International Journal of Basic and Applied Sciences, 14 (8) (2025) 520-530

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/tc3wv638
Research paper

A Hierarchical and Finite Mixture Modeling Approach to Mathematics Achievement: Evidence from PISA 2022

Pei-Ching Chao [1], Joseph Meng-Chun Chin [2], Gregory S. Ching [2] *

[1] *Center for Teacher Education, Fu Jen Catholic University, New Taipei City, Taiwan*
[2] *Graduate Institute of Educational Administration and Policy, National Chengchi University, Taipei City, Taiwan*
**Corresponding author E-mail: gching@nccu.edu.tw*

Received: November 21, 2025, Accepted: December 19, 2025, Published: December 23, 2025

Abstract

This study investigates the combined effects of cognitive, psychosocial, and cross-subject academic indicators on students' mathematics achievement across 38 countries that participated in the PISA 2022 creative thinking assessment. Drawing on data from 144,446 students, we employed hierarchical linear modeling to examine how reading and science proficiency, engagement in creative activities (in and out of school), perseverance, curiosity, and socioeconomic status (ESCS) predict mathematical performance. The results show that reading and science are robust predictors of mathematics scores. ESCS and perseverance also demonstrated consistent positive effects, while creativity showed context-specific associations, positive in some clusters and negative in others. To identify latent cross-national typologies, we applied both K-means clustering and Gaussian Mixture Modeling (GMM) to country-level aggregates. Model comparison using the Bayesian Information Criterion (BIC) favored the GMM solution, which was subsequently used to group countries for multigroup structural equation modeling (MG-SEM). Results revealed significant variations in predictor effects across clusters, highlighting heterogeneity in pathways to mathematics success. This study contributes to comparative education research by integrating hierarchical regression and latent classification techniques, offering implications for instructional design and international education policy aimed at promoting mathematical literacy across diverse systems.

Keywords: *Mathematics Achievement; Hierarchical Linear Modeling; Gaussian Mixture Model; Clustering and Classification; Bayesian Information Crite-Rion; Multigroup Structural Equation Modeling; Educational Data Science.*

1. Introduction

Understanding the predictors of student achievement in mathematics has long been a central problem in educational measurement and learning analytics (Namoun & Alshanqiti, 2021; Rajak et al., 2020). Formally, this can be posed as a prediction and classification problem in discrete mathematical sciences, where the task is to model the mapping $X \mapsto Y$ with nested dependence across countries and to partition systems into latent classes based on observed covariates. With the growing availability of international large-scale assessment datasets such as Programme for International Student Assessment (PISA), mathematical modeling techniques can be used not only to identify performance factors (Korres & Tsami, 2010; Wu et al., 2020), but also to characterize structural variation across national education systems (Bayirli et al., 2023). The 2022 cycle of PISA provides a rich opportunity to analyze the relationships between mathematics proficiency and a range of cognitive and psychosocial indicators, including cross-subject achievement (reading, science), perseverance, and creativity-related behavior and traits (OECD, 2024). From a modeling standpoint, this setting invites the application of hierarchical and mixture-based approaches to address nested and heterogeneous data structures.
This study applies a multi-method quantitative framework to predict mathematics achievement among 144,446 students from 38 countries who participated in the PISA 2022 creative thinking module (OECD, 2024). We modeled the relationship between mathematics performance and its predictors using Hierarchical Linear Modeling (HLM) to account for the nested structure of students within countries.

$$Y_i = \alpha_{g(i)} + X_i^T \beta + \epsilon_i, \alpha_g \sim N(\alpha_0, \tau^2), \epsilon_i \sim N(0, \sigma^2)$$

This mixed-effects formulation provides both fixed estimates of predictors and random intercepts capturing country-level variation.
To explore latent typologies of education systems, we compared K-means and Gaussian Mixture Modeling (GMM) using country-level means of cognitive and psychosocial indicators. Based on model fit criteria, particularly the Bayesian Information Criterion (BIC), GMM was selected and used to generate clusters of educational systems.

$$f(z) = \sum_{k=1}^{K} \pi_k \phi\,(z|\mu_k, \Sigma_k)$$

$\sum_{k=1}^{K} \pi_k = 1$

where Z_g are country-level aggregates and BIC is used to select K. This yields a discrete partition of systems into clusters with probabilistic membership. These clusters then served as grouping variables in Multigroup Structural Equation Modeling (MG-SEM) to examine whether the strength and direction of predictor effects varied across educational profiles.

$$Y_i^{(k)} = \gamma_{0k} + X_i^{(k)\top}\gamma_k + \xi_i^{(k)}, i \in G_k, k = 1, \dots, K$$

where $Y_i^{(k)}$ is math achievement for student i in cluster k, $X_i^{(k)}$ are predictors, γ_k are group-specific path coefficients, $\xi_i^{(k)}$ are residuals. We therefore compared an unconstrained model $H_1: \gamma_1, \dots, \gamma_K$ free against a constrained model $H_0: \gamma_1 = \dots = \gamma_K$, with fit evaluated through Δx^2, CFI, and RMSEA. This framework is well-suited in detecting both fixed and random effects, as well as latent heterogeneity in global learning environments (Khine et al., 2020; Lee & Stankov, 2023).

