

Throwing Out the Bathwater but Keeping the Baby: Extending Campbell-Fiske's Multitrait-Multimethod Framework

Advances in Methods and Practices in Psychological Science
April-June 2026, Vol. 9, No. 2,
pp. 1–24
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459251407652
www.psychologicalscience.org/AMPPS



Herbert W. Marsh^{1,2} , Jiesi Guo¹ , Oliver Lüdtke³ ,
and Reinhard Pekrun^{1,4,5} 

¹Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia; ²Department of Education, Oxford University, Oxford, England; ³Department of Educational Measurement, Leibniz Institute for Science and Mathematics Education, Kiel, Germany; ⁴Department of Psychology, University of Essex, Colchester, England; and ⁵Department of Psychology, Ludwig Maximilian University, Munich, Germany

Abstract

The 65th anniversary of Campbell and Fiske's multitrait-multimethod (MTMM) framework provides a timely opportunity to revisit and modernize this foundational model for construct validation. Although structural-equation-modeling-based MTMM approaches have enhanced the field, their widespread application remains constrained by convergence problems, ambiguous trait-method distinctions, and a lack of consensus regarding optimal model specifications. We propose an extended Campbell-Fiske framework that resolves these limitations while preserving the original guidelines' conceptual strengths. Our key innovation is to apply the MTMM logic to a fully latent multitrait-multidomain correlation matrix derived from a rigorously tested multiple-indicator measurement model. Our approach treats traits and methods (i.e., domains, occasions, informants, contexts, or another method facet) as fully symmetrical, substantive facets, eliminates reliance on manifest correlations, corrects for measurement error, and introduces formal asymptotic parameter comparisons to test each validity criterion. This framework provides a formatively heuristic structure that retains the original appeal of the MTMM logic for applied research while meeting current psychometric standards for transparency, reproducibility, and inferential rigor expected by leading academic journals. We illustrate the method using a large, multidimensional data set ($N = 18,047$), but the framework generalizes across domains of psychological science. The extended framework offers applied researchers a flexible, powerful tool for evaluating convergent and discriminant validity, diagnosing trait-domain interactions, and clarifying measurement quality. By "keeping the baby" while refreshing the empirical implementation, our approach affirms the enduring value of the Campbell-Fiske logic while aligning it with the demands of modern research practice.

Keywords

multitrait-multimethod (MTMM) analysis, extended Campbell-Fiske framework, latent correlations, convergent and discriminant validity, trait-method symmetry

Received 6/14/25; Revision accepted 10/17/25

Construct validity remains one of psychology's most enduring concerns, foundational to assessing traits, motivations, and abilities across clinical, educational, and organizational contexts (Cronbach & Meehl, 1955; Marsh, 2007). Despite significant psychometric advances, applied researchers continue to struggle with transparent and rigorous validation strategies—especially in complex designs involving multiple traits, informants, or time

points (Campbell & O'Connell, 1967; Lance et al., 2010; Lance & Vandenberg, 2010). A central challenge is disentangling target constructs from extraneous method

Corresponding Author:

Herbert W. Marsh, Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia
Email: Herb.Marsh@acu.edu.au



variance (Eid, 2006; Podsakoff et al., 2003), particularly when researchers aim to establish both convergent and discriminant validity simultaneously.

The multitrait-multimethod (MTMM) paradigm, introduced by Campbell and Fiske (1959), directly addressed these concerns by offering a matrix-based structure and five guidelines for evaluating construct validity. This approach remains highly cited and widely used (Eid & Nussbeck, 2009; Kenny, 2022), although difficulties in implementation and interpretation have grown as expectations for statistical precision have evolved (Marsh, 1988; Millsap, 2014).

To advance this legacy, we present the extended Campbell-Fiske (EC-F) framework, which integrates the conceptual logic of the original guidelines with contemporary expectations for psychometric rigor. Our EC-F framework retains the matrix-based reasoning of Campbell and Fiske (1959) while embedding it within latent-variable modeling, allowing for more defensible inferences in applied validation contexts.

Legacy and Logic of the Campbell-Fiske Guidelines

The year 2024 marked the 65th anniversary of Campbell and Fiske's (1959) seminal article on the MTMM paradigm. Campbell and Fiske's article introduced the MTMM matrix as a systematic approach to construct validation via analysis of multiple traits measured across multiple methods. Their core contribution was articulating expected correlation patterns to support convergent and discriminant validity—providing a heuristic lens to separate true construct variance from method effects.

Although not formalized as a statistical model, the guidelines offered intuitive appeal and diagnostic value, especially for applied researchers. However, reliance on manifest correlations introduced several limitations: inability to separate error from true variance, lack of formal testing, and blurred distinctions between trait and method variance. These issues prompted decades of methodological refinements (Marsh, 1988; Millsap, 1990). The distinction between true and error variance is sharpened in latent-variable models, in which trait-method correlations can be disaggregated from measurement error. We explicitly build on Marsh (1989) and Kenny and Kashy (1992), who emphasized that manifest indicators often obscure true construct-method variance partitioning.

Later extensions (e.g., Marsh, 1988; Millsap, 1990) added criteria such as profile consistency and mean-level agreement across methods, broadening the guidelines' diagnostic scope. These refinements underscore the need for flexible models that preserve the guidelines' conceptual clarity while addressing their statistical limitations.

Ambiguity in What Constitutes a Method Effect

MTMM:structural equation models (SEMs) reflect a two-facet design prioritizing the trait facet over the method facet. However, a persistent complexity in MTMM designs is the ambiguous boundary between trait and method facets. Campbell and Fiske (1959) acknowledged this, noting that “what is an unwanted response set for one tester may be a trait for another” (p. 85; also see Cronbach & Meehl, 1955; Lance et al., 2010; Lance & Vandenberg, 2010). For instance, in studies involving multiple raters, time points, or subject domains, both facets may represent substantive constructs rather than bias or error.

This ambiguity challenges assumptions that one facet is merely “noise.” In many real-world applications, both traits and methods carry meaningful information. Accordingly, our EC-F framework treats both trait and domain/method facets as conceptually symmetric and structurally parallel—a crucial advance for applied construct-validation models, especially when domains such as academic subject, informant, or time are methodologically meaningful. As we elaborate in Extension 3, our EC-F framework explicitly acknowledges this ambiguity by treating both trait and domain/method facets as conceptually and analytically equivalent—a crucial step for construct validity in applied social-science studies.

Dilemma for Basic and Applied Researchers

The Campbell-Fiske (1959) guidelines (Fig. 1) gained widespread use because of their intuitive appeal and heuristic value, particularly in applied-research settings. However, they relied on observed (manifest) correlations that conflate true trait-method relationships with measurement error, provided no formal significance tests, and often blurred the boundary between traits and methods in complex designs. These limitations have persisted in practice, motivating decades of methodological refinement. In response, researchers developed various SEM-based approaches to the MTMM design. These models are powerful tools for decomposing trait and method variance and testing specific hypotheses in MTMM studies (e.g., Byrne, 1998, 2012; Eid, 2006; Helm, 2022; Marsh & Hocevar, 1988). Advances in statistical software (e.g., Mplus, Lavaan, R) and methodological innovations have made SEM-based MTMM models more accessible to applied researchers, supporting a more rigorous and nuanced analysis of trait-method structures.

Among these approaches, the correlated-trait-correlated-method (CT-CM) model was the original

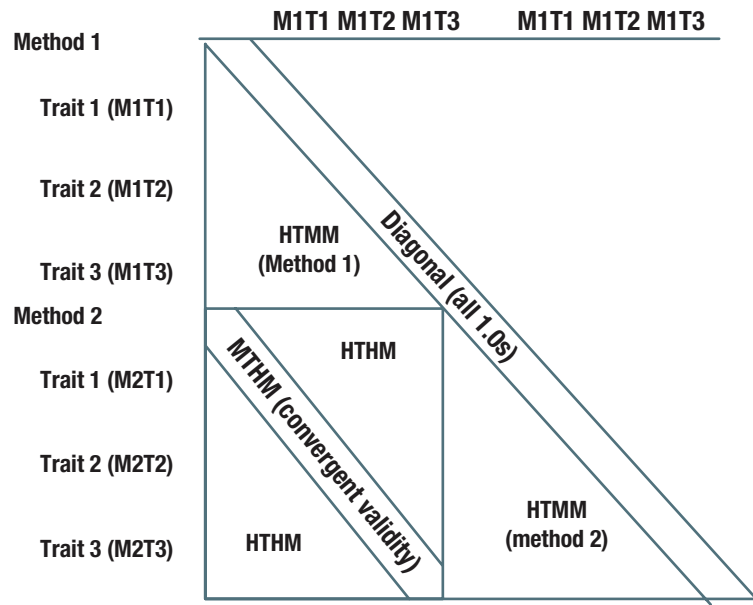


Fig. 1. Conceptual diagram of a multitrait-multimethod matrix illustrating Campbell-Fiske (1959) guidelines. Each variable represents a combination of one trait (T1, T2, T3) and one method (M1, M2, M3). The Campbell-Fiske guidelines considered here are as follows. Guideline 1 is convergent validity, in which correlations between the same trait measured by different methods (MTHM) should be substantial. This indicates that the construct is consistently captured across different measurement approaches. Guideline 2 is discriminant validity-trait emphasis, in which convergent-validity (MTHM) correlations should be stronger than HTHM correlations. This ensures that the instruments measure distinct traits and that the observed associations are not simply due to shared method variance. Guideline 3 is discriminant-validity-method emphasis, in which convergent-validity (MTHM) correlations should also be stronger than HTMM correlations. This demonstrates that methods effects (e.g., response style or rater tendencies) do not overshadow the distinctiveness of the traits being measured. Guideline 4 is pattern consistency across methods, which proposes that valid trait constructs should show a consistent pattern of intercorrelations across different methods. To operationalize this, Marsh (1988) and Millsap and Everson (1993) introduced the profile-similarity index (PSI) as a descriptive indicator of structural consistency. PSI is computed as the Pearson correlation between vectors of intertrait correlations obtained separately within each method. For each method, the correlations among traits are arranged into a vector (excluding the diagonal). PSI values are then computed by correlating these vectors across methods. A high PSI (approaching +1.0) indicates that the configuration of trait interrelations is preserved across methods, supporting structural generalizability and the internal coherence of the constructs (for more detail, see Section S3 in the Supplemental Material available online). Guideline 5, mean consistency of correlations across methods, is a corollary to Campbell and Fiske’s Guideline 4, meaning that HTMM correlations among traits within each method (HTMM, e.g., Method 1) should also be similar across methods (HTMM, Method 2). Whereas the Guideline 4 focuses on the pattern of correlations, large mean differences suggest that method effects are substantial for at least some methods; small mean differences suggest the size of method effects is similar across methods or absent altogether. MTHM = monotrait-heteromethod (same trait, different methods); HTHM = heterotrait-heteromethod (different traits, different methods); HTMM = heterotrait-monomethod (different traits, different methods).

model (Jöreskog, 1971; Kenny, 1979) and is often depicted as the conceptual “gold standard” (e.g., Eid, 2006; Marsh, 1989; Widaman, 2022). It preserves the fundamental symmetry between trait and method factors envisioned by Campbell and Fiske (1959), allowing all

traits and methods to be estimated simultaneously without structural prioritization. In this sense, it aligns most closely with the original MTMM logic, particularly in designs in which both trait and method facets are substantively meaningful.

However, in practice, the CT-CM model presents serious estimation challenges, including frequent nonconvergence, negative residual variances, and non-positive-definite covariance matrices—even in simulated data sets specifically constructed to support it (e.g., Castro-Schilo et al., 2014; Grayson & Marsh, 1994; Marsh, 1989; Marsh & Grayson, 1991, 1995). Thus, although the CT-CM model serves as a conceptual “gold standard,” it is not an empirical “gold standard.” These technical difficulties have led many researchers to adopt more constrained alternatives—such as the correlated-trait-uncorrelated-method (CT-UM) model, which constrains methods to be uncorrelated; the correlated-trait-correlated-method-minus-1 (CT-C[M-1]) model, which eliminates one of the method factors; and the correlated-trait-correlated-uniqueness (CT-CU) model, which represents methods as correlated uniquenesses. Although these alternative models can improve estimation stability and convergence, they often do so at the cost of constraining the representation of method effects and compromising the conceptual symmetry that makes the CT-CM model (and the Campbell-Fiske, 1959, guidelines; see Fig. 1) appealing. Furthermore, even among experts, there is no consensus on which SEM is most appropriate (e.g., Helm, 2022).

This presents an enduring dilemma: On the one hand, applied researchers are encouraged to adopt sophisticated SEM-based MTMM frameworks to meet contemporary psychometric expectations; on the other hand, the technical demands, estimation fragility, and lack of consensus across models make these approaches challenging to implement in basic and applied research. As a result, some researchers revert to the original Campbell and Fiske (1959) guidelines—despite their acknowledged limitations—because of their intuitive structure, conceptual accessibility, and diagnostic utility, or they abandon the MTMM paradigm altogether.

We emphasize that the purpose of this article is not to juxtapose the strengths and weaknesses of alternative SEM-based MTMM models or to position our EC-F framework as a replacement for SEM-based approaches. Instead, our goal is to highlight the complementary role that our EC-F framework can play in bridging the gap between the technical rigor of SEM-based models and the practical interpretability and diagnostic utility offered by the original Campbell-Fiske (1959) framework. We argue that our EC-F framework is sufficiently rigorous and robust to stand alone in both basic and applied research while also serving as a valuable complement to more statistically oriented SEM studies. Indeed, applying our EC-F framework—rooted in a fully latent measurement model—leads naturally to pursuing more complex SEMs. Both perspectives are needed for a comprehensive evaluation of construct validity, particularly in applied-research contexts in

which large-scale survey data and complex measurement structures often coexist.

Our EC-F Framework

Our EC-F framework comprises three key extensions that modernize and strengthen the classic Campbell-Fiske guidelines (Campbell & Fiske, 1959; see Fig. 1) while retaining their conceptual clarity. Figure 1 illustrates the foundational structure of a hypothetical 3 (Trait) \times 2 (Method) correlation matrix and the relevant Campbell-Fiske guidelines categories that inform our extensions. We intend these extensions to be stand-alone contributions, offering a specific advance that is rigorous and accessible for applied researchers.

Extension 1: a fully latent matrix via a measurement model with multiple indicators

The first enhancement involves specifying and testing a fully latent measurement model that captures each trait-facet combination using multiple observed indicators. This measurement model tests the underlying factor structure and corrects for measurement error at the outset, ensuring that inferences are based on properly specified latent constructs. The multiple-indicator strategy, initially proposed by Marsh and Hocevar (1988; also see Castro-Schilo et al., 2014; Eid, 2006; Geiser et al., 2014; Kenny, 2022), remains underused despite its critical importance. Our approach explicitly tests the measurement model, thereby avoiding the implicit assumptions and potential biases associated with single-item or composite-score representations. This strategy employs a measurement model based on confirmatory factor analysis (CFA), aligning seamlessly with SEM frameworks to ensure methodological rigor and accessibility. Once the measurement model is estimated, the resulting latent correlation matrix—the MTMM matrix—is the basis for evaluating the Campbell-Fiske guidelines and their statistical extensions (Fig. 1).

Extension 2: statistical tests of the Campbell-Fiske guidelines

The second key enhancement transforms the original descriptive Campbell-Fiske guidelines into formal statistical hypotheses. As part of testing the measurement model, we first estimate a fully latent correlation matrix using a CFA-based measurement model. This ensures that the model fits well and the latent constructs are appropriately measured and account for measurement error. From this latent MTMM matrix, we then specify

formal statistical tests of the Campbell-Fiske criteria (Fig. 1)—using asymptotic parameter comparisons (e.g., Oehlert, 1992; Raykov, 2011).

In practical terms, we employ Mplus's (Muthén & Muthén, 2017) model-constraint function to compute aggregate estimates (e.g., the mean of the convergent-validity correlations) and compare these with other aggregates (e.g., the mean of heterotrait-heteromethod [HTHM] correlations). These comparisons can be conducted separately for each trait or domain and then averaged across all traits or domains (for the Mplus syntax, see the Supplemental Material available online). Although these calculations could also be performed manually outside the model, the asymptotic parameter-comparison approach provides standard errors and significance tests, enabling researchers to formally test each of the Campbell-Fiske guidelines within the measurement model itself.

This formalization enhances transparency, replicability, and diagnostic clarity while preserving the intuitive interpretability of the latent-correlation matrix illustrated in Figure 1. The approach is straightforward to implement with contemporary statistical software, leverages robust-estimation methods, and ensures that the tests are both statistically valid and accessible to applied researchers.

Extension 3: clarifying symmetry between facets

The third enhancement reinforces the importance of treating the MTMM structure as a fully symmetrical two-facet design. Historically, even the application of the Campbell-Fiske guidelines—although acknowledging the potential overlap between trait and method facets—often treated method variance as subordinate (see earlier discussion). This tendency has persisted in many SEM-based MTMM models, in which method variance is often constrained or treated as residual error.

In contrast, our EC-F framework treats both facets—traits and the second design dimension—as fully symmetric and analytically equivalent, consistent with the fluidity emphasized by Campbell and Fiske (1959). In many studies, the second facet may represent a dimension that deserves to be treated as substantively meaningful rather than simply labeled as “M = method.” Examples include multitrait-multioccasion (MTMO), multitrait-multirater, multitrait-multiinstrument, multitrait-multisituation, or multitrait-multidomain (MTMD) designs. Campbell and O'Connell (1967) operationalized this strategy by positing MTMO designs with time as the method effect, and Kenny (2022) noted that this strategy could apply to other design features. Thus, studies have explicitly treated the second facet as substantively meaningful—such as occasions (Campbell & O'Connell, 1967), raters (Hoyt &

Melby, 1999), and instruments (Widaman, 1985)—highlighting that what is labeled “method” may actually capture meaningful systematic variance.

Although our EC-F framework advocates evaluating both facets symmetrically, it does not preclude treating the method facet as subordinate when this aligns with the research aims (i.e., retaining the traditional M = method perspective). However, it is crucial to retain this symmetry whatever the source of the method facet. This flexibility allows researchers to apply the EC-F framework to their specific needs while maintaining rigorous construct-validation practices.

Illustrative Empirical Application Using Data From the Trends in International Mathematics and Science Study: Overview

To demonstrate the utility and generalizability of the EC-F framework, we illustrate its application using data from the 2007 Trends in International Mathematics and Science Study (TIMSS). This large-scale, multidimensional data set includes responses from students across four countries and provides comprehensive measurement across five science-subject domains: mathematics, physics, chemistry, earth science, and biology. We focus on five theoretically grounded motivational- and achievement-related constructs: task value, self-concept, interest, enjoyment, and achievement. Each construct is assessed across all five domains, yielding a full 5 (Traits) \times 5 (Domains) design and producing 25 distinct trait-domain constructs. Each construct is represented as a unique latent factor capturing a particular trait measured in a specific content area.

This design exemplifies an MTMD structure—a natural extension of the traditional MTMM framework. The MTMD design's second facet (domain) represents substantively meaningful content variation rather than an artifact of method effects. This perspective aligns with Campbell and Fiske's (1959) insight that the distinction between trait and method is often context-dependent and that both facets may be substantively meaningful in applied-research settings. By treating both facets—traits and domains—as equally substantive, the MTMD design naturally supports a symmetrical two-facet approach, consistent with our EC-F framework (see Fig. 1).

For example, five self-concept factors—one each for mathematics, physics, chemistry, earth science, and biology—reflect students' self-concept in each discipline. Likewise, there are five domain-specific factors for mathematics, each capturing a different motivational trait (e.g., interest in math, enjoyment of math, math self-concept). These 25 trait-domain constructs form the core of the MTMD matrix and serve as the foundation for all subsequent analyses.

As shown in Table 1, the MTMD correlation matrix retains the structural logic of a traditional MTMM matrix, with convergent validities (monotrait-heteromethod correlations) appearing in the diagonal blocks corresponding to each trait. Highlighting methodological rigor, we emphasize that constructs are ideally measured using multiple observed indicators so that the correlations in the matrix reflect relationships among latent factors rather than manifest variables. A traditional CFA based on this latent structure enables empirical testing of the a priori measurement model, corrects for measurement error, yields appropriate standard errors for each estimate, and provides a foundation for formal statistical tests of the Campbell-Fiske guidelines in a latent-variable framework. Accordingly, Table 1 presents the MTMD latent correlation matrix estimated from the CFA model.

Both the trait and domain facets in this design are conceptually grounded and central to the study's theoretical aims. Treating either facet as a method effect would misrepresent its substantive role. The motivational traits capture psychologically meaningful constructs, and the domain facet represents structured content variation across sciences. This factorial structure exemplifies the type of applied-research setting in which the original MTMM logic—and our symmetrical extension—are especially valuable.

Accordingly, we begin by specifying and validating the a priori measurement model and then apply asymptotic parameter comparisons to evaluate each of the five Campbell-Fiske guidelines in testable statistical form. We conducted all analyses using standard tools for latent-variable modeling to ensure transparency and accessibility for applied researchers.

Although the empirical example is drawn from educational psychology, the same principles apply broadly across psychological and social-science research. The EC-F framework offers a rigorous, generalizable framework for evaluating construct validity in any two-facet design in which both dimensions represent meaningful sources of theoretical variation.

Method

Data and measures

We analyzed data from 18,047 eighth-grade students drawn from nationally representative samples in four OECD countries (Australia, Hong Kong, Slovenia, and the United States). The data set included measures across five science-subject domains: mathematics, physics, chemistry, earth science, and biology. Achievement constructs were derived from plausible values provided by TIMSS item-response-theory scales, and all other

constructs were assessed using Likert-type self-report items. As detailed in Table S2 in the Supplemental Material, items were worded in parallel across traits and domains (e.g., “I enjoy learning [biology/chemistry/etc.]”) with three to five items per construct, except for achievement and aspirations, each measured by a single indicator per domain. Although these single-indicator factors limited full latent modeling for those constructs, the overall MTMD structure remained applicable across the full 5 × 5 design.

All data used in the present investigation are freely available from TIMSS (IEA TIMSS, n.d.). For statistical code and description of the variables, see the Supplemental Material and the IEA TIMSS (n.d.).

Analytic procedures

Cross-country measurement invariance. We addressed the complex sampling design of TIMSS by incorporating the HOUWGT weight and treating class as the primary clustering unit. Five imputed data sets were constructed using the five plausible values for achievement provided by TIMSS. This strategy enabled proper estimation of standard errors and model fit while also accommodating the minimal item-level missingness (average < 2%).

We evaluated the measurement model using CFA in Mplus (Version 8; Muthén & Muthén, 2017), specifying 25 latent factors representing the trait-domain constructs. Following our EC-F framework, we began with a CFA model excluding method effects (Model M1a) and progressively added a priori correlated uniquenesses (CUs) to account for negative item wording and parallel item stems.

