Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cross-construct Optimization of TIMSS Mathematics Affective Scales

Jihyun Hwang

jihyun-hwang@knue.ac.kr

Korea National University of Education

Research Article

Keywords: mathematics attitude, bridge centrality, cross-construct measurement, scale shortening, cross-construct optimization

Posted Date: July 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9964636/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Students' attitudes toward mathematics—interest, confidence, and perceived value—are important predictors of educational outcomes, yet assessing all three dimensions simultaneously often imposes substantial survey burden. Using TIMSS 2023 data from 232,970 Grade 8 students across 44 countries, this study examined whether cross-construct information among mathematics affective scales can be used to develop a compact screening instrument.

All 26 items from the TIMSS Students Like Learning Mathematics (SLM), Students Confident in Mathematics (SCM), and Students Value Mathematics (SVM) scales were pooled and analyzed using a Gaussian Graphical Model. Items were ranked by betweenness centrality and sequentially added to predict TIMSS official three-level classifications (High, Intermediate, Low) through multinomial logistic regression. Results showed that SVM items occupied the most central bridge positions in the affective network. The item "I would like a job that involves using mathematics" exhibited the highest betweenness centrality and was selected first across all three dimensions. Predictive performance plateaued with seven items for SLM, six items for SCM, and two items for SVM. The SLM and SCM solutions shared a common six-item core, yielding a seven-item battery capable of simultaneously screening all three dimensions with high accuracy (AUROC = .889–.939), representing a 73% reduction from the original 26-item instrument. Cross-national analyses demonstrated generally stable predictive performance across 44 countries despite moderate variation in bridge-item structure. The findings identify mathematics value beliefs as the structural hub of the mathematics affective network and provide a practical seven-item instrument for efficient affective monitoring in research and educational settings.

Introduction

Students' affective engagement with mathematics — their interest in the subject, confidence in their ability, and beliefs about its value — has long been recognized as a consequential domain of educational research. Large-scale international assessments such as the Trends in International Mathematics and Science Study (TIMSS) devote substantial questionnaire space to measuring these dimensions, precisely because affective engagement predicts patterns of participation, course selection, and ultimately career trajectories in mathematics-intensive fields (Campos et al., 2025). TIMSS 2023 assessed three distinct affective scales for Grade 8 students: Students Like Learning Mathematics (SLM), Students Confident in Mathematics (SCM), and Students Value Mathematics (SVM), together comprising 26 items administered to over 230,000 students across more than 40 countries. Yet despite the scale and richness of this data, a fundamental measurement question has not been addressed: do these three scales, and their constituent items, operate as an integrated affective network, or as independent constructs that happen to be assessed together?

Three research questions guide the study. First, which items in the pooled 26-item TIMSS affective battery occupy bridge positions in the conditional dependence network, and which affective dimension

do they primarily represent? Second, what is the minimum cross-construct item set needed to simultaneously predict SLM, SCM, and SVM classifications, and what does the composition of this set reveal about the structural organization of mathematics affect? Third, how consistently does the identified bridge item structure replicate across countries, and what are the boundary conditions of the cross-construct framework?

The study makes three contributions. Empirically, it provides the first item-level characterization of the conditional dependence structure across TIMSS SLM, SCM, and SVM items in a large international sample (N = 232,970; 44 countries), identifying SVM items as the structural hub of the mathematics affective network — a finding that extends and empirically specifies the TMA framework. Methodologically, it introduces the cross-construct optimization approach to large-scale educational assessment, demonstrating that a seven-item battery derived through GGM bridge centrality can replace 26 items with negligible loss of predictive accuracy. Practically, it provides a freely available compact instrument — with both a batch-prediction Python module and a browser-based individual feedback tool — that enables periodic monitoring of student affect across all three TMA dimensions simultaneously in research and classroom contexts.

Theoretical Background

Attitude toward Mathematics as a Multidimensional Construct

Research on affect in mathematics education has long grappled with the conceptual ambiguity surrounding the term *attitude*. Early definitions treated attitude as a simple positive–negative evaluation of mathematics (Di Martino & Zan, 2003), but this unidimensional view proved insufficient to account for the complex, sometimes contradictory, ways students respond to mathematics — excelling in a subject they openly dislike, or expressing strong interest while doubting their own competence. In an effort to resolve this conceptual ambiguity, Di Martino and Zan (2010) developed the Three-dimensional Model of Attitude toward mathematics (TMA) through a qualitative analysis of over 1,600 autobiographical essays written by students from Grade 2 to Grade 13. Their analysis converged on three constitutive dimensions: Emotional Dimension (ED), reflecting students' positive or negative affective responses to mathematics; Perceived Competence (PC), reflecting self-assessments of mathematical ability; and Vision of Mathematics (VM), reflecting students' beliefs about the utility and relevance of mathematics. Importantly, the TMA does not treat these three dimensions as independent but rather as mutually influencing: a student's emotional disposition may shape perceived competence, and both may be modulated by beliefs about the usefulness of mathematics in life and career contexts.

The TMA maps closely onto the three TIMSS affective scales used in the present study. As Author (2017) established through a direct comparison of Di Martino and Zan's (2010) coded student statements with the TIMSS questionnaire items, the SLM scale aligns with the ED, the SCM scale aligns

with PC, and the SVM scale aligns with the VM dimension. This theoretical correspondence provides the primary justification for treating SLM, SCM, and SVM as a theoretically coherent, mutually complementary set of affective constructs rather than an arbitrary selection of scales (see also Choi & Choi, 2025).

A critical implication of the TMA framework for the present study is what it does not specify: the framework identifies the three dimensions as mutually influencing but does not designate any single dimension as the primary integrative hub of the affective network. Whether ED, PC, or VM plays the most central connective role — in the sense of carrying the most cross-construct predictive information — remains an open empirical question that the present study addresses for the first time at the item level using a Gaussian Graphical Model (GGM).

Interdependence of Affective Constructs

Eccles and colleagues' expectancy-value theory (Wigfield & Eccles, 2000; Eccles & Wigfield, 2020) provides the motivational architecture within which the three TMA dimensions — and by extension SLM, SCM, and SVM — are theoretically connected. In this framework, students' achievement-related choices and persistence are primarily shaped by two sets of beliefs: expectancy of success (the student's expectation of doing well, corresponding to PC and SCM) and subjective task value (the student's assessment of how important, useful, or interesting the task is, corresponding to ED/VM and SLM/SVM). Crucially, these beliefs are also not treated as independent predictors as discussed by Di Martino and Zan (2010) but as elements of a reciprocal motivational system in which competence self-perceptions influence value beliefs and vice versa. A student who anticipates success in mathematics is more likely to find it intrinsically interesting and to value it for future goals; conversely, strong utility value beliefs may sustain engagement even when immediate affective enjoyment is limited.

The implication of expectancy-value theory for the present study is directional. Because SVM captures beliefs about the utility and relevance of mathematics, it carries instrumental information that spans both the affective (interest) and cognitive self-assessment (confidence) dimensions. A student who endorses items such as “I would like a job that involves using mathematics” is simultaneously encoding information about positive affect toward mathematics (a high-SLM-like orientation) and a self-belief that mathematics is achievable enough to be worth pursuing (a high-SCM-like orientation). This theoretical prediction — that SVM items should exhibit disproportionate cross-construct predictive information — is directly testable through the GGM betweenness centrality analysis reported in the present study.

Evidence and Implications for Measurement. One of the most persistent methodological challenges in mathematics attitude research is the tension between the theoretical claim that SLM, SCM, and SVM are distinct dimensions and the empirical observation that they co-vary substantially and interact in complex, sometimes non-linear ways (Michaelides et al., 2019). Author (2017), using hierarchical linear modeling of TIMSS 2011 data from US and Finnish Grade 8 students, found that confidence in mathematics (SCM) had the strongest relationship with mathematical reasoning across all levels of

liking and valuing, but that this relationship was modulated by the other two dimensions through significant two- and three-way interaction effects. In particular, at high levels of liking mathematics, increasing confidence was associated with a negative relationship between valuing mathematics and reasoning — a pattern inconsistent with a simple additive model in which each dimension contributes independently. The authors concluded that the paradoxical and often inconsistent findings in prior attitude–achievement research could be attributed precisely to the independent analysis of each TMA dimension without comprehensive consideration of multidimensionality.

This insight motivates the cross-construct design of the present study. If the three affective dimensions operate as an interrelated system rather than as independent constructs, then a measurement strategy that pools items across scales and explicitly seeks cross-construct informational bridges is theoretically more appropriate than one that treats each scale in isolation. The analogy to Author's (2022) analysis of teacher epistemic orientation is instructive: just as those authors found that beliefs about learning, knowledge, and teaching were not separable entities but collectively shaped a unidimensional orientation, the present study treats SLM, SCM, and SVM items as elements of a single network in which information flows across construct boundaries. The identification of bridge items — items whose conditional dependencies span multiple construct boundaries — directly captures the empirical signature of this interdependence.

Network Psychometrics: Gaussian Graphical Models and Bridge Centrality

The analytical framework of the present study draws on network psychometrics, specifically the Gaussian Graphical Model (GGM; Epskamp & Fried, 2018; Borsboom et al., 2021). In contrast to latent variable models, which represent observed items as imperfect indicators of unobserved constructs and model the relationships between those constructs as correlations or path coefficients, GGMs model the conditional dependence structure among observed items directly. Each node in a GGM represents an item, and each edge represents the partial correlation between two items after controlling for all other items in the network. An edge that survives regularisation — here, graphical LASSO with EBIC tuning (Epskamp & Fried, 2018) — indicates a genuine conditional relationship between two items that is not simply an artifact of their shared loadings on a latent factor.

This item-level conditional dependence structure is theoretically informative in the present context because it reveals which items function as bridges across construct boundaries. An item that shows strong partial correlations with items from multiple scales — after controlling for all other items — cannot be explained solely by within-scale variance; it must carry information that is relevant to more than one affective dimension. Betweenness centrality, which quantifies the extent to which a node lies on the shortest paths between other nodes in the network, provides an empirically grounded index of bridge importance: items with high betweenness centrality are informational hubs through which variance flows across the network.

A methodological strength of applying GGMs to TIMSS data in particular is the large sample size available for regularised estimation. With N = 232,970 students across 44 countries, the graphical LASSO-estimated network is expected to be stable, and cross-national replication analyses are statistically well powered. The bridge item consistency index introduced in the present study (building on Li et al., 2024) — the proportion of overall top-k bridge items reproduced in each country's own GGM — provides a novel cross-national reliability metric for network-based item selection that has not previously been reported in the mathematics education literature.

Synthesis and Rationale for the Present Study

The three theoretical strands reviewed above — the TMA's specification of mutually influencing affective dimensions, the expectancy-value model's account of how value beliefs mediate competence and interest, and network psychometrics' tools for characterizing conditional dependence at the item level — converge on a common implication: the items comprising SLM, SCM, and SVM should not be treated as three independent pools but as an integrated affective network in which cross-construct information flows across scale boundaries. Realizing this implication for measurement practice, however, requires an explicit analytical framework for identifying and exploiting cross-construct information. The present study adopts the cross-diagnostic optimization framework introduced by Lim et al. (2026) for simultaneous screening across multiple psychiatric disorder classifications. In that framework, items are pooled across diagnostic scales, ranked by their betweenness centrality in a Gaussian Graphical Model, and added sequentially in bridge-first order until predictive performance plateaus — at which point the identified item set constitutes a minimum cross-diagnostic battery. The critical innovation is that items from one scale are permitted to serve as predictors for the classification of another scale, exploiting the cross-construct conditional dependencies that conventional within-scale shortening ignores. We apply this framework, for the first time, to mathematics attitude measurement in large-scale educational assessment.

Despite the theoretical motivation for cross-construct measurement and the methodological availability of the framework, no prior study has applied cross-construct optimization to the TIMSS mathematics affective scales, nor has the bridge item structure of SLM, SCM, and SVM been empirically characterized at the item level across a large international sample. The present study addresses this gap through three research questions: (1) which items in the pooled 26-item TIMSS affective battery occupy the highest bridge positions in the conditional dependence network, and which affective dimension do they primarily represent; (2) what is the minimum cross-construct item set needed to simultaneously predict SLM, SCM, and SVM classifications, and what does the composition of this set reveal about the structural organization of mathematics affect; and (3) how consistently does the identified bridge item structure replicate across countries, and what are the boundary conditions of the cross-construct framework. The urgency of such monitoring is underscored by evidence that students’ interest, enjoyment, and confidence in mathematics decline substantially across grade levels (Campbell et al., 2025).

Method

This study applies a cross-diagnostic optimization framework (Lim et al., 2026) to the TIMSS 2023 mathematics affective scales to identify a minimum set of cross-construct bridge items capable of reproducing the official three-level classification (High / Intermediate / Low) with equivalent accuracy across countries. Rather than selecting items within each scale independently, we pooled all 26 items from the three affective scales and ordered them by their betweenness centrality in a Gaussian Graphical Model (GGM). Items were then added sequentially in this bridge-first order until predictive performance plateaued. Crucially, items from one scale were permitted to contribute to the prediction of another scale's official classification — the defining feature of a cross-construct design.

Data Description

Source and Sample. Data were drawn from the TIMSS 2023 Grade 8 international database (Mullis et al., 2024). The analytic sample comprised N = 232,970 students from 45 countries after listwise deletion of cases with missing or invalid item responses. Country-level sample sizes ranged from 2,706 (Romania) to 15,077 (Brazil). One country (Malaysia) was excluded from country-level analyses due to severe class imbalance on one or more official scale indices (proportion < .05 in at least one level), leaving 44 countries for country-level replication analyses.

Measures. Three TIMSS 2023 Grade 8 student mathematics affective scales were examined. Items are rated on a four-point scale (1 = Agree a lot, 4 = Disagree a lot). Reverse-scored items were recoded (5 − x) prior to analysis.

Table 1
TIMSS 2023 Grade 8 Mathematics Affective Scales

Scale	Variable prefix	N items	Reverse-scored items	Official index
Students Like Learning Mathematics (SLM)	BSBM19A–I	9	BSBM19B, BSBM19C	BSDGSLM
Students Confident in Mathematics (SCM)	BSBM22A–H	8	BSBM22B, BSBM22C, BSBM22G, BSBM22H	BSDGSCM
Students Value Mathematics (SVM)	BSBM23A–I	9	None	BSDGSVM

TIMSS 2023 provides official IRT-based three-level indices for each scale (1 = High, 2 = Intermediate, 3 = Low), derived from each scale's full item set. These indices served as classification targets, avoiding the need for arbitrary percentile-based cut-offs and providing internationally validated reference labels.

Gaussian Graphical Model Estimation

A GGM was estimated on the pooled 26 items to characterize conditional dependence among items across all three scales. Each edge in the GGM represents the partial correlation between two items after controlling for all remaining items, enabling empirical identification of items that bridge across construct boundaries. The GGM was estimated using the graphical LASSO with Extended Bayesian Information Criterion (EBIC; γ = 0.5) tuning, implemented via the estimateNetwork function in the bootnet package (Epskamp et al., 2018) with default = "EBICglasso". Edges with |r| < .03 were treated as negligible and excluded from subsequent centrality analyses. A separate GGM was estimated for each of the 44 included countries.

Bridge Centrality

Betweenness centrality was computed for each item to quantify its role as an informational bridge within the network. Items with high betweenness centrality that also link nodes from different scales are designated bridge items. Betweenness centrality was normalized to [0, 1] using the *igraph* package (Csárdi & Nepusz, 2006). For cross-national analyses, a bridge item consistency index was computed for each country as the proportion of the overall top-5 bridge items that also appeared in that country's own top-5 bridge items, ranging from 0 (no overlap) to 1 (complete overlap).

Cross-construct Sequential Item Selection

Item pooling and bridge-first ordering. All 26 items from SLM, SCM, and SVM were pooled into a single candidate set and ranked by betweenness centrality derived from the overall GGM. This bridge-first ordering operationalizes the expectation that items occupying central bridge positions carry disproportionate cross-construct predictive information.

Sequential addition and plateau detection. For each of the three official scale indices, items were added one-by-one in bridge-first order. At each step k (k = 1, 2, ..., 26), a multinomial logistic regression model was fitted using the k selected items as predictors and the TIMSS official three-level index as the outcome. The minimum item set was defined as the smallest k at which the test-set macro-averaged AUROC reached or exceeded 99% of the highest single-item baseline AUROC achievable within the target scale. This 99% threshold was selected to allow a negligible performance decrement (< 1 percentage point) in exchange for a substantially more parsimonious instrument, consistent with scale-shortening conventions in applied measurement research. Items from scales other than the target scale were permitted and tracked separately as cross-scale items.

Model evaluation. Predictive performance was evaluated using a stratified 80/20 train–test split with stratified 5-fold cross-validation within the development set. Performance metrics on the held-out test set included: (a) macro-averaged AUROC (one-vs-rest across all three levels); (b) quadratically-weighted Cohen's kappa; and (c) overall classification accuracy. Multinomial logistic regression was implemented using the multinom function in the *nnet* package (Venables & Ripley, 2002); AUROC was computed with pROC (Robin et al., 2011).

Cross-national Consistency Analysis

The GGM estimation and sequential selection procedure were replicated independently for each of the 44 included countries. Country-level GGM similarity was quantified as the Pearson correlation between upper-triangle entries of pairs of partial correlation matrices. Hierarchical clustering (Ward's linkage) was applied to the resulting distance matrix (1 − r) to identify groups of countries with similar affective network structures. One representative country per geographic region was selected for detailed presentation based on bridge item consistency and sample size (argmax: Consistency × log(N)).

Results

Overall GGM Structure and Bridge Centrality

The GGM estimated on the full sample (N = 232,970) yielded 102 non-zero edges (|r| > .03) across the 26 items. As shown in Fig. 1, items within each scale formed coherent subnetworks while a number of cross-scale connections were evident, indicating shared conditional dependence across SLM, SCM, and SVM.

The five strongest cross-construct partial correlations are presented in Table 2. The strongest cross-scale edge linked BSBM19G ("Like to solve mathematics problems"; SLM) and BSBM22E ("I am good at working out difficult mathematics problems"; SCM; r = .111), reflecting the close empirical relationship between enjoyment of problem-solving and self-perceived competence. The second strongest cross-scale connection linked BSBM19D ("Learn many interesting things in mathematics"; SLM) and BSBM23A ("Math will help me in my daily life"; SVM; r = .107), suggesting that perceiving mathematics as interesting and perceiving it as useful are conditionally related even after accounting for all other items.

Table 2
Five Strongest Cross-construct Partial Correlations

Item 1	Scale	Item 2	Scale	Content 1	Content 2	Partial r
BSBM19G	SLM	BSBM22E	SCM	Like to solve problems	Good at difficult problems	.111
BSBM19D	SLM	BSBM23A	SVM	Learn interesting things	Helps in daily life	.107
BSBM19C	SLM	BSBM22H	SCM	Math is boring (R)	Math confuses me (R)	.105
BSBM19A	SLM	BSBM22A	SCM	Enjoy learning math	Usually do well	.093
BSBM19F	SLM	BSBM23E	SVM	Like schoolwork w/ numbers	Would like job using math	.092

Betweenness centrality values for all 26 items are displayed in Fig. 2. BSBM23E ("I would like a job that involves using mathematics"; SVM) exhibited the highest betweenness centrality (β = .174), followed by BSBM23A (β = .070), BSBM19D (β = .069), BSBM23H (β = .063), and BSBM22A (β = .056). Notably, three

of the top-five bridge items belong to SVM, despite SVM containing nine items of comparable number to SLM. This indicates that value items disproportionately serve as informational hubs connecting the three affective constructs at the item level.

Cross-construct Sequential Item Selection

Table 3 and Fig. 3 present the overall sequential selection results. The plateau criterion was defined as 99% of the highest AUROC achievable by any single own-scale item — that is, the smallest item set whose predictive accuracy falls within one percentage point of the best possible single-item performance, using only items from the target scale as the reference. This threshold reflects a deliberate trade-off: accepting a negligible loss in accuracy (< 1%p) in exchange for a substantially more parsimonious instrument. The minimum item sets required to reach this criterion differed substantially across scales: k = 7 for SLM, k = 6 for SCM, and k = 2 for SVM.

Table 3 Cross-construct Sequential Selection Results (Overall Sample, N = 232,970)

	SLM	SCM	SVM
Full items	9	8	9
Plateau k	7	6	2
Own-scale	2	2	2
Cross-scale	5	4	0
Cross %	71%	67%	0%
Baseline AUROC	.908	.839	.853
Plateau AUROC	.939	.914	.889
Δ AUROC	+.031	+.075	+.036
Plateau Kappa	.720	.641	.583
Selected items	BSBM23E, 23A, 19D, 23H, 22A, 22B, 19A	BSBM23E, 23A, 19D, 23H, 22A, 22B	BSBM23E, 23A

Table 4
The 7-item Cross-construct Battery

No.	Item	Scale	TMA Dimension	Item content	R	Predicts
1	BSBM23E	SVM	*Vision of Mathematics*	I would like a job that involves using mathematics.		SLM / SCM / SVM
2	BSBM23A	SVM	*Vision of Mathematics*	Learning mathematics will help me in my daily life.		SLM / SCM / SVM
3	BSBM19D	SLM	*Emotional Dimension*	I learn many interesting things in mathematics.		SLM / SCM
4	BSBM23H	SVM	*Vision of Mathematics*	My parents think it is important that I do well in mathematics.		SLM / SCM
5	BSBM22A	SCM	*Perceived Competence*	I usually do well in mathematics.		SLM / SCM
6	BSBM22B	SCM	*Perceived Competence*	Mathematics is harder for me than for many of my classmates.	R	SLM / SCM
7	BSBM19A	SLM	*Emotional Dimension*	I enjoy learning mathematics.		SLM

Note. Items are presented in bridge-first order (ranked by GGM betweenness centrality). R = reverse-scored. Predicts = TIMSS official scale indices for which the item contributes to the cross-construct plateau set.

Three findings are noteworthy. First, the plateau item sets for SLM and SCM were dominated by cross-scale items. For SLM prediction, five of the seven plateau items (71%) originated from SVM or SCM. For SCM prediction, four of the six items (67%) came from other scales — and, critically, the first four items added in the bridge-first order were all cross-scale items (k = 1–4: own-scale n = 0), demonstrating that student confidence in mathematics is more efficiently predicted by items assessing interest and value than by confidence items alone. This constitutes the strongest evidence for cross-construct predictive information in the present dataset.

Second, plateau item sets showed substantial overlap across scales: the same six items (BSBM23E, BSBM23A, BSBM19D, BSBM23H, BSBM22A, BSBM22B) were selected for both SLM and SCM prediction, with SLM additionally requiring BSBM19A. This means that a single compact instrument of six to seven items in Table 4 — rather than 26 — can simultaneously screen students across all three affective dimensions.

Third, SVM reached plateau at k = 2, with both selected items drawn from within the SVM scale itself (BSBM23E: "I would like a job that involves using mathematics"; BSBM23A: "Math will help me in my daily life"). This stands in sharp contrast to SLM and SCM, and indicates that mathematics value beliefs are structured as a relatively self-contained construct, requiring no cross-scale information for efficient

prediction. The contrast between SVM (k = 2, 0% cross-scale) and SCM (k = 6, 67% cross-scale) suggests meaningful differences in the degree to which these constructs are embedded within the broader affective network.

Plateau AUROC values ranged from .889 (SVM) to .939 (SLM), all substantially above the .80 threshold conventionally used to indicate acceptable discriminative accuracy in classification research. Compared with the single-item baseline, AUROC gains from cross-construct integration were modest for SLM (Δ = +.031) and SVM (Δ = +.036) but meaningful for SCM (Δ = +.075), further supporting the interpretation that student confidence benefits most from cross-construct information.

Cross-national Consistency

Separate GGMs and sequential selection analyses were conducted for each of the 44 included countries. The bridge item consistency index — the proportion of the overall top-5 bridge items reproduced in each country's own top-5 — had a mean of .509 (SD = .158) across countries, indicating that approximately half of the overall bridge structure is reliably recovered at the country level.

Four countries achieved the highest consistency index of .80: Czech Republic (CZE; Europe), Korea (KOR; East Asia), New Zealand (NZL; Oceania), and the United States (USA; Americas). Despite their differing geographic and cultural contexts, all four countries exhibited similar patterns: cross-scale items contributed to SLM and SCM prediction at early steps, while SVM reached plateau rapidly with own-scale items only. Table 4 summarises the plateau k and AUROC for each country and scale. At the other extreme, consistency indices of .20 were observed for Saudi Arabia (SAU), Romania (ROM), Turkey (TUR), and Uzbekistan (UZB), suggesting that the overall bridge structure generalizes less readily to these contexts. Countries with lower consistency are discussed as boundary conditions in the Discussion.

Table 4
Representative Country Results (top-4 by bridge item consistency index)

Region	CZE	KOR	NZL	USA
	Europe	East Asia	Oceania	Americas
N	6,915	3,916	3,115	6,572
Consistency	.80	.80	.80	.80
SLM AUROC	.931	.927	.952	.952
SLM k	5	3	7	5
SCM AUROC	.886	.920	.899	.900
SCM k	4	2	10	7
SVM AUROC	.880	.902	.864	.886
SVM k	4	4	2	4

Korea (KOR) showed the most parsimonious SCM prediction (plateau k = 2), suggesting that in the Korean context, student confidence in mathematics can be reliably screened with only two items. The United States required the most items for SCM (k = 7), reflecting a somewhat more distributed predictive structure. Across all four representative countries, SVM plateau k did not exceed 4, consistent with the overall finding that mathematics value beliefs are more efficiently captured than interest or confidence.

Plateau AUROC values were high and relatively consistent across countries (Fig. 4: median AUROC across all 44 countries and three scales ranged from .87 to .95), suggesting that the predictive utility of the bridge item framework generalizes well internationally even where the specific bridge items selected differ.

Discussion

This study set out to examine whether a minimum set of cross-construct bridge items could simultaneously reproduce the TIMSS official three-level classification of student interest, confidence, and value in mathematics — the three dimensions of Di Martino and Zan's (2010) Three-dimensional Model of Attitude (TMA). The answer is affirmative, and the results carry both theoretical and practical implications. Three findings stand out.

The first concerns the role of mathematics value beliefs. Items from the SVM scale — operationalizing the Vision of Mathematics dimension of the TMA — occupied three of the seven plateau positions and anchored the bridge-first ordering, appearing before any own-scale items in both the SLM and SCM selection sequences. Theoretically, this asymmetry is consistent with Eccles and colleagues' expectancy-value model (Wigfield & Eccles, 2000), in which task value beliefs are posited to mediate the relationship between competence expectations and engagement outcomes. Items such as "I would like

a job that involves using mathematics" and "Learning mathematics will help me in my daily life" appear to carry cross-construct signal precisely because they encode not only value judgements but implicit information about interest and self-perception: a student who aspires to a mathematics-related career is unlikely to simultaneously dislike the subject or feel incompetent. Di Martino and Zan (2010) proposed the three TMA dimensions as conceptually distinct but mutually influencing, without specifying which dimension serves as the primary integrative node. The present GGM evidence identifies the Vision of Mathematics as that node — at least as operationalized in TIMSS items — contributing new empirical content to the TMA framework.

The second finding concerns the internal structure of confidence prediction. Student confidence in mathematics was the most cross-construct dependent of the three dimensions: the first four items added in bridge-first order for SCM prediction were all drawn from SLM or SVM, and confidence items only entered the set at the fifth and sixth steps. This means that knowing a student's interest and value orientation predicts their confidence classification more efficiently than does any single confidence item alone. The interpretation is straightforward within the TMA framework: perceived competence is not an isolated self-assessment but is shaped by emotional experience and utility beliefs (Di Martino & Zan, 2010), and this co-variation is captured more efficiently by items that straddle construct boundaries than by items embedded within a single cluster. The practical implication is that instruments assessing student confidence in isolation may overlook cross-construct predictive information that is readily accessible through a small number of well-positioned bridge items.

Third, in sharp contrast to confidence, mathematics value beliefs were the most structurally self-contained dimension: the SVM plateau was reached at just two items, both from within the SVM scale itself. This independence is theoretically plausible. Utility and relevance beliefs — the primary content of SVM items — are cognitively distinct from affective enjoyment and from performance self-assessment; a student may strongly value mathematics for career reasons without enjoying it particularly or feeling competent. This separability has been observed in expectancy-value research (e.g., Guo et al., 2015) and is reflected here in the relatively weaker cross-construct edges connecting SVM items to SLM and SCM nodes in the GGM. For applied purposes, it implies that a two-item SVM screener achieves acceptable accuracy (AUROC = .889) independently of any interest or confidence items — a finding with direct implications for survey economies when only value orientation is of interest.

Cross-national replication yielded a more nuanced picture. The bridge item consistency index — the proportion of overall top-5 bridge items reproduced in each country's own top-5 — averaged .509 (SD = .158) across 44 countries, indicating that approximately half of the overall bridge structure is reliably recovered at the country level. However, plateau AUROC values were uniformly high across countries regardless of consistency level, ranging from .87 to .95 across scales and countries. This dissociation between structural consistency and predictive stability has a practically important implication: the seven-item battery performs well as a screening instrument across diverse national contexts even where the specific bridge items driving prediction differ. Four countries achieved the highest observed consistency of .80 — Czech Republic, Korea, New Zealand, and the United States — and their sequential addition

curves closely replicated the overall pattern, supporting the robustness of the framework in these contexts.

Countries with lower consistency indices — notably Saudi Arabia, Romania, Turkey, and Uzbekistan (consistency ≤ .20) — represent meaningful boundary conditions of the framework rather than analytical failures (Gao et al., 2026). In these contexts, cultural and institutional factors may structure the conditional dependencies among affective items differently: where mathematics engagement is predominantly externally regulated — by family expectations, examination pressures, or vocational imperatives — the mutual reinforcement of interest, confidence, and value beliefs posited by both the TMA and the expectancy-value model may be attenuated (cf. Ryan & Deci, 2000). Hierarchical clustering of country GGMs further revealed that structural similarity across countries does not align cleanly with geographic proximity, suggesting that instructional culture and curriculum structure may play a more determining role than regional identity in shaping the affective network. Practitioners in lower-consistency countries are advised to treat the seven-item battery as a starting point and to validate it with local data before operational use.

The convergence of SLM and SCM plateau item sets into a shared six-item core, with a seventh item extending coverage to SLM, yields a single instrument that replaces 26 items across three scales while retaining discriminative accuracy. In analogy to Lim and colleagues' (2026) application of cross-diagnostic optimization in digital mental health monitoring — where a compact item set enables more frequent and sustainable symptom labeling — the seven-item battery enables periodic re-administration across a school year without the survey burden of full-scale administration. Teachers could administer the battery at the beginning of each term to track changes in student affect across all three TMA dimensions simultaneously, supporting timely pedagogical response. To reduce barriers to adoption, this study provides a Python module with embedded model coefficients for batch prediction in research or administrative contexts, and a browser-based tool through which individual students can self-administer the battery and receive immediate probabilistic feedback, available at [repository URL].

Limitations

Three limitations warrant acknowledgement. The plateau threshold of 99% is operationally defined; while it produced stable and interpretable results, sensitivity analyses comparing alternative thresholds would strengthen the justification for this criterion and are planned for revision if requested by reviewers. The multinomial logistic regression models assume linear additive effects, and more flexible classifiers might achieve higher accuracy in countries with complex affective structures, at the cost of the interpretability and deployability that motivated the present modeling choice. Finally, TIMSS sampling weights were not applied, meaning that larger countries contribute proportionally more to the overall GGM; future work should examine whether weighted estimation alters bridge centrality rankings in substantively meaningful ways.

Conclusion

This study set out to examine whether the three TIMSS mathematics affective scales — SLM, SCM, and SVM — contain items that carry cross-construct predictive information, and whether that information can be exploited to produce a principled, compact instrument for simultaneous monitoring of student affect across all three dimensions. The answer on both counts is affirmative, and the findings carry implications for the measurement of mathematics attitude in large-scale international assessments.

The most consequential empirical finding is the identification of SVM items as the structural hub of the mathematics affective network. Three of the seven cross-construct plateau items originated from SVM, and the two SVM items with the highest betweenness centrality — BSBM23E ('I would like a job that involves using mathematics') and BSBM23A ('Learning mathematics will help me in my daily life') — were the first items selected for SLM, SCM, and SVM prediction alike. This asymmetry was not imposed by design; it emerged from the conditional dependence structure of the GGM estimated on 26 items across N = 232,970 students. Di Martino and Zan (2010) proposed the three TMA dimensions as mutually influencing but left open the question of which dimension serves as the primary integrative node. The present GGM evidence answers that question at the item level: the Vision of Mathematics dimension, operationalized in TIMSS through utility and relevance beliefs, functions as the hub through which interest and confidence information is most efficiently transmitted. This is consistent with expectancy-value theory's prediction that value beliefs carry distinctive cross-domain information, and with the observation by Author (2017) that the relationship between valuing mathematics and reasoning outcomes is mediated by the other two affective dimensions.

The structural contrast between SVM and the other two dimensions is equally informative. SCM required the most items (k = 6) and relied most heavily on cross-scale information, with the first four items in bridge-first order all drawn from SLM or SVM — meaning that student confidence is more efficiently captured by items assessing interest and utility value than by confidence items alone. This finding reflects the theoretically expected interdependence of competence beliefs and value beliefs in the expectancy-value framework. SVM, by contrast, reached plateau at k = 2 with only own-scale items, indicating that utility value beliefs are sufficiently self-contained to be screened with two items independently of interest and confidence information. These structural differences between dimensions were consistent across the majority of the 44 included countries, with moderate average bridge item consistency (mean = .509), although predictive accuracy was high and stable cross-nationally regardless of bridge item overlap (median AUROC .87–.95).

The practical output of the study is a seven-item cross-construct battery that replaces 26 items across three scales with an instrument requiring approximately two to three minutes to complete. Validated on N = 232,970 students from 44 countries, the battery achieves AUROC values of .889 to .939 for simultaneous prediction of High, Intermediate, and Low classifications on all three TIMSS indices. Implementation resources — a Python module with embedded model coefficients for batch institutional use and a browser-based self-assessment tool for individual students — are provided as supplementary materials. These resources lower the barrier to periodic affective monitoring in educational contexts where the full 26-item administration is impractical.

Several boundary conditions should be noted. Countries with lower bridge item consistency indices — notably Saudi Arabia, Romania, Turkey, and Uzbekistan — represent contexts where the overall bridge structure generalizes less readily, likely reflecting cultural and institutional differences in the organization of mathematics affect. In these contexts, the seven-item battery should be treated as a starting point requiring local validation rather than a directly deployable instrument. The plateau threshold of 99% is operationally defined rather than statistically derived, and sensitivity analyses across alternative thresholds are a natural extension of the present work. Finally, the multinomial logistic regression models used for prediction assume linearity, and more flexible classifiers may yield modest gains in accuracy, particularly in structurally divergent contexts.

Taken together, these findings advance the measurement of mathematics affect in large-scale assessments in three ways: by providing the first item-level empirical mapping of the conditional dependence structure across TIMSS SLM, SCM, and SVM; by demonstrating that cross-construct bridge items — predominantly from the Vision of Mathematics dimension — carry sufficient information to screen students on all three TMA dimensions simultaneously; and by delivering a validated compact instrument with open-access implementation tools. Future work should examine whether the bridge item structure identified here replicates in TIMSS Grade 4 data, in other large-scale assessments such as PISA, and longitudinally across TIMSS cycles, to determine the stability of the Vision of Mathematics hub role across developmental stages and over time.

Declarations

Ethics approval and consent to participate

This study used secondary analysis of publicly available de-identified data from the TIMSS 2023 international database (IEA). No individual student data were collected by the authors. Ethical approval was not required.

Consent for publication

Not applicable.

Availability of data and materials

The TIMSS 2023 international database is publicly available at https://www.iea.nl/data-tools/repository/timss.

Competing interests

The author declares no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

[Blinded]

Use of AI tools

AI-assisted tools were used for code generation and copy editing during the preparation of this manuscript. All analytical decisions, theoretical interpretations, and final text are the sole responsibility of the author.

References

1. Author. (2017). *Title*.
2. Author. (2022). *Title*.
3. Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1*, 58. https://doi.org/10.1038/s43586-021-00055-w
4. Campbell, T. G., Vela, K. N., & Powell, T. (2025). Interest, enjoyment, and confidence in mathematics in the United States: exploring patterns across age, gender, race, and time. *Large-scale Assessments in Education*, *13*(1), 22. https://doi.org/10.1186/s40536-025-00258-7
5. Campos, M., Eryilmaz, N., & Strietholt, R. (2025). Beyond achievement gaps: inequalities in affective components of math learning. *Large-scale Assessments in Education*, *13*(1), 12. https://doi.org/10.1186/s40536-025-00249-8
6. Choi, H. W., & Choi, Y.-J. (2025). Validity study on the students' attitudes toward mathematics scale for English-speaking countries. *SAGE Open, 15*(2). https://doi.org/10.1177/21582440251339984
7. Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695, 1–9*. https://igraph.org
8. Di Martino, P., & Zan, R. (2003). What does 'positive' attitude really mean? In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 27th Conference of PME* (Vol. 4, pp. 451–458). University of Hawai'i.
9. Di Martino, P., & Zan, R. (2010). 'Me and maths': Towards a definition of attitude grounded on students' narratives. *Journal of Mathematics Teacher Education, 13*(1), 27–48. https://doi.org/10.1007/s10857-009-9134-z
10. Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

11. Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1
12. Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617–634. https://doi.org/10.1037/met0000167
13. Gao, H., Evans, T., & Brown, G. T. L. (2026). The relations between motivation and mathematics achievement: A cross-cultural study of Asia-Pacific jurisdictions using TIMSS 2023 data. *ZDM – Mathematics Educatio*n. https://doi.org/10.1007/s11858-026-01773-9
14. Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology, 51*(8), 1163–1176. https://doi.org/10.1037/a0039440
15. Li, M., Johnson, T., & Adegoke, A. S. (2024). The structure of psychosocial factors in academic success: A gaussian graphical model approach. *Social Psychology of Education*, *27*(6), 3175-3201. https://doi.org/10.1007/s11218-024-09917-z
16. Lim, M., Suh, S., Byeon, J., et al. (2026, April 28). *One item per disorder: A cross-diagnostic optimization framework for simultaneous prediction of six psychiatric disorders* (Version 1) [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-9356755/v1
17. Michaelides, M. P., Brown, G. T. L., Eklöf, H., & Papanastasiou, E. C. (2019). *Motivational profiles in TIMSS mathematics: Exploring student clusters across countries and time* (IEA Research for Education, Vol. 7). Springer. https://doi.org/10.1007/978-3-030-26183-2
18. Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2024). *TIMSS 2023 international results in mathematics and science.* TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2023/
19. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77. https://doi.org/10.1186/1471-2105-12-77
20. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68
21. Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://doi.org/10.1007/978-0-387-21706-2
22. Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Figures