Grounded in an integrative modeling framework that combines statistical inference and unsupervised machine learning (Lezhnina & Kismihók, 2022), the study draws on a multilevel regression foundation (Sun et al., 2012; You et al., 2021) and applies clustering methods to derive typologies based on country-level aggregates (Belciu et al., 2024). The GMM derived clusters were then analyzed using MG-SEM to assess structural invariance (Singh & Kathuria, 2023). While the conceptual foundation draws on psychological and educational theories: such as Cross-Domain Transfer (Doumas et al., 2022) and Metacognitive Engagement (Li & Lajoie, 2022), the novelty lies in its methodological synthesis. This includes linear mixed-effects modeling, model evaluation using information criteria, and the application of finite mixture modeling to uncover latent structural variation. Such techniques are especially pertinent to mathematical modeling in the social sciences, where precision and generalizability are both essential (Toker & Green, 2021). The contribution of this study lies in treating the educational prediction task as a discrete classification-and-regression framework, integrating (i) multilevel modeling, (ii) mixture-based clustering with explicit likelihood and information criteria, and (iii) group-specific structural modeling. The framework demonstrates how discrete structures such as partitions, overlap matrices, and invariance constraints can be combined to yield policy-relevant insights.

Overall, the study advances the use of integrated regression and classification techniques in international education research. By combining HLM, GMM, and MG-SEM, it demonstrates how nested dependence structures, latent heterogeneity, and contextualized predictor effects can be addressed within a unified framework. The findings offer empirical insight into mathematics learning across global contexts and establish a transferable modeling paradigm for interdisciplinary educational data science. From a discrete mathematical perspective, the research objectives (ROs) can be formalized as follows:

- RO1: To model the relationship between reading performance and mathematics achievement using a two-level hierarchical linear model.
- RO2: To extend the model by incorporating science achievement and evaluating its incremental predictive power.
- RO3: To construct a full model including creativity-related activity engagement, perseverance, curiosity, and ESCS, and estimate their combined effects on mathematics achievement.
- RO4: To identify and compare latent clusters of countries based on cognitive and psychosocial indicators using K-means and GMM, and to adopt the optimal solution for further analysis.
- RO5: To test the invariance of predictor effects across the derived clusters using MG-SEM.

2. Theoretical Background

2.1. Cross-subject predictors of mathematics achievement

Previous studies have established the predictive relationship between proficiency in reading and science and students' performance in mathematics (Korpershoek et al., 2015; Zhu, 2022). In fact mathematics and science achievement are quite correlated (Hansen & Gonzalez, 2014). While reading comprehension have also shown to have significant relationship with both mathematics and sciences performance (Akbasli et al., 2016; Caponera et al., 2016; Peng et al., 2020). Some other studies have even noted the impact of technology resources or competencies with mathematics achievement within the PISA 2022 data (Chao & Ching, 2025). Besides the studies using data from PISA, other international assessments such as Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS), have consistently shown that reading skills contribute to mathematical problem comprehension, especially for word problems and multi-step reasoning tasks (Gomez et al., 2020; Vilenius-Tuohimaa et al., 2008). Likewise, science proficiency correlates with math achievement due to overlapping demands in logical reasoning, data interpretation, and conceptual abstraction (Jonsson et al., 2020; Pasigon, 2024). The notion of cross-domain transfer underpins these relationships, suggesting that foundational skills in one subject can enhance learning in others through shared cognitive structures. In this study, reading and science scores are modeled hierarchically as predictors of mathematics achievement across countries, forming the cognitive backbone of the predictive framework (RO1–RO2). From a modeling perspective, these cross-domain effects are treated as covariates in a hierarchical regression framework, where their coefficients can be estimated while accounting for the nested dependence of students within countries.

2.2. Psychosocial predictors: perseverance, curiosity, and socioeconomic status

Beyond cognitive skills, affective and dispositional traits play a significant role in shaping mathematics outcomes (Awofala et al., 2022; Kamid et al., 2021). In addition, perseverance; often operationalized as grit or task persistence, has been positively associated with sustained engagement in mathematics tasks and improved outcomes in challenging settings (DiNapoli, 2023; Yu et al., 2021). While, curiosity; a motivational trait reflecting interest in exploring novel problems (Spielberger & Starr, 1994), has mixed effects, such as supporting exploratory learning (Tang et al., 2025), but potentially unclear distinction with curiosity from interest and for supporting motivation in learning mathematics (Peterson & Cohen, 2019). Meanwhile, socioeconomic status (SES) remains one of the most consistent predictors of academic success with multiple pathