachers in the Program 1 group. The ability to detect differential growth across PD contexts supports this proposed use.

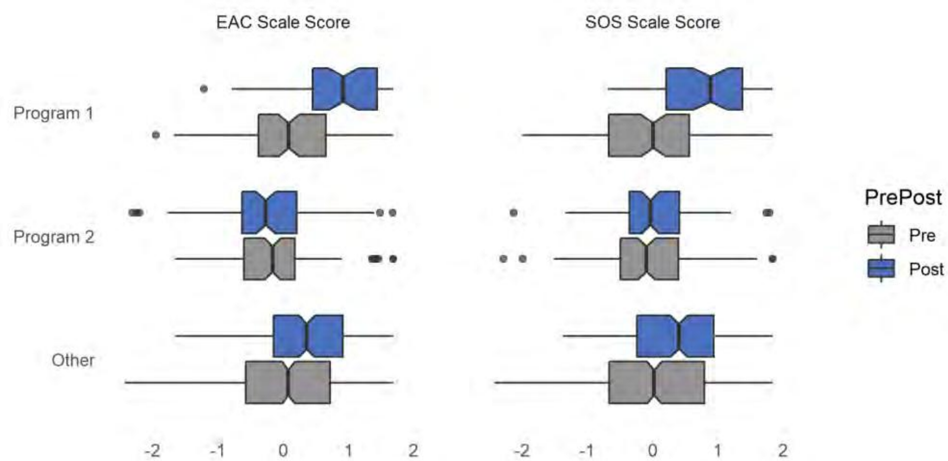

Figure 4. Pre and post distributions of EAC and SOS beliefs by PD Group. (Non-overlapping central notches indicate statistically different group medians.)

Social Desirability Response Bias (Assumption 4)

The paired pre-post EAC and SOS scores suggest minimal risks of social desirability response bias at the individual or group levels. Though teachers tended to shift toward the EAC&SOS quadrant after participating in PD (see Figure 5), the magnitudes and directions of those shifts varied greatly, with greater variability within groups than across. This variability supports the assumption that the social desirability of the response options is not obvious to respondents following PD that includes a focus on EAC and SOS.

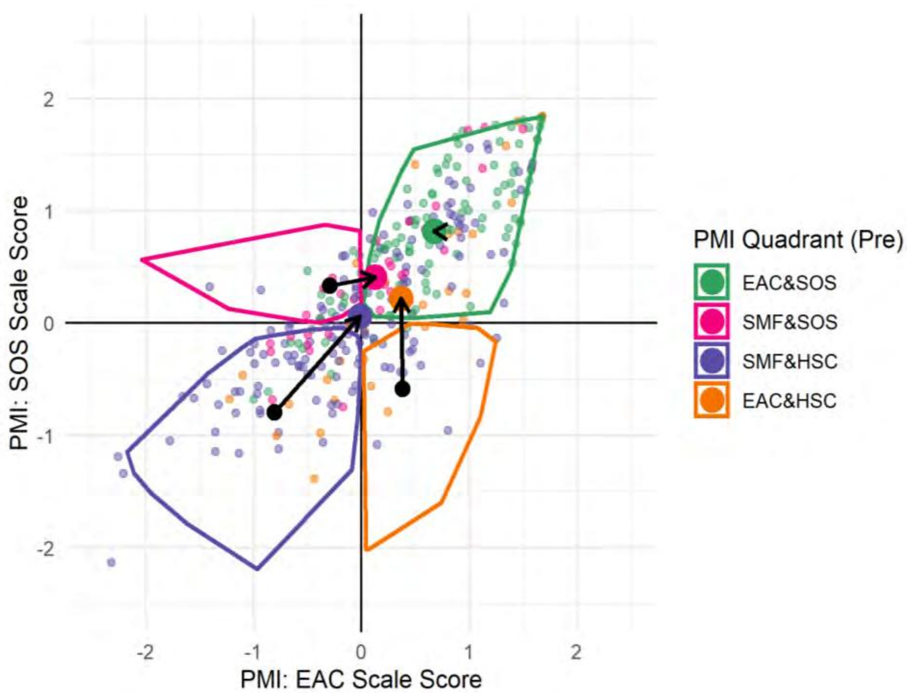

Figure 5. Post EAC & SOS Scores by Pre PMI quadrant. (Polygons capture the middle 90% of points in each group, arrows indicate mean changes.)

Olanoff, D., Johnson, K., & Spitzer, S. (2021). *Proceedings of the forty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Philadelphia, PA.

In addition, pre-post changes in EAC and SOS beliefs differed across the PD groups, with substantial changes in Program 1 (EAC: M = 0.8(SD = 0.9), SOS: 1.1(1.0)), insignificant changes in Program 2 (EAC: 0.0(0.8), SOS: 0.1(0.6)), and moderate changes in Other (EAC: 0.4(0.8), SOS: 0.5(0.8)). As shown in Figure 6, changes in PMI Quartiles differed significantly across the PD contexts. All three PD contexts included information about why EAC and SOS are important for classroom practice, and Program 2 in particular emphasized engaging in activities that make use of EAC and SOS in the classroom. However, there was significant variability in the amount of change in EAC and SOS across PD contexts with Program 2 having the least change and most focus on EAC and SOS. This evidence of variability across PD contexts supports the assumption that social desirability in favor of EAC and SOS is not impacting responses to the survey items. If it were, we would have expected the Program 2 scores to have shifted to reflect this bias.

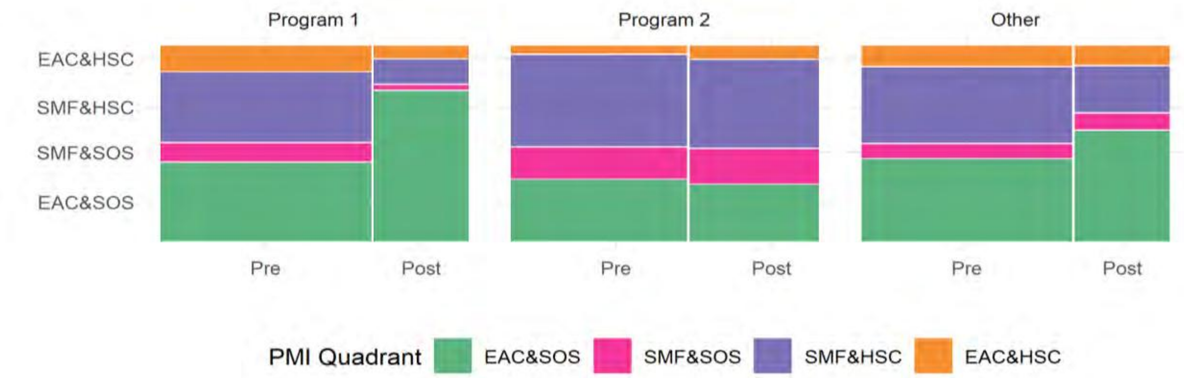

Figure 6. Pre-post changes in distributions of PMI quadrants across PD Groups.

Discussion

Instrument validation is an iterative process. This work presents an initial set of evidence for the interpretation and use of the PMI survey scale scores for *PMI: SMF-EAC* beliefs and *PMI: HSC-SOS* beliefs. Following the recommendations of the *Standards* (AERA, APA, NCME, 2014) we stated the interpretation and use of the two survey scales and identified the most critical assumptions to investigate. In particular, we investigated the following assumptions, and examined the associated evidence.

- The operationalization aligned with our theory (assumption 1). The CFA indicated a good fit which provides support for the unique operationalization of SOS to HSC and EAC to SMF as separate continuums of competing priorities.
- The survey scales scores are broadly relevant to the mathematics education community (assumption 2). The grounding of the scales in the work of Hiebert & Grouws (2007) and Stein and colleagues (2017), in addition to the finding of similar measures of center, variability and quadrant percentages across PD contexts, provide evidence in support of this assumption.
- The survey scales are sensitive enough to identify group and individual changes (assumption 3). The evidence of scale score changes from the perspective of both the pre-PD quadrant and three different PD contexts provides support for this assumption.

- Social desirability did not impact post survey responses (assumption 4). The strongest evidence in support of this assumption is that Program 2 participants – where the primary focus of the PD is EAC and SOS – had the least changes in pre-post scale scores.

Taken together the evidence in support of the four critical assumptions provides an important initial investigation into the interpretation and use of the PMI survey scale scores. We see this evidence as sufficient for recommending the use of the survey scales more broadly within the mathematics education community and hope that others will make use of the instrument and conduct additional validity investigations.

It is important to note a few key limitations. We did not complete a full investigation of the validity argument. There are additional assumptions that need to be examined and as the survey is used we anticipate others might have additional interpretation and use ideas that expand upon what was stated here. These would require further investigation. Finally, this work occurred during the COVID-19 pandemic, which likely impacted survey responses in complicated ways.

Note

[1] The Standards explicitly state "It is incorrect to use the unqualified phrase "the validity of the test" (p. 11)."

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. American Educational Research Association.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). Standards for educational and psychological testing. American Educational Research Association.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. Journal of the Society for Social Work and Research, 1(2), 99-103.

Carney, M., Crawford, A., Siebert, C., Osguthorpe, R., & Thiede, K. (2019). Comparison of Two Approaches to Interpretive Use Arguments. Applied Measurement in Education, 32(1), 10-22.

Cronbach, L. J. (1988). Five perspectives on validity argument. Test validity, 3-17.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning*, *1*, 371-404.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychological methods, 14(1), 6.

Kane, M. T. (1992). An argument-based approach to validity. Psychological bulletin, 112(3), 527.

Kane, M. T. (2001). Current concerns in validity theory. Journal of Educational Measurement, 38(4), 319–342. doi:10.1111/jedm.2001.38.issue-4

Kane, M. T. (2016). Validation strategies: Delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. Raymond, & T. M. Haladyna (Eds.), Handbook of test development (Vol. 2, pp. 80–96). New York, NY: Routledge.

Leatham, K. R. (2006). Viewing mathematics teachers' beliefs as sensible systems. *Journal of Mathematics Teacher Education*, *9*(1), 91-102.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. Sociological Methods & Research, 18(2-3), 292-326.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), 741.

Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teacher's pedagogical content beliefs in mathematics. *Cognition and instruction*, *6*(1), 1-40.

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. *Second handbook of research on mathematics teaching and learning*, *1*, 257-315.

R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48(2), 1–36. https://www.jstatsoft.org/v48/i02/.

Stein, M. K., Correnti, R., Moore, D., Russell, J. L., & Kelly, K. (2017). Using theory and measurement to sharpen conceptualizations of mathematics teaching in the common core era. *AERA Open*, *3*(1), 2332858416680566.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and teacher education*, *17*(2), 213-226.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. https://www.jstatsoft.org/v45/i03/.