Finally, we tested the Campbell-Fiske guidelines using asymptotic parameter comparisons in Mplus via the MODEL CONSTRAINT command, focusing on key contrasts among latent correlations. This approach provides precise statistical tests using standard errors derived from the model-implied asymptotic covariance matrix, allowing rigorous evaluation of convergent and discriminant validity within the MTMD structure. For details of the specific tests—including the contrasts corresponding to Campbell and Fiske's (1959) five guidelines—see Figure 1 and Section 6 in the Supplemental Material.

Goodness of fit and model evaluation. We evaluate model fit using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR), interpreted according to conventional thresholds (e.g., CFI > .95, RMSEA < .05; Hu & Bentler, 1999; Marsh et al., 2004). Although the full

technical details of model specifications, imputation strategies, and incremental model-fit steps are provided in the Supplemental Material, they are not the focus of the present article. We include them for transparency and reproducibility, but our primary aim here is to demonstrate the conceptual and practical value of the EC-F framework.

As shown in Table S1 in the Supplemental Material, the model fit improved substantially with both sets of CUs included (Model M1d: RMSEA = .027, CFI = .957, Tucker-Lewis index [TLI] = .944). These effects were consistent across countries and retained in the final model.

Cross-country measurement invariance

We tested measurement invariance (Meredith, 1993; Millsap, 2010) across countries (configural, weak, strong, and strict) with a series of nested CFA models (M2a–M2d) reported in the Supplemental Material (see Section M1.7 and Table S1 in the Supplemental Material). Model M2d imposed equality constraints on factor loadings, intercepts, variances, and covariances across countries, thereby testing strict invariance. It demonstrated excellent fit to the data (CFI = .953, TLI = .941, RMSEA = .042), with minimal changes in fit indices across models (Δ CFI, Δ TLI < .003; RMSEA stable at .042).

Given this strong support for full measurement invariance, we selected Model M2d as the basis for all subsequent analyses. The establishment of full cross-national invariance (variances, covariances, factor loadings, and intercepts) justified the pooling of data across countries rather than conducting separate analyses in each national sample.

Standardized factor loadings for the 25 constructs—averaged across the five imputed data sets and constrained to equality across countries—ranged from .45 to .89; most constructs loaded above .80 on at least one indicator. These results support the robustness of the latent structure (see Table S2 in the Supplemental Material).

The findings confirm that a well-defined measurement model that is invariant over countries. The latent-correlation matrix from this measurement model is the basis of all subsequent analyses demonstrating the application of the EC-F framework.

Results

Having established the measurement model and specified the EC-F-framework statistical tests, we now turn to the empirical results. Using asymptotic parameter-comparison techniques, we tested differences between individual latent correlations and averages of correlation subsets in the MTMD matrix. These comparisons address, for example, whether self-concept correlations across

domains are significantly greater than zero and whether they exceed correlations among different traits or domains. A central strength of this approach is its flexibility: It allows for rigorous statistical tests at both micro (individual correlations) and macro (pattern-level summary) levels—across 300 latent correlations derived from a unified model. Tables 1 to 5 summarize the key results across traits and domains.

Advanced organizer: worked example—self-concept and mathematics

As an advanced organizer, Table 2 (see also Section 2 in the Supplemental Material) illustrates the logic of asymptotic parameter comparisons for one trait (self-concept) and one domain (mathematics), using values drawn from the latent MTMD matrix (Table 1; see also Table 6). Because this example focuses on a single trait and domain, only Guidelines 1 through 3 are relevant (Campbell & Fiske, 1959). Guidelines 4 and 5 (Campbell & Fiske, 1959) apply to broader patterns across multiple traits and methods and are therefore evaluated in subsequent analyses (for details on the Campbell-Fiske guidelines, see Fig. 1).

Trait Facet: self-concept (based on Table 1)

Convergent validity (Guideline 1). The correlation between math self-concept and physics self-concept (monotrait-heterodomain; MTHD) is strong and statistically significant, $r = .51$ ($SE = .011$).

Discriminant validity (Guideline 2). This MTHD correlation is significantly larger than the heterotrait-heterodomain (HTHD) correlation between math self-concept and physics extrinsic motivation, $r = .20$; difference = .31 ($SE = .012$).

Discriminant validity (Guideline 3). The convergent correlation also exceeds the heterotrait-monodomain (HTMD) correlation between math self-concept and math extrinsic motivation by .16 ($SE = .015$).

Domain facet: mathematics (based on Table 5)

Convergent validity (Guideline 1). The correlation between math self-concept and math extrinsic motivation (HTMD) is $r = .35$ ($SE = .011$).

Discriminant validity (Guideline 2). This correlation exceeds the same HTHD correlation noted above by .15 ($SE = .013$).

Discriminant validity (Guideline 3). However, the domain-based convergent validity is significantly lower than the cross-domain self-concept correlation ($r = .51$), yielding a negative difference of $-.16$ ($SE = .015$).

Table 2. Applying the Extended Campbell-Fiske Guidelines Based on Single Correlations From the Latent MTMD Matrix (Table 1): Illustrative Asymptotic Parameter Comparisons for the Self-Concept Trait and the Math Domain

Facet comparison		Estimate	SE	Description
Trait facet: SC	Convergent validity (Math SC×Physics SC), MTHD	.51	.011	Strong convergent validity (Guideline 1)
	Discriminant validity: convergent vs. HTHD (Math SC×Physics Extrinsic Motivation)	.31	.012	Convergent validity exceeds HTHD correlations (Guideline 2)
	Discriminant validity: convergent vs. HTMD (Math SC×Math Extrinsic Motivation)	.16	.015	Convergent validity exceeds HTMD correlations (Guideline 3)
Domain facet: math	Convergent validity (Math SC×Math Intrinsic Motivation), HTMD	.35	.011	Strong convergent validity for domain facet (Guideline 1)
	Discriminant validity: convergent vs. HTHD (Math SC×Physics Extrinsic Motivation)	.15	.013	Convergent validity exceeds HTHD correlations (Guideline 2)
	Discriminant validity: convergent vs. MTHD (Math SC×Physics SC)	−.16	.015	Convergent validity is lower than MTHD (violation of Guideline 3)

Note: This table provides a didactic illustration of how asymptotic parameter comparisons are used to test the Campbell-Fiske guidelines (Campbell & Fiske, 1959; see Fig. 1) using selected examples from the latent MTMD matrix (Tables 1 and 6). We do not intend this summary to be exhaustive but rather to demonstrate the logic underlying the broader analyses (also see Section 2 in the Supplemental Material available online). For each facet (trait: self-concept; domain: math), we illustrate the following: convergent-validity tests (Campbell & Fiske, 1959, Guideline 1), showing the magnitude of the focal correlations (e.g., MTHD when the focus is on traits or HTMD when the focus is on domain), and discriminant-validity tests comparing convergent correlations with HTHD and HTMD correlations (Campbell & Fiske, 1959, Guidelines 2 and 3 for the trait focus). Critically, the MTHD correlation (.51) serves as the convergent-validity benchmark for the trait facet but becomes the discriminant-validity comparison for the domain facet. Conversely, the HTMD correlation (.35) serves as convergent validity for the domain facet but as the discriminant comparison for the trait facet. Because of this interplay, the test estimates for Guideline 3 (discriminant validity; Campbell & Fiske, 1959) for the trait and domain facets must sum to zero (+.16 and −.16). Likewise, the HTHD correlation (.20) is the same for both trait and domain facets when testing discriminant validity (Campbell & Fiske, 1959, Guideline 2). This structural juxtaposition is inherent in all MTMD matrices and generalizes across all comparisons. Critical is the symmetry between trait and domain facets (e.g., that the same correlation might serve as a benchmark for one facet and as a comparison for another). For completeness, detailed results for all trait and domain comparisons are provided in the Supplemental Material. MTMD = multitrait-multidomain; MTHD = monotrait-heterodomain; HTMD = heterotrait-multidoman; HTHD = heterotrait-heterodomain.

This worked example highlights a key structural implication of the MTMD design: Guideline 3 discriminant effects for trait and domain facets (Campbell & Fiske, 1959; for details, see Fig. 1) are necessarily reciprocal. Discriminant validity of the trait facet comes at the cost of a lack of discriminant validity of the domain facet and vice versa. Here, the positive difference for the trait facet (+.16) and the negative difference for the domain facet (−.16) must sum to zero. Likewise, the same HTHD correlation ($r = .20$) anchors both facets when testing Guideline 2 (Campbell & Fiske, 1959; for details, see Fig. 1). These interdependencies are built into all MTMD matrices and form a core rationale for using a symmetric framework.

Unified matrix, dual perspectives: juxtaposing trait and domain facets

We present two arrangements of the same latent MTMD matrix: one emphasizing the trait facet (e.g., self-concept

across domains, Table 1) and the other emphasizing the domain facet (e.g., motivation constructs within a domain, Table 5). These are not separate analyses but alternative views of the same underlying latent structure. We based all results on this one unified CFA model.

The distinction between trait and domain perspectives is conceptual rather than structural. The same latent correlations serve different roles depending on orientation: For instance, MTHD coefficients represent convergent validity in trait-based comparisons but function as discriminant comparisons when viewed from the domain perspective. This symmetry is essential for interpreting the pattern of results and understanding the trade-offs inherent in Guideline 3 (Campbell & Fiske, 1959; for details, see Fig. 1) as illustrated above.

To aid clarity, we first present detailed results from the trait perspective (focusing on Guidelines 1–3; Campbell & Fiske, 1959; for details of guidelines, see Fig. 1), followed by a briefer summary of domain-oriented results.

Table 3. Summary of Tests of Convergent and Discriminant Validity for the Trait Facet Based on the Latent MTMD Matrix

Domain pair	CV (<i>SE</i>)	HTHD (<i>SE</i>)	G2 (<i>SE</i>)	G3A (<i>SE</i>) ^a	G3B (<i>SE</i>) ^a	Difference (<i>SE</i>)
Math-physics	.50 (.005)	.22 (.006)	.28 (.005)	.10 (.007)	.10 (.007)	.00 (.007)
Math-chemistry	.44 (.005)	.19 (.006)	.25 (.005)	.03 (.008)	.03 (.008)	-.00 (.007)
Math-earth sciences	.36 (.006)	.13 (.006)	.24 (.005)	-.04 (.008)	-.00 (.008)	.03 (.007)
Math-biology	.30 (.007)	.09 (.007)	.22 (.005)	-.01 (.009)	-.08 (.008)	.02 (.007)
Physics-chemistry	.52 (.009)	.20 (.007)	.32 (.007)	.12 (.010)	.12 (.010)	-.01 (.007)
Physics-earth sciences	.43 (.007)	.14 (.006)	.29 (.005)	.03 (.010)	.06 (.008)	.03 (.007)
Physics-biology	.41 (.007)	.13 (.006)	.28 (.005)	.01 (.009)	.02 (.010)	.01 (.006)
Chemistry-earth sciences	.42 (.007)	.12 (.006)	.01 (.003)	.00 (.010)	.04 (.010)	.04 (.007)
Chemistry-biology	.50 (.007)	.20 (.007)	.29 (.005)	.10 (.008)	.11 (.008)	.02 (.007)
Earth sciences-biology	.45 (.007)	.13 (.006)	.32 (.005)	.08 (.009)	.06 (.009)	-.02 (.006)
Summary						
Math	.40 (.004)	.16 (.004)	.25 (.004)	.00 (.007)		
Physics	.46 (.005)	.17 (.005)	.29 (.004)	.06 (.007)		
Chemistry	.47 (.005)	.18 (.005)	.29 (.004)	.07 (.007)		
Earth sciences	.41 (.005)	.13 (.006)	.28 (.004)	.04 (.007)		
Biology	.42 (.005)	.14 (.004)	.28 (.003)	.03 (.004)		
Grand total	.43 (.004)	.15 (.004)	.28 (.003)	.04 (.004)		

Note: Each cell represents the relation between a pair of domains (e.g., math-physics) averaged across the five traits (self-concept, intrinsic motivation, extrinsic motivation, achievement, aspiration) based on the MTMD latent matrix in Table 1. The reported difference is between HTMD and HTHM correlations, computed separately for each pair of domains (Campbell & Fiske, Guideline 5). Summary is the aggregate associated with each domain. CV = convergent validity (Campbell & Fiske, Guideline 1); HTHD = heterotrait-heterodomain correlation; HTMD = heterotrait-monodomain correlation; G2 and G3 = Campbell and Fiske (1959) Guidelines 2 and 3 for discriminant validity. For details on the Campbell-Fiske guidelines, see Figure 1.

^aFor individual trait pairs, tests for G3 (Campbell & Fiske, 1959) were conducted separately for each domain (e.g., math and physics). For the total summary scores, tests were aggregated across all domains.

We present a more detailed discussion of their implications in the Discussion section.

Trait-facet results. For the trait facet, we assessed convergent validity through MTHD correlations—associations among the same construct across different content domains (e.g., self-concept in math, physics, chemistry, earth science, and biology). Discriminant validity is supported when these MTHD correlations are larger than two key alternatives: (a) HTHD correlations involving different traits across domains (Guideline 2; Campbell & Fiske, 1959; see Fig. 1) and (b) HTMD correlations involving different traits within the same domain (Guideline 3; Campbell & Fiske, 1959; see Fig. 1). Table 1 presents the MTMD latent correlation matrix organized by trait, and Table 3 summarizes key asymptotic comparisons.

Convergent validity (Guideline 1). MTHD correlations were consistently strong and statistically significant. The grand mean convergent validity (CV_{tot}) was .43 ($SE = .004$), indicating substantial consistency for each trait across academic domains.

Discriminant validity (Guideline 2). MTHD correlations significantly exceeded the corresponding HTHD values. The overall mean difference ($G2_{\text{tot}}$) was .28 ($SE = .003$), reflecting robust discriminant validity across traits and domains.

Discriminant validity (Guideline 3). MTHD values also exceeded HTMD correlations, although differences were modest. The grand mean contrast ($G3_{\text{tot}}$) was .04 ($SE = .005$); domain-specific variation (e.g., $G3_{\text{math}} = .00$) indicated weaker evidence in some subject areas.

Although not the primary focus here, results for Guidelines 4 and 5 (which require considering patterns across multiple traits and methods; Campbell & Fiske, 1959; see Fig. 1) are summarized in Table 4. Guideline 4 showed high profile similarity across domains. Following Marsh (1988) and Millsap and Everson (1993), we operationalize Guideline 4 with the profile similarity index (PSI) as a descriptive indicator of structural consistency. PSI is computed as the Pearson correlation between vectors of intertrait correlations obtained separately within each method (for description of PSI, see

Table 4. Pattern-Similarity Indices Tests of Campbell-Fiske Guideline 4 for Domain Facet and Construct Facet

	Trait facet (Table 1)					Domain facet (Table 5)				
	Math	Physics	Chemistry	Earth sciences	Biology	Self-concept	Intrinsic	Extrinsic	Achievement	Aspirations
Math	1					1				
Physics	.86	1				.95	1			
Chemistry	.83	.99	1			.55	.51	1		
Earth sciences	.84	.99	.99	1		.08	-.05	.72	1	
Biology	.84	.99	.99	.99	1	.86	.91	.76	.22	1

Note: For each domain (math, physics, etc.) in the trait-facet PSI matrix, we computed the 10 intercorrelations among the five constructs (self-concept, intrinsic, etc.) and then correlated these domain-specific correlation profiles. The resulting PSI values indicate how similarly the constructs interrelate across domains. The pattern of correlations among traits is very similar for the five subject domains (PSI r s > .8), supporting the generalizability of the structure. For each construct in the domain-facet PSI matrix, we computed the 10 intercorrelations among the five domains and then correlated these construct-specific domain profiles. The resulting PSI values index the consistency of domain-specific patterns across constructs. For details on the Campbell and Fiske (1959) guidelines, see Figure 1. PSI = pattern-similarity index.

Table 1; for more detail, see Section S3 in the Supplemental Material). PSI values ranged from .86 to .99 (Table 4), suggesting stable trait interrelations across academic contexts.

Practical implications. Strong support for Guidelines 1 and 2 (Campbell & Fiske, 1959; for details, see Fig. 1) affirms the generalizability of motivational constructs such as self-concept and task value across academic subjects. More modest effects for Guideline 3 (Campbell & Fiske, 1959; for details, see Fig. 1) point to conceptual overlap among traits within specific domains—a plausible pattern in real-world educational settings.

Domain-facet results for trait facet: self-concept (based on Table 1). Tables 5 and 6 reorganize the same MTMD matrix to emphasize the domain facet—examining how motivation constructs cluster within specific academic subjects. In this perspective, the logic of comparison is reversed: HTMD correlations reflect convergent validity, and MTHD and HTHD correlations serve as discriminant benchmarks.

Convergent validity (Guideline 1). HTMD correlations were substantial across all five domains. The grand mean convergent validity (CV_{tot}) was .39 ($SE = .004$), supporting the coherence of motivational constructs within subject areas.

Discriminant validity (Guideline 2). HTMD correlations significantly exceeded HTHD comparisons, with an overall $G2_{tot} = .24$ ($SE = .003$), consistent with expectations of domain-specific validity.

Discriminant validity (Guideline 3). HTMD values were slightly lower than MTHD correlations, resulting in a negative grand mean difference ($G3_{tot} = -.04$, $SE = .005$). This negative value mirrors the positive $G3_{tot}$ in the trait-facet analysis, reflecting the structural reciprocity inherent in the MTMD design.

Results for Guidelines 4 and 5 (Campbell & Fiske, 1959; for details, see Fig. 1) revealed important insights. Guideline 4 (Table 4) revealed consistent intertrait patterns across domains, particularly for constructs such as self-concept, intrinsic motivation, and aspirations. However, PSIs were lower for extrinsic motivation and especially achievement, suggesting weaker cross-domain consistency for these constructs. Guideline 5 indicated substantial mean differences among traits as a function of domain. The mean correlation among achievement traits (Table 6) was substantially higher than correlations among traits for any of the other constructs—highlighting the weak discriminant validity of the achievement measures. For a more detailed analysis of this issue, including additional tests of domain specificity, see Section 3 in the Supplemental Material.

Practical implications. The domain perspective highlights the extent to which constructs cohere within academic subjects while also revealing cross-domain generalizability for certain traits. For instance, self-concept may correlate more strongly across physics and chemistry than with unrelated constructs within these domains. This underscores the importance of analyzing both domain-specific and trait-general patterns to understand the full structure of motivational constructs in educational settings. The very high correlations among achievement scores representing the different domains are of particular relevance.

Trait-domain juxtaposition and structural symmetry

Critically, we derived both the trait- and domain-organized MTMD matrices (Tables 1 and 5) from the same latent model. Their structural equivalence is especially relevant for Guideline 3, in which convergent validity in one perspective serves as the discriminant benchmark in the other (Campbell & Fiske, 1959). As a result, strong support for convergent validity in one facet necessarily constrains support for discriminant validity in the opposing view—an implication not always explicitly addressed in previous MTMM research.

Rather than viewing trait and domain perspectives as competing, we interpret them as complementary. The trait orientation emphasizes construct coherence across contexts, and the domain orientation captures conceptual clustering within specific subject areas. A comprehensive understanding of convergent and discriminant validity in multidimensional constructs requires attention to both perspectives—and their inherent interplay.

Demonstrating the flexibility of the extended MTMD framework: illustrative applications

In this section, we demonstrate the flexibility of the extended MTMD framework through a series of illustrative applications. These examples showcase the framework's capacity to address a range of substantive research questions, highlighting its versatility across different constructs, contexts, and levels of analysis.

Adapting a classic analysis-of-variance framework using the MTMD matrix

Historically, MTMM data were analyzed using a three-factor unrepeated random-effects analysis-of-variance (ANOVA) model, partitioning variance into orthogonal sources attributed to persons, traits, and methods (Kavanagh et al., 1971; Stanley, 1961; see also Marsh & Hocevar, 1983; Millsap, 2014; Schmitt & Stults, 1986). Conceptually aligned with

Table 5. Latent Multitrait-Multidomain Matrix For the Domain Facet: Relations Among Five Domains and Five Traits

	SC					INT					EXT					ACH					ASP																			
	M	P	C	E	B	M	P	C	E	B	M	P	C	E	B	M	P	C	E	B	M	P	C	E	B															
SC																																								
M	1																																							
P	.51	1																																						
C	.42	.42	1																																					
E	.27	.34	.27	1																																				
B	.13	.31	.41	.35	1																																			
INT																																								
M	.69	.36	.29	.16	.07	1																																		
P	.31	.80	.27	.20	.18	.45	1																																	
C	.24	.28	.80	.15	.30	.37	.38	1																																
E	.12	.20	.14	.79	.20	.26	.28	.23	1																															
B	-.03	.13	.25	.15	.76	.17	.25	.40	.27	1																														
EXT																																								
M	.35	.26	.22	.17	.14	.43	.28	.23	.20	.16	1																													
P	.20	.41	.23	.14	.14	.26	.49	.27	.18	.18	.50	1																												
C	.11	.19	.40	.11	.22	.20	.25	.49	.16	.29	.40	.65	1																											
E	.05	.13	.13	.33	.13	.12	.17	.19	.45	.19	.38	.50	.53	1																										
B	.04	.13	.21	.10	.38	.14	.18	.29	.18	.50	.34	.46	.63	.56	1																									
ACH																																								
M	.58	.32	.26	.21	.16	.30	.16	.12	.07	-.02	.09	.06	-.02	-.05	-.04	1																								
P	.42	.31	.24	.20	.21	.18	.17	.12	.07	.03	.05	.08	-.01	-.05	-.03	.70	1																							
C	.44	.31	.25	.20	.25	.21	.16	.14	.08	.08	.07	.07	.00	-.03	.01	.71	.78	1																						
E	.39	.30	.23	.23	.25	.16	.15	.10	.09	.07	.04	.07	-.02	-.03	-.01	.68	.77	.78	1																					
B	.40	.28	.25	.22	.29	.16	.12	.12	.07	.11	.04	.03	-.01	-.05	.00	.71	.80	.81	.82	1																				
ASP																																								
M	.41	.24	.20	.10	.05	.68	.33	.27	.21	.14	.34	.22	.18	.12	.13	.16	.08	.11	.07	.08	1																			
P	.18	.57	.20	.12	.11	.31	.75	.30	.21	.18	.22	.44	.25	.18	.18	.05	.07	.06	.05	.03	.33	1																		
C	.15	.22	.62	.09	.22	.25	.30	.78	.16	.31	.19	.27	.44	.19	.27	.04	.06	.06	.04	.05	.28	.38	1																	
E	.07	.14	.09	.60	.13	.17	.21	.16	.76	.19	.16	.18	.15	.40	.17	.02	.03	.03	.05	.03	.21	.26	.21	1																
B	-.06	.09	.19	.11	.57	.09	.18	.31	.20	.75	.12	.17	.27	.18	.45	-.05	.00	.04	.03	.06	.17	.22	.35	.25	1															

Note: In each triangular submatrix of the latent multitrait-multidomain matrix are correlations between the same trait representing different domains (monotrait-heterodomain). In square submatrix are correlations among traits representing different domains (heterotrait-heterodomain) with convergent validities (heterotrait-monodomain; shaded in gray) in the diagonals. The latent multitrait-multidomain matrix is based on Model M2d (see Table S1 in the Supplemental Material available online); see factor loadings in Table 3. Each correlation is averaged over five imputed data sets and constrained to be equal across all four countries (see Model M2d, Table S1 in the Supplemental Material). SC = self-concept; INT = intrinsic motivation; EXT = extrinsic motivation; ACH = achievement; ASP = aspiration; M = math; P = physics; C = chemistry; E = earth sciences; B = biology.

Table 6. Summary of Tests of Convergent and Discriminant Validity for the Domain Facet (Based on the Latent MTMD matrix)

Trait pair	CV (SE)	HTHD (SE)	G2 (SE)	G3A (SE) ^a	G3B (SE) ^a	Difference (SE)
Self-concept–intrinsic	.77 (.004)	.20 (.006)	.57 (.006)	.43 (.008)	.46 (.008)	.04 (.007)
Self-concept–extrinsic	.38 (.007)	.15 (.007)	.22 (.005)	.03 (.009)	-.08 (.008)	-.12 (.008)
Self-concept–achievement	.33 (.008)	.28 (.007)	.06 (.004)	-.01 (.008)	-.42 (.010)	-.41 (.008)
Self-concept–aspirations	.55 (.005)	.13 (.005)	.42 (.005)	.21 (.009)	.20 (.008)	.08 (.007)
Intrinsic–extrinsic	.47 (.006)	.21 (.006)	.27 (.005)	.20 (.009)	.03 (.007)	-.15 (.007)
Intrinsic–achievement	.16 (.007)	.11 (.006)	.05 (.004)	-.11 (.009)	-.59 (.009)	-.45 (.008)
Intrinsic–aspirations	.74 (.003)	.22 (.005)	.52 (.006)	.47 (.007)	.48 (.007)	.04 (.005)
Extrinsic–achievement	.02 (.009)	.01 (.009)	.01 (.003)	-.43 (.010)	-.74 (.012)	-.30 (.008)
Extrinsic–aspirations	.41 (.005)	.19 (.005)	.22 (.005)	-.03 (.007)	.15 (.007)	.19 (.008)
Aspirations–achievement	.08 (.007)	.04 (.007)	.04 (.003)	-.67 (.009)	-.18 (.009)	-.49 (.008)
Summary self-concept	.51 (.004)	.19 (.005)	.32 (.004)	.17 (.007)		
Summary intrinsic	.54 (.003)	.19 (.004)	.35 (.004)	.29 (.007)		
Summary extrinsic	.36 (.004)	.14 (.005)	.22 (.004)	-.10 (.006)		
Summary achievement	.25 (.005)	.11 (.006)	.14 (.004)	-.51 (.008)		
Summary aspirations	.45 (.003)	.15 (.004)	.30 (.004)	.18 (.007)		
Grand total	.39 (.004)	.15 (.004)	.24 (.003)	-.03 (.005)		

Note: Each cell represents the relation between a pair of traits (e.g., self-concept–intrinsic) averaged across the five domains (math, physics, chemistry, earth sciences, biology) based on the MTMD latent matrix in Table 5. The difference is between the HTMD and HTHM correlations. These were tested separately for each pair of traits. G2 tests CV against HTHD, and G3 tests CV against HTMD. CV = convergent validity (Campbell & Fiske, 1959, Guideline 1); HTHD = heterotrait-heterodomain correlation; HTMD = heterotrait-monodomain correlation; G2 and G3 = Campbell-Fiske Guidelines 2 and 3 for discriminant validity. For details on the Campbell-Fiske guidelines, see Figure 1.

^aFor individual trait pairs, G3 tests were conducted separately for each domain (e.g., math and physics). For the total summary scores, tests were aggregated across all domains.

generalizability theory (Woehr et al., 2012), this framework provided a global assessment of variance components. However, like the original Campbell-Fiske guidelines, these ANOVA-based models relied on manifest correlations and did not correct for measurement error. Moreover, they summarized variance patterns globally, without explicitly modeling trait-by-method unit interactions central to MTMM logic.

However, when applied to a latent MTMD matrix, the classic ANOVA decomposition becomes a powerful diagnostic tool for probing assumptions embedded in MTMM-SEM frameworks—most notably, the widely used CT-CM model. Critically, all standard CFA-based MTMM models (e.g., CT-CM, CT-UM, CT-C[M–1], and CT-CU) impose additive structures that effectively preclude the detection of trait-by-method (or in our case, trait-by-domain) interactions. That is, they assume uniform validity patterns across combinations, potentially masking meaningful sources of variation.

One of the formative insights of the original Campbell-Fiske guidelines—particularly in Guidelines 4 (profile similarity across methods) and 5 (domain specificity of trait relationships)—is that such interaction effects may, in fact, be present (Campbell & Fiske, 1959). Our adaptation of the ANOVA model within a fully latent MTMD framework enables an explicit test of this possibility by including a trait-by-domain interaction term.

We applied a two-factor (Trait \times Domain) ANOVA decomposition to the latent MTMD matrix using Mplus's MODEL CONSTRAINT functionality. We computed the mean latent correlation for each of the 25 MTMD cells and statistically tested for main effects of trait and domain and their interaction. The residual three-way interaction (Person \times Trait \times Domain) captures nonadditive variance not accounted for by the additive main effects—analogue to the residual error in traditional unreplicated ANOVA models.

Results revealed statistically significant main effects for both traits and domains, consistent with theoretical expectations: Motivational traits showed greater generalizability across domains, whereas achievement constructs were more domain-specific (Table 7). Critically, the trait-by-domain interaction was also significant. This finding indicates that trait-domain associations vary systematically and nonuniformly—challenging the assumption of additivity that underpins most MTMM-SEM approaches.

From a practical standpoint, this interaction highlights a key limitation of standard CFA-based models: They cannot accommodate or detect local variations in construct-validity patterns across MTMD matrices. In contrast, we demonstrated that the latent-variable ANOVA approach can directly test and interpret these patterns meaningfully.

Table 7. ANOVA Results for the MTMD Matrix

Source	df	Computation of variance components	Variance-component estimates
Person (σ_p^2)	$N = 18,046$	$M(\text{HDHT})$.15 (.004)
Domain \times Person (σ_{pd}^2)	$N \times \text{Domain} = 72,184$	$M(\text{MDHT}) - M(\text{HDHT})$.24 (.003)
Trait \times Person (σ_{pt}^2)	$N \times \text{Trait} = 72,184$	$M(\text{HDMT}) - M(\text{HMHT})$.28 (.003)
Domain \times Trait \times Person + residual ($\sigma_{pdt,r}^2$)	$N \times \text{Trait} \times \text{Domain} = 288,736$	$1 - M(\text{MDHT}) - M(\text{HDMT}) + M(\text{HDHT})$.33 (.003)

Note: Number of domains is 5, and number of traits is 5. $N = 18,046$ (number of individuals). Results are based on the latent MTMD matrix (Table 1). The same equations and computations generalize to the traditional MTMM matrix, with traits and methods instead of domains. Because the values are based on a latent correlation matrix, there is no residual variance associated with measurement error. All values in this table are based on asymptotic parameter comparisons applied to the latent MTMD matrix (Tables 1 and 6), providing appropriate standard errors for each estimate. A unique feature of this asymptotic framework is the ability to formally test the substantively important Domain \times Trait (Domain \times Construct \times Person) interaction—capturing nonadditive effects across facets—something that is not easily accommodated by other multitrait-multimethod approaches. ANOVA = analysis of variance; MTMD = multitrait-multidomain; HDHT = heterodomain-heterotrait correlations; MDHT = monodomain-heterotrait correlations; HTMD = heterotrait-monodomain correlations; HDMT = heterodomain-monotrait; HMHT = heteromethod-heterotrait.

In summary, this illustrative application highlights the diagnostic value of the MTMD framework when combined with formal asymptotic comparisons and latent-variable modeling. The results reveal nuanced validity structures—particularly trait-by-domain interactions—that would remain undetected in traditional confirmatory MTMM models. This reinforces the broader utility and flexibility of the EC-F framework within modern construct-validation research.

Substantive application: motivation-achievement relations across domains

In the preceding sections, we applied the EC-F framework to evaluate construct validity within the MTMD matrix. In this section, we illustrate the broader diagnostic potential of the framework by examining a substantively important application: the domain-specific relationships between motivational traits and academic achievement. Crucially, the same unified MTMD structure that supports formal tests of convergent and discriminant validity also enables targeted substantive analyses—demonstrating the flexibility of our EC-F framework. Here, we use the latent MTMD matrix to explore how the strength and specificity of motivation-achievement associations vary across academic subjects and motivational constructs.

Research on academic self-concept has long emphasized domain specificity in its relationship with achievement. For instance, math self-concept typically correlates strongly with math achievement but weakly with verbal achievement—even when achievement measures are highly intercorrelated (Marsh, 2007). High domain-matched correlations support convergent validity, whereas low cross-domain correlations support discriminant validity. Here, we extend this analysis to include

additional motivational traits, assessing their pattern of relations with achievement across domains.

Although these relations are not directly part of the Campbell-Fiske guidelines, they can be meaningfully examined using the MTMD matrix—primarily through the lens of trait-by-domain interactions. Table 8 summarizes results based on 100 correlations involving at least one achievement variable drawn from the full 300-cell MTMD matrix (see Tables 1 and 5).

Motivational-trait facet. For motivational traits, we defined convergent validity as the correlation between a given trait and achievement in the corresponding domain (e.g., math self-concept and math achievement). Nonconvergent validity compared that trait’s correlation with achievement in nonmatching domains (e.g., math self-concept and biology achievement). Self-concept exhibited the most evident domain-specific pattern, characterized by the highest mean convergent validity (.34) and the highest nonconvergent correlations (.27), resulting in a modest but meaningful discriminant difference (.07). Intrinsic motivation also demonstrated moderate convergent validity (.16) and discriminant validity (.05). In contrast, extrinsic motivation and aspirations showed weak domain-specific correlation patterns (.03–.07 for convergent, .02 for discriminant validity), suggesting they may be more general or less directly predictive of domain-specific performance.

Domain facet. We defined convergent validity as the correlation between a motivation construct and achievement within the same domain (e.g., biology self-concept with biology achievement). Nonconvergent correlations involved mismatched domains (e.g., physics self-concept with biology achievement). Results revealed clear domain differences. The mean convergent validity across traits was highest for mathematics (.28), substantially higher than for other

Table 8. Correlations Between Motivation Constructs and Achievement: Convergent and Discriminant Validity in Relation to Subject Domain and Motivation construct

	Convergent		Nonconvergent		Difference	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Subject domain						
Biology	.12	.011	.11	.007	.00	.010
Chemistry	.13	.012	.13	.009	.00	.007
Earth science	.09	.009	.12	.009	-.04	.011
Math	.28	.008	.08	.007	.20	.008
Physics	.16	.011	.10	.006	.06	.010
Total mean	.15	.006	.11	.006	.04	.003
Trait construct						
Extrinsic	.03	.009	.01	.006	.02	.004
Self-concept	.34	.008	.27	.007	.07	.005
Aspirations	.07	.006	.05	.007	.02	.004
Intrinsic	.16	.007	.11	.006	.05	.004
Total mean	.15	.006	.11	.006	.04	.002

Note: All values are based on correlations in Table 1 (or equivalently, Table 5). For the subject-domain facet, convergent estimates are the mean estimates for each domain when the subject domains for the motivation construct and for achievement match (e.g., biology achievement and biology motivations). Nonconvergent estimates are mean estimates for each domain when the subject domain does not match (e.g., biology achievement and motivations on all domains other than biology). The difference between the two is a test of discriminant validity in relation to the subject domain. Likewise, for each trait construct, convergent estimates are the mean estimates for each trait when the subject domain for the motivation construct and achievement domain match (e.g., biology self-concept with biology achievement, physics self-concept with physics achievement). Other estimates are mean estimates for each trait when the subject domain and achievement domain do not match (e.g., biology self-concept with physics achievement, physics self-concept with biology achievement). The difference between the two is a test of discriminant validity concerning the trait domain.

domains (.12–.16) and lower nonconvergent correlations (.08). This produced a strong discriminant-validity effect (.20) for math that we did not observe in other domains.

Summary. Across both facets, the most substantial evidence for convergent and discriminant validity emerged for the self-concept trait and the mathematics domain. The strongest individual convergent correlation was between math self-concept and math achievement ($r = .58$), followed by math intrinsic motivation and math achievement ($r = .31$). Math-related motivation constructs consistently showed stronger correlations with achievement than did constructs in other domains. Self-concept exhibited more robust and generalizable achievement relations than other motivational traits.

These findings underscore the diagnostic utility of the EC-F framework in exploring domain-specific relations between motivation and achievement. Treating the domain facet merely as a “method” would obscure the uniquely strong predictive value of math-based motivation constructs. Likewise, treating motivational traits as

interchangeable would overlook the particular relevance of academic self-concept. Our EC-F framework enables a more nuanced understanding of construct validity—essential for theoretical refinement and applied interventions in educational research. Although the focus here is necessarily on motivation-achievement relations, the underlying logic—testing domain-specific relations within an MTMD matrix—may also be relevant in other areas of psychology in which domain-specific processes are important (e.g., emotions, quality of life, self-efficacy, job satisfaction, well-being, basic psychological needs, self-compassion). We return to this point in the general discussion section.

Substantive application: domain-specific gender differences in motivation and achievement

In our substantive-methodological synergy, our EC-F framework also readily examines theoretically and policy-relevant correlates, such as gender—an especially

Table 9. Gender Differences: Breakdown as a Function of Domain and Construct

Trait	Domain					<i>M</i>	<i>SS</i>
	Math	Physics	Chemistry	Earth sciences	Biology		
Self-concept	.114 (.010)	.187 (.011)	.065 (.011)	.115 (.010)	−.075 (.011)	.081 (.007)	.038 (.004)
Intrinsic	.017 (.010)	.064 (.010)	.001 (.011)	.049 (.010)	−.148 (.011)	.010 (.007)	.051 (.004)
Extrinsic	.060 (.011)	.136 (.009)	−.046 (.010)	−.002 (.010)	−.138 (.010)	.003 (.007)	.043 (.004)
Achievement	.006 (.010)	.097 (.012)	.010 (.010)	.072 (.011)	−.030 (.010)	.029 (.009)	.012 (.002)
Aspirations	.019 (.010)	.167 (.010)	.043 (.010)	.082 (.010)	−.092 (.010)	.036 (.007)	.039 (.003)
<i>M</i>	.026 (.007)	.150 (.007)	.014 (.007)	.064 (.007)	−.097 (.007)	.032 (.005)	
<i>SS</i>	.014 (.002)	.005 (.002)	.007 (.001)	.007 (.001)	.009 (.002)		.202 (.013)
ANOVA							
Trait							.019 (.003)
Domain							.159 (.011)
Trait × Domain							.023 (.002)

Note: The set of 25 gender differences consists of correlations between gender and each of the 25 (5 Domain × 5 Construct) latent factors. Values in parentheses are asymptotic standard errors. All values in this table are based on asymptotic parameter comparisons applied to the latent multitrait-multidomain matrix (Tables 1 and 6), providing appropriate standard errors for each estimate. We present the sums of squared deviations (*SS*) and tests of statistical significance for each row and column (e.g., variation in self-concept ratings across different domains). The “*SS*” column summarizes the sum of squares for gender differences across domains and traits. The lower ANOVA section decomposes variance into trait, domain, and trait-by-domain (Trait × Domain) interaction components. The ANOVA results report the *SS* and tests of statistical significance for gender differences associated with the main effects of domains, constructs, and their interaction. A unique feature of this asymptotic framework is the ability to formally test the substantively important Domain × Construct interaction—capturing nonadditive effects across facets—something that is not easily accommodated by other multitrait-multimethod approaches. ANOVA = analysis of variance.

important factor in educational research, particularly in studies on science motivation and achievement. Gender differences in science motivation and achievement remain a central concern, with persistent disparities in motivational constructs such as self-concept and aspirations, especially in mathematics and the physical sciences (Eccles, 2014; Marsh & Yeung, 1998; Mullis et al., 2008). These patterns carry critical implications for equity, educational interventions, and STEM-pipeline policies.

Decomposing gender variance: ANOVA framework.

We applied a two-way (Trait × Domain) ANOVA decomposition to the gender-related correlations to examine the relative contributions of traits, domains, and their interaction. The total variance (*SS* = .202, *SE* = .013) was primarily attributable to domain effects (*SS* = .159), followed by smaller but significant trait effects (*SS* = .019, *SE* = .003) and a meaningful trait-by-domain interaction (*SS* = .023, *SE* = .002). These results demonstrate that gender differences in motivation and achievement are not uniform but systematically vary across domains and traits—a pattern uniquely accessible through the MTMD framework.

Latent gender differences across traits and domains.

Table 9 summarizes the gender correlations (coded 0 = girls, 1 = boys) across all 25 trait-domain combinations. The overall mean correlation was modest (*r* = .032, *SE* = .005), slightly favoring boys—however, this average masked substantial domain-specific variation.

Self-concept showed the largest overall gender effect (mean *r* = .081). Domain-specific results varied considerably: Boys reported significantly higher self-concept in physics (.187), earth science (.115), math (.114), and chemistry (.065), whereas girls reported higher self-concept in biology (−.075). This variability produced a significant domain-level sum of squares (*SS* = .038, *SE* = .004), underscoring the importance of disaggregated, facet-level analyses. Other traits also exhibited domain-specific variation: Intrinsic motivation showed a small mean difference (*r* = .010) but the largest domain variation (*SS* = .051) in which boys favored physics and girls favored biology. Extrinsic motivation and aspirations displayed mixed polarity, with higher levels for boys in the physical sciences and for girls in biology. Achievement differences were modest overall (*r* = .029), with consistent advantages for boys in physics and girls in biology.

Implications for broader application.

This extension highlights the practical utility of our EC-F framework combined with the traditional multiple-indicators/multiple-causes model (Jöreskog & Goldberger, 1975) for addressing substantive questions beyond traditional construct validation. For addressing gender equity in STEM education, treating domains as substantively meaningful facets—rather than simply as “method” factors—reveals critical patterns, such as boys’ higher self-concepts in physics and girls’ higher self-concepts in biology. These insights

demonstrate the framework's value in exploring nuanced, policy-relevant patterns of equity. Moreover, gender differences are likely to emerge in other multidimensional constructs that exhibit meaningful domain specificity (e.g., self-efficacy, well-being, emotions, psychological needs, quality of life, Big Five personality traits, coping strategies, locus of control, social support networks). A similar approach could also be used to investigate differences related to other grouping variables, such as socioeconomic status, age, or educational background, offering further opportunities to understand systematic patterns in construct validity and to inform applied interventions.

General Discussion, Summary, and Implications

This investigation revisited the enduring relevance of the Campbell-Fiske guidelines for construct validation (Campbell & Fiske, 1959) and proposed a pragmatic modernization grounded in contemporary psychometric practice. Across a series of analyses, we introduced and applied the EC-F framework, which was implemented within an MTMD structure. This framework is rooted in a fully latent measurement model and extends the logic of traditional MTMM approaches. Our aim was not merely to defend the original guidelines but also to preserve their conceptual strengths while rigorously updating their empirical implementation for contemporary research needs.

Revisiting and extending the Campbell-Fiske logic

By reframing the MTMM structure as an MTMD structure—or more generally, as an MTMX structure, where X can be any relevant second facet (e.g., domains, occasions, raters, contexts, instruments)—we emphasize the importance of treating both facets as theoretically meaningful. This aligns with Campbell and Fiske's (1959) original insight that the distinction between traits and methods is often context-dependent and not always straightforward and with Campbell and O'Connell's (1967) discussion of MTMO designs. Our demonstration using TIMSS data highlights the critical role of domain facets (e.g., biology, chemistry, physics) in the interpretation of constructs. This perspective is also relevant across other areas of psychology, including studies of self-efficacy, well-being, psychological needs, and achievement emotions, cases in which the second facet may capture contexts, occasions, or other relevant dimensions. Highlighting the flexibility of our extended framework, we show that it offers important advantages and is applicable to MTMM designs in which the method effect is subordinate. This reframing underscores the flexibility

and adaptability of our EC-F framework, making it a powerful tool for contemporary applied research across diverse psychological and educational domains.

The extended guidelines as a complementary tool

We emphasize that the purpose of this article is not to juxtapose the strengths and weaknesses of alternative SEM-based MTMM models or to position the EC-F framework as a replacement for SEM-based approaches. Instead, our goal is to highlight the complementary role that the framework can play in bridging the gap between the technical rigor of SEM-based models and the practical interpretability and diagnostic utility offered by the original Campbell-Fiske guidelines (Campbell & Fiske, 1959). We argue that the framework is itself sufficiently rigorous and robust to stand alone in both basic and applied research while also serving as a valuable complement to more statistically oriented SEM studies. Indeed, the application of the EC-F framework—rooted in a fully latent measurement model—naturally leads to more complex SEM applications. Thus, both perspectives are complementary for a comprehensive evaluation of construct validity, particularly in large-scale, applied-social-science research with complex measurement structures.

Centrality of the multiple-indicator measurement model

A core strength of the EC-F framework lies in its explicit commitment to the multiple-indicator measurement model. By starting with a well-fitting CFA model, we ensure that the latent-correlation matrix reflects reliable constructs and properly tests the a priori factor structure that is the focus of the study. If there is no empirical support for this structure, the application of MTMM analyses becomes questionable. This measurement model distinguishes our approach from older manifest-score MTMM designs and from some SEM-based variants that omit or oversimplify the measurement model.

Although the importance of specifying and testing a multiple-indicator measurement model has been noted in the literature (e.g., Marsh & Hocevar, 1988; also see Castro-Schilo et al., 2014; Eid, 2006; Geiser et al., 2014), it remains underused in practice. Our approach explicitly tests this structure and corrects for measurement error at the outset, ensuring that inferences are based on well-specified latent constructs. This aligns with best practices in contemporary psychometrics, ensuring that subsequent analyses accurately reflect true trait-method relationships rather than artifacts of measurement error or rater bias. Once the CFA-based measurement model is estimated, the

resulting latent-correlation matrix—our extended MTMD matrix—serves as the empirical foundation for evaluating the Campbell-Fiske guidelines (Campbell & Fiske, 1959) and their extensions.

Asymptotic parameter comparisons: a practical alternative

The second key enhancement of the EC-F framework transforms the original descriptive Campbell-Fiske guidelines (Campbell & Fiske, 1959) into formal statistical hypotheses. As part of testing the measurement model, we first estimate a fully latent correlation matrix using a CFA-based measurement model. This ensures that the model fits well, the latent constructs are appropriately measured, and measurement error is accounted for. From this latent MTMM matrix, we then specify formal statistical tests of the Campbell-Fiske criteria (Fig. 1) using asymptotic parameter comparisons (e.g., Oehlert, 1992; Raykov, 2011; Raykov & Marcoulides, 2004).