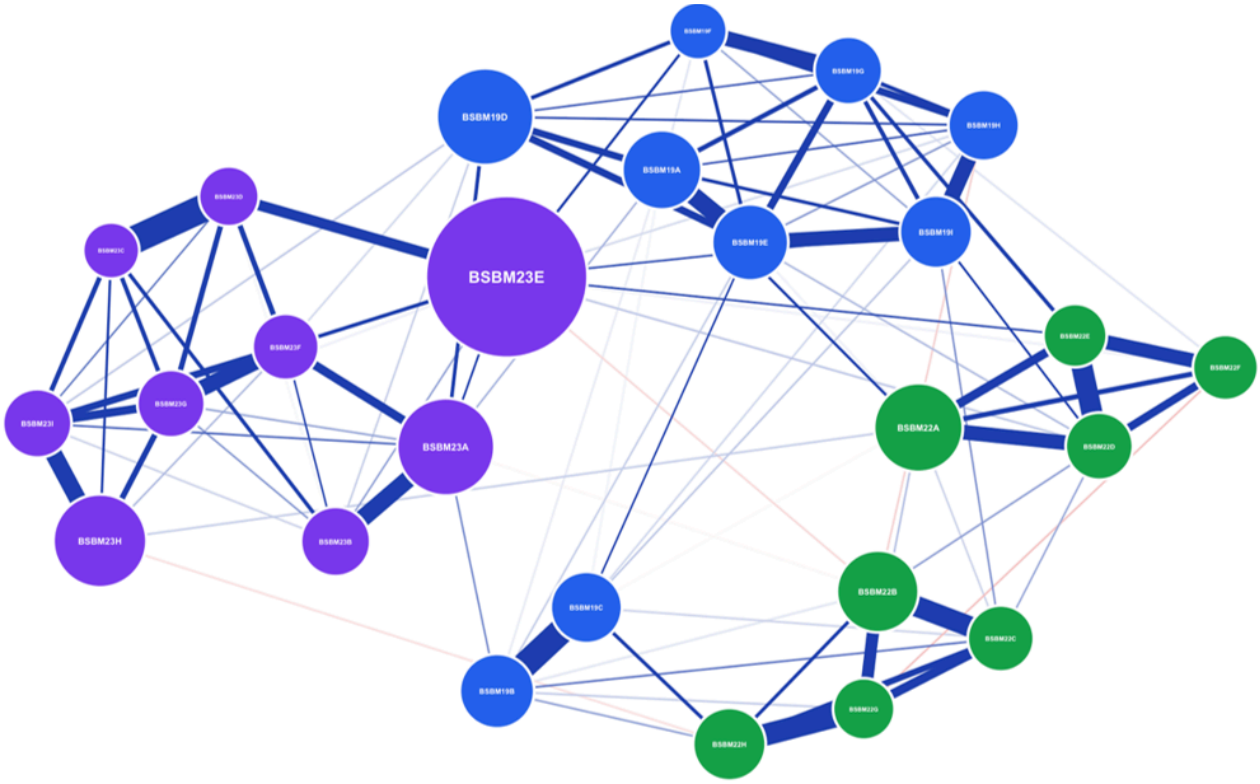


Figure 1

Overall GGM network of TIMSS 2023 Mathematics Affective Sca

Note. Node size is proportional to betweenness centrality; edge width is proportional to partial correlation.

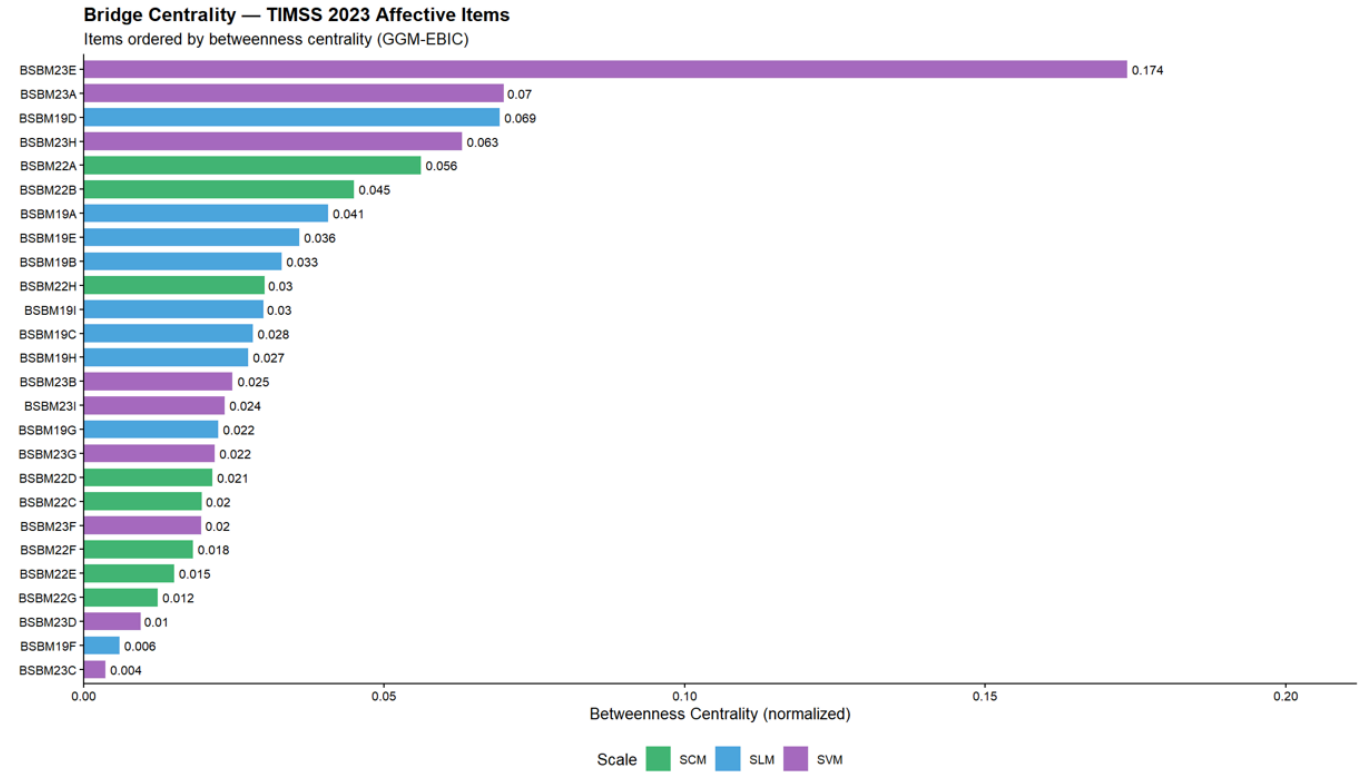


Figure 2

Betweenness centrality for all 26 TIMSS affective items.