In practical terms, we employ Mplus's (Muthén & Muthén, 2017) model-constraint function to compute aggregate estimates (e.g., the mean of the convergent-validity correlations) and compare these with other aggregates (e.g., the mean of HTHM correlations). These comparisons can be conducted separately for each trait or domain and then averaged across all traits or domains. Although these calculations could also be performed manually outside the model, the asymptotic-parameter-comparison approach has the critical advantage of providing standard errors and significance tests, enabling researchers to formally test each of the Campbell-Fiske guidelines (Campbell & Fiske, 1959; also see Fig. 1) in the measurement model itself.

A key advantage of using the model-constraint function in the SEM environment is that it automatically accounts for the standard errors of parameter estimates and the underlying model structure. This ensures that significance tests are valid even in the presence of model dependencies. In contrast, manually computing differences between observed correlations—especially in manifest-score MTMM matrices—can conflate sampling and measurement error, violate statistical assumptions, and provide no standard errors or confidence intervals for formally testing the comparisons. The asymptotic-parameter-comparison approach aligns seamlessly with the CFA-based measurement model, yielding robust and statistically valid tests that meet contemporary psychometric standards.

This formalization enhances transparency, replicability, and diagnostic clarity while preserving the intuitive interpretability of the latent-correlation matrix illustrated in Figure 1. The approach is straightforward to implement with contemporary statistical software, leverages

robust-estimation methods, and ensures that the tests are both statistically valid and accessible to applied researchers.

Addressing additivity and trait-domain interactions

A key advantage of the EC-F framework is the capacity to assess trait-by-domain interactions—an important aspect of construct validity that Campbell and Fiske (1959) implicitly acknowledged and that underpins Guidelines 4 and 5 (see Fig. 1). Although Campbell and Fiske did not formally define interactions as a statistical term, they conceptually sought to assess potential trait-by-method interactions through empirical patterns in the correlation matrix. In contrast, most SEM-based MTMM models (e.g., CT-CM, CT-UM, CT-CU) assume additivity between trait and method factors, treating them as independent and additive sources of variance. This means these models cannot test or detect systematic interactions between traits and domain or method facets even though such interactions may have substantial theoretical and practical relevance. Our application of the ANOVA decomposition to the latent MTMD matrix (Table 7) explicitly includes a trait-by-domain interaction term, enabling us to test for systematic, nonuniform validity patterns across domains—a critical step for capturing the complexity of construct validity in real-world research settings.

Broader applicability to applied psychological research: empirical versatility of the framework

We designed our EC-F framework to offer applied researchers a robust, accessible, and versatile tool for evaluating construct validity across diverse psychological contexts. Although our case study focused on STEM motivation, the framework applies broadly to multidimensional constructs assessed across substantively meaningful facets—such as measurement method, instrument, time, or context.

Applications in this study demonstrating the framework's flexibility. For the motivational construct validity, we evaluated trait-facet relationships across TIMSS domains (Tables 3 and 6). For the domain-specific motivation-achievement relations, we examined how motivational constructs relate to achievement within and across domains (Table 8). For gender differences, we used the latent MTMD matrix to detect nuanced gender differences across motivational constructs (Table 9). The trait-by-domain interactions were formally tested and captured systematic variation in trait relationships across domains

(Table 7). In the split-half validity analysis, we assessed discriminant validity of domain-specific achievement measures using random split-halves, highlighting the framework's diagnostic power for test development (see Section 3 in the Supplemental Material).

Applications in diverse subfields of psychological science. Our framework lends itself to numerous research possibilities in psychological science, such as the following: personality research—testing Big Five personality traits across contexts or measurement instruments to evaluate trait-by-method interactions (Biesanz & West, 2004; Sims, et al., 2012); organizational psychology—assessing leadership styles or job satisfaction from multiple perspectives (e.g., self, peer, supervisor) across facets (Lance et al., 2002); social psychology—examining self-other agreement to inform self-understanding and interpersonal processes (Marsh & Shavelson, 1985); clinical psychology—assessing symptoms of anxiety or depression across multiple methods and settings (e.g., self-report, clinician ratings, outpatient vs. inpatient) in which trait-by-method interactions can be substantial (Campbell & Fiske, 1959); instructional effectiveness—juxtaposing class-average student ratings with teacher self-evaluations of instructional effectiveness (Marsh & Roche, 1997); developmental psychology—analyzing developmental changes in children's depression and anxiety using self-reports and parent reports (Geiser et al., 2010); sport psychology—comparing multidimensional self-beliefs of elite and the general public across physical, academic, and social domains (Marsh et al., 1995); item-wording effects—testing whether positively and negatively worded items on multidimensional instruments represent substantive distinctions or wording-method artifacts (Marsh et al., 2023); instrument validation—assessing discriminant validity through repeated test-retest administrations (Marsh et al., 2010) or split-half designs (see Section 3 in the Supplemental Material); and jingle-jangle fallacies—evaluating whether parallel labels assigned to factors by different instruments (or different labels for factors measuring the same construct) are appropriate; Marsh, 1994).

These examples underscore the empirical versatility of the EC-F framework in applied psychological research. Its compatibility with standard SEM software facilitates implementation using existing tools, promoting immediate adoption across subfields. By explicitly addressing trait-by-method interactions, the framework enhances diagnostic precision and supports nuanced interpretations of construct validity—critical for advancing theory, informing interventions, and improving measurement practices in contemporary psychological science.

Retirement plans are premature

Our study addresses recent suggestions that the Campbell-Fiske guidelines (Campbell & Fiske, 1959) may

no longer be needed given the prominence of MTMM:SEM approaches (e.g., Helm, 2022; Widaman, 2022). Although we share criticisms of the original guidelines—particularly their reliance on manifest correlations and lack of formal statistical testing—we believe these limitations can be addressed through our latent-variable adaptation. By applying the guidelines within a fully latent MTMD matrix, supported by multiple-indicator models and formal asymptotic comparisons, we preserve the formative spirit of the Campbell and Fiske guidelines while enhancing precision and interpretability.

Rather than discarding the guidelines, we aim to “refresh the bathwater while keeping the baby”—updating their empirical implementation without abandoning their conceptual foundation. Their core contribution—a structured, criterion-based approach to evaluating convergent and discriminant validity—remains uniquely valuable. In this sense, retirement plans are premature. Properly modernized, the Campbell and Fiske guidelines (Campbell & Fiske, 1959) continue to offer a robust and accessible framework for construct validation in contemporary basic and applied research.

Practical guidance for the application of the extended framework

In response to recurring concerns about convergence and practical application of MTMM models, particularly for researchers working with large-scale survey data, in this section, we outline applied guidance for using the EC-F approach. Drawing on both psychometric theory and simulation evidence, we summarize key recommendations regarding the number of traits, methods, indicators, sample size, and treatment of CUs. We aim to clarify minimal requirements and best practices that can support the broader adoption of EC-F models in applied-research contexts.

Number of traits, methods, and indicators. SEM approaches to MTMM data—particularly the “gold standard” CT-CM model—are notoriously subject to convergence problems and improper solutions. This does not apply to our EC-F approach, which involves fitting only a simple measurement model that rarely has convergence issues.

As with all MTMM models, the minimal requirements are at least two traits and at least two methods. However, it is useful to have three or more methods and particularly three or more traits. Like any other CFA model, we recommend having three or more indicators per factor, although it is sometimes possible to fit a model with only two indicators. However, it is always good to have more high-quality indicators of each factor because this enhances estimation stability, reduces bias, and improves construct coverage. Of course, it is technically possible to have one or more factors based on a single indicator;

however, this detracts from the value of the approach and warrants discussion as a limitation (e.g., the single-indicator measures of achievement for each domain in our study).

Sample-size recommendations. Minimal-sample-size recommendations remain a contentious issue in CFA and SEM. Because our EC-F model is merely a CFA measurement model, the same general considerations apply. The only truly uncontroversial advice is that more is always better (Marsh et al., 1995). There is no universally accepted rule for determining the minimum sample size in SEM; however, many ad hoc guidelines exist in the literature. A widely endorsed benchmark is that $N > 200$ generally suffices for fitting moderately complex measurement models, particularly when latent factors have at least three indicators and are well defined (Kline, 2016; Marsh et al., 1995).

Various heuristics are sometimes proposed—for example, 10 to 20 participants per indicator (Marsh et al., 1995; Wolf et al., 2013) or 10:1 ratios of cases to parameters (Bentler & Chou, 1987; Marsh et al., 1995). However, Marsh et al. (1995) criticized such rules for oversimplification and lack of generalizability. In their Monte Carlo simulations, Marsh et al. demonstrated that increasing the number of indicators per factor substantially improves model convergence, estimation precision, and the recovery of population parameters—particularly when sample sizes are modest. Their findings suggest that the number of indicators per trait or method factor is often more important than arbitrary participant-to-parameter ratios.

Accordingly, we adopted a conservative minimum-sample-size guideline of $N > 200$ for our extended MTMD models while emphasizing that appropriate sample size should be evaluated in relation to model complexity, factor reliability, and the extent of missing data. For future applications, we encourage researchers to consider simulation-based power analysis tailored to the model specifications of their own study (e.g., Muthén & Muthén, 2002).

Treatment of CUs. Whereas some MTMM:SEM models use CUs to overcome convergence problems—for example, in “correlated-traits, correlated-uniquenesses” variants of the CT-CM model—this is not necessary in our EC-F approach. As in any CFA measurement model, CUs should be used sparingly and only when they are theoretically justified a priori and supported empirically. Typical a priori justifications include cases in which different latent constructs are measured using parallel-worded items that may share residual variance beyond the latent traits or methods or in longitudinal models in which identical or similarly worded items are administered across time. In such cases, shared residual variance across repeated item measurements is not captured by the trait factors and

should be modeled as CUs to avoid inflating stability estimates or misattributing change (Jöreskog, 1971; Little et al., 2007; Marsh & Hau, 1996). CUs should not be added post hoc merely to improve model fit or aid convergence because this undermines the interpretability of trait-method decompositions. In our EC-F models, all CUs were prespecified based on item wording, consistent with best practices.

Multiple-group, longitudinal, and multi-informant considerations. When applying our EC-F framework in multigroup, longitudinal, or multiinformant settings, it is generally appropriate to test for measurement invariance (Meredith, 1993; Millsap, 2011) to ensure that latent traits are measured equivalently across the relevant groups, occasions, or informants. In multigroup designs (e.g., across countries, as in the present investigation), invariance testing supports valid cross-group comparisons.

In some situations, invariance tests in the EC-F model are appropriate for testing invariance across multiple methods. In longitudinal applications—in which waves of data collection (time) serve as methods—tests of longitudinal invariance are appropriate to ensure that trait scores retain the same meaning across time.

The EC-F model is also well suited to multi-informant contexts, which can involve either interchangeable informants (e.g., several employees rating the same supervisor) or identifiably distinct informants (e.g., self, parent, and teacher ratings of each child; see Kenny, 2022; Widaman, 2022). In each case, tests of measurement invariance are informative: For interchangeable informants, they assess whether informants are indeed interchangeable; for distinguishable informants, they provide a basis for testing theoretically predicted patterns of agreement and divergence across perspectives.

However, some scholars have cautioned against the uncritical use of invariance tests in multi-informant designs. For example, De Los Reyes et al. (2022) argued that informant discrepancies may reflect meaningful, domain-relevant differences—such as variations in observation context or informant perspective—rather than mere measurement bias. In such cases, strict invariance constraints may obscure substantively important differences. Accordingly, we recommend that researchers interpret invariance tests within the broader theoretical context of informant divergence and consider model variants, such as partial invariance, sensitivity analyses, or Trait \times Method interactions, as appropriate.