Note. Items are ordered from highest to lowest. Color indicates original scale (SLM = blue; SCM = green; SVM = purple). The five highest-centrality items are highlighted as bridge items.

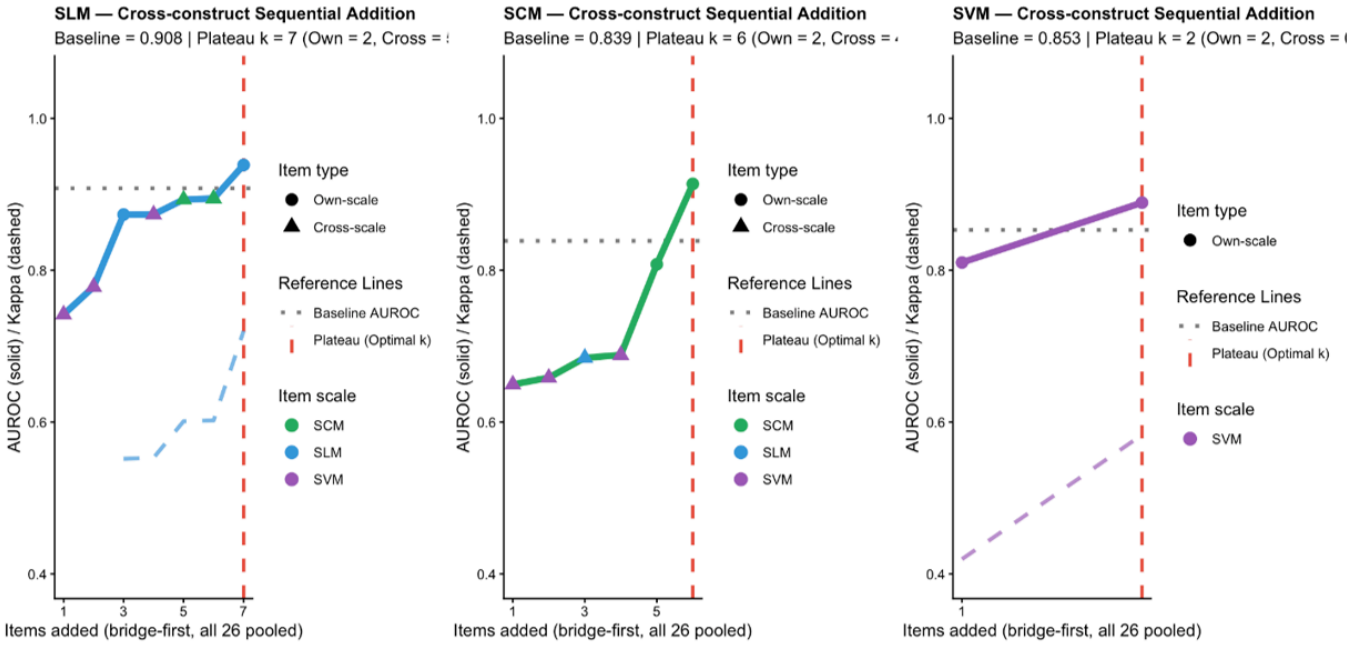


Figure 3

Cross-construct sequential addition curves for SLM, SCM, and SVM

Note. Solid line = macro-averaged AUROC; dashed line = quadratic weighted Kappa. Circles = own-scale items added; triangles = cross-scale items added. The vertical dashed line marks the plateau k.

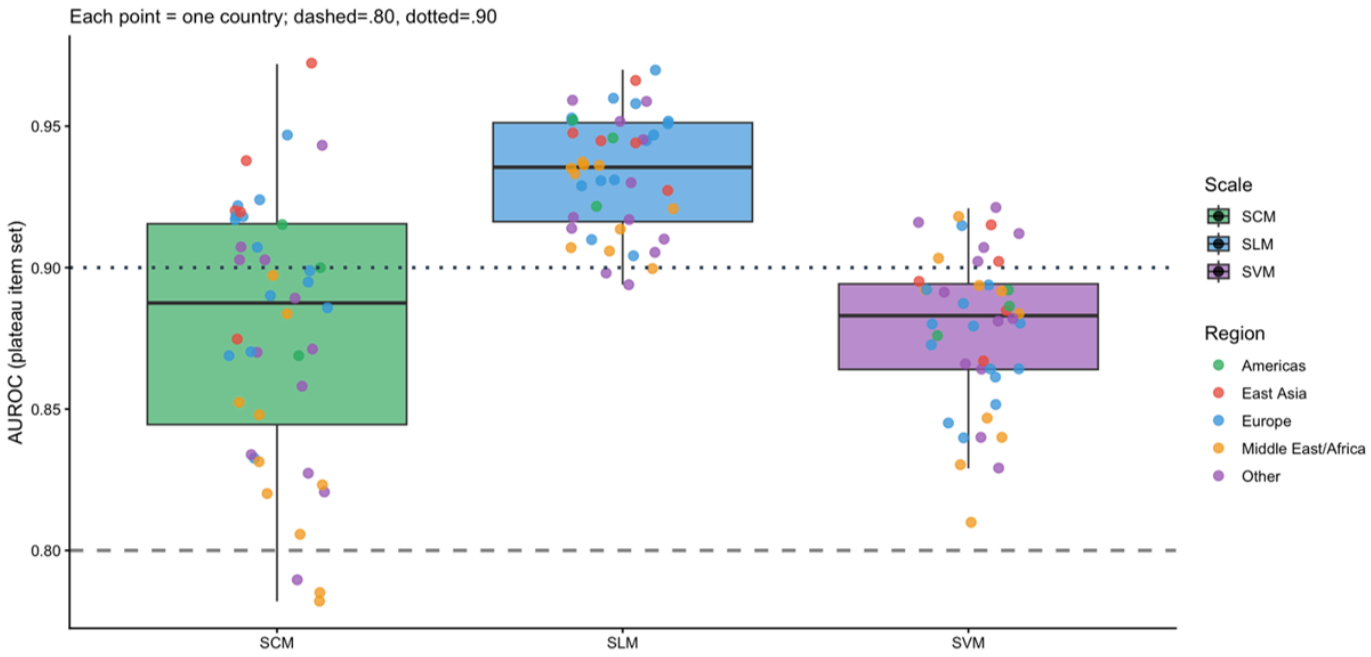


Figure 4

Distribution of plateau AUROC values across 44 countries by scale (SLM, SCM, SVM)

Note. The dashed horizontal line indicates AUROC = .80; the dotted line indicates AUROC = .90. Points are color-coded by geographic region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- timssmathaffectpredictor.py
- timssanalysisclean.html