Conclusion and Recommendations

Our EC-F guidelines offer an empirically grounded, conceptually coherent alternative for construct validation. They preserve the diagnostic clarity of the original guidelines while addressing long-standing psychometric

limitations by using a well-specified latent-measurement model and formal tests of statistical significance to evaluate each guideline.

Although the EC-F framework fully supports rigorous construct validation as a stand-alone approach, it is also complementary to SEM-based MTMM models. By beginning with a robust, multiple-indicator measurement model and providing formal statistical tests of validity criteria, our framework aligns naturally with SEM principles. This complementarity ensures that researchers can flexibly address a wide range of research questions—using the extended framework when conceptual clarity and diagnostic transparency are paramount—and integrate it with SEM-based MTMM models as appropriate. In this way, the extended framework preserves its status as a stand-alone approach while remaining fully compatible with more advanced SEM strategies.

We aim to provide a flexible, interpretable, and practical approach supporting applied researchers in diverse settings. By grounding construct validation in a latent MTMM (or MTMD) correlation matrix, the framework enables researchers to foreground domain-specific interpretations, conduct formal asymptotic comparisons, and evaluate trait-by-domain interactions—all while building on a validated measurement model that strikes a balance between interpretability and rigor.

To support best practices, we recommend that future MTMM-based studies (a) begin with a multiple-indicator CFA model to ensure a valid measurement structure, (b) use the latent MTMM (or MTMD) matrix as the empirical foundation for construct validation, (c) apply the extended guidelines to formalize convergent and discriminant validity criteria, (d) treat both facets (e.g., traits and methods/domains) as theoretically meaningful unless theory justifies asymmetry, and (e) use asymptotic parameter comparisons for transparent, statistically grounded evaluations.

The EC-F guidelines provide a practical roadmap that strikes a balance between conceptual fidelity and empirical rigor. They support decisions in scale development, program evaluation, and policy interpretation—particularly in fields such as psychology, education, public health, and the social sciences, in which multidimensional constructs and diverse data sources are common. The approach's interpretability, low reliance on complex modeling, and capacity to diagnose construct validity at the facet level make it especially well suited for applied contexts. At the same time, because all analyses are anchored in a validated latent-measurement model and supported by formal statistical comparisons, results generated using this approach meet the evidentiary standards of leading scholarly journals—ensuring both scientific rigor and practical utility.

Transparency

Action Editor: Patrick Manapat

Editor: David A. Sbarra

Author Contributions

Herbert W. Marsh: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Jiesi Guo: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing.


Oliver Lüdtke: Conceptualization; Investigation; Methodology; Writing - original draft; Writing – review & editing.

Reinhard Pekrun: Conceptualization; Investigation; Methodology; Project administration; Writing – original draft; Writing – review & editing.


Declaration of Conflicting Interests


The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iDs

Herbert W. Marsh  <https://orcid.org/0000-0002-1078-9717>

Jiesi Guo  <https://orcid.org/0000-0003-2102-803X>

Oliver Lüdtke  <https://orcid.org/0000-0001-9744-3059>

Reinhard Pekrun  <https://orcid.org/0000-0003-4489-3827>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/25152459251407652>

References

- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117. <https://doi.org/10.1177/0049124187016001004>
- Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality*, *72*(4), 845–876. <https://doi.org/10.1111/j.0022-3506.2004.00282.x>
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press. <https://doi.org/10.4324/9780203774762>
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge/Taylor & Francis. <https://doi.org/10.4324/9780203807644>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. <https://doi.org/10.1037/h0046016>
- Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, *2*, 409–426.
- Castro-Schilo, L., Widaman, K. F., & Grimm, K. J. (2014). Toward improving the modeling of MTMM data: A response

- to Geiser, Koch, and Eid (2014). *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 524–533. <https://doi.org/10.1080/10705511.2014.919818>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. <https://doi.org/10.1037/h0044919>
- De Los Reyes, A., Tyrell, F. A., Watts, A. L., & Asmundson, G. J. G. (2022). Conceptual, methodological, and measurement factors that disqualify used of measurement invariance techniques to detect informant discrepancies in youth mental health assessments. *Frontiers in Psychology*, 13, Article 931296. <https://doi.org/10.3389/fpsyg.2022.931296>
- Eccles, J. S. (2014). Gender and achievement choices. In E. T. Gershoff, R. W. Mistry, & D. A. Crosby (Eds.), *Societal contexts of child development: Pathways of influence and implications for practice and policy* (pp. 19–36). Oxford University Press.
- Eid, M. (2006). Methodological approaches for analyzing multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 223–230). American Psychological Association.
- Eid, M., & Nussbeck, F. W. (2009). The multitrait-multimethod matrix at 50! *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 5(3), Article 71. <https://doi.org/10.1027/1614-2241.5.3.71>
- Geiser, C., Eid, M., & Nussbeck, F. W. (2010). On the meaning of latent variables in psychological research. *Psychological Methods*, 15(2), 113–131. <https://doi.org/10.1037/a0018577>
- Geiser, C., Koch, T., & Eid, M. (2014). Data-generating mechanisms versus constructively defined latent variables in multitrait-multimethod analysis: A comment on Castro-Schilo, Widaman, and Grimm (2013). *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 509–523. <https://doi.org/10.1080/10705511.2014.919816>
- Grayson, D., & Marsh, H. W. (1994). Identification with deficient rank loading matrices in confirmatory factor analysis: Multitrait-multimethod models. *Psychometrika*, 59, 121–134. <https://doi.org/10.1007/BF02294271>
- Helm, J. L. (Ed.). (2022). *Advanced multitrait-multimethod analyses for behavioral and social sciences*. Routledge.
- Hoyt, W. T., & Melby, J. N. (1999). Dependability of measurement in counseling psychology: An introduction to generalizability theory. *The Counseling Psychologist*, 27(3), 325–352. <https://doi.org/10.1177/0011000099273003>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- IEA TIMSS. (n.d.). *TIMSS 2007*. <https://www.iea.nl/studies/iea/timss/2007>
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2), 109–133.
- Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351), 631–639. <https://doi.org/10.1080/01621459.1975.10482485>
- Kavanagh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. *Psychological Bulletin*, 75, 34–49. <https://doi.org/10.1037/h0030412>
- Kenny, D. A. (1979). *Correlation and causality*. Wiley.
- Kenny, D. A. (2022). Multitrait-multimethod matrix: Method in the madness. In J. L. Helm (Ed.), *Advanced multitrait-multimethod analyses for the behavioral and social sciences* (pp. 16–27). Routledge.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165–172.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Lance, C. E., Dawson, B., Birklebach, D., & Hoffman, B. J. (2010). Method effects, measurement error, and substantive conclusions. *Organizational Research Methods*, 13, 435–455.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 7, 228–244.
- Lance, C. E., & Vandenberg, R. J. (2010). If it ain't trait it must be method: (Mis)Application of the multitrait-multimethod design in organizational research. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *A handbook of international large-scale assessment data analysis* (pp. 317–344). Routledge.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31(4), 357–365. <https://doi.org/10.1177/0165025407077757>
- Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook* (pp. 570–578). Pergamon Press.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13, 335–361. <https://doi.org/10.1177/014662168901300401>
- Marsh, H. W. (1994). Sport motivation orientations: Beware of jingle-jangle fallacies. *Journal of Sport and Exercise Psychology*, 16(4), 365–380. <https://doi.org/10.1123/jsep.16.4.365>
- Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. British Psychological Society.
- Marsh, H. W., Fraser, M. I., Rakhimov, A., Ciarrochi, J., & Guo, J. (2023). The bifactor structure of the Self-Compassion Scale: Bayesian approaches to overcome exploratory structural equation modeling (ESEM) limitations. *Psychological Assessment*, 35(8), 674–691. <https://doi.org/10.1037/pas0001247>
- Marsh, H. W., & Grayson, D. (1991). *Multitrait-multimethod data: An evaluation of five analytic approaches*. ERIC. <https://files.eric.ed.gov/fulltext/ED342797.pdf>
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Sage.
- Marsh, H. W., & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, 64, 364–390. <https://doi.org/10.1080/00220973.1996.10806604>

- Marsh, H. W., Hau, K.-T., & Balla, J. R. (1995). Is more ever too much: The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 30*, 1–23.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.
- Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20*(3), 231–248. <https://doi.org/10.1111/j.1745-3984.1983.tb00202.x>
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*, 107–117. <https://doi.org/10.1037/0021-9010.73.1.107>
- Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport and Exercise Psychology, 32*(4), 438–482. <https://doi.org/10.1123/jsep.32.4.438>
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187–1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*(3), 107–123. <https://doi.org/10.1207/s15326985ep2003>
- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal, 35*(4), 705–738. <https://doi.org/10.3102/00028312035004705>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Millsap, R. E. (1990). A cautionary note on the detection of method variance in multitrait-multimethod data. *Journal of Applied Psychology, 75*, 350–353. <https://doi.org/10.1037/0021-9010.75.3.350>
- Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives, 4*(1), 5–9. <https://doi.org/10.1111/j.1750-8606.2009.00109.x>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
- Millsap, R. E. (2014). The statistical analysis of method effects in multitrait-multimethod data: A review. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 93–110). Psychology Press. <https://doi.org/10.4324/9781315806815-8>
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*(4), 297–334. <https://doi.org/10.1177/014662169301700401>
- Mullis, I. V. S., Martin, M. O., Olson, J. F., Berger, D. R., & Stanco, G. M. (Eds.). (2008). *TIMSS 2007 encyclopedia: A guide to mathematics and science education around the world* (Vols. 1 and 2). TIMSS & PIRLS International Study Center, Boston College.
- Muthén, B., & Muthén, L. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Oehlert, G. W. (1992). A note on the delta method. *The American Statistician, 46*(1), 27–29. <https://doi.org/10.1080/00031305.1992.10475842>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Raykov, T. (2011). Evaluation of convergent and discriminant validity with multitrait-multimethod correlations. *The British Journal of Mathematical and Statistical Psychology, 64*(1), 38–52. <https://doi.org/10.1348/000711009X478616>
- Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling, 11*, 621–637. https://doi.org/10.1207/s15328007sem1104_6
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*(1), 1–22. <https://doi.org/10.1177/014662168601000101>
- Şimşek, Ö. F., Koydemir, S., & Schütz, A. (2012). A multigroup multitrait-multimethod study in two countries supports the validity of a two-factor higher order model of personality. *Journal of Research in Personality, 46*(4), 442–449. <https://doi.org/10.1016/j.jrp.2012.04.005>
- Stanley, J. C. (1961). Analysis of unreplicated three-way classifications to rater bias and trait independence. *Psychometrika, 26*, 205–219. <https://doi.org/10.1007/BF02289209>
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1–26. <https://doi.org/10.1177/014662168500900101>
- Widaman, K. F. (2022). Musings on alternate confirmatory factor models for multitrait-multimethod data. In J. L. Helm (Ed.), *Advanced multitrait-multimethod analyses for the behavioral and social sciences* (pp. 51–79). Routledge. <https://doi.org/10.4324/9780429320989-4>
- Woehr, D. J., Putka, D. J., & Bowler, M. C. (2012). An examination of G-theory methods for modeling multitrait-multimethod data: Clarifying links to construct validity and confirmatory factor analysis. *Organizational Research Methods, 15*(1), 134–161. <https://doi.org/10.1177/1094428111425699>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934. <https://doi.org/10.1177/0013164413495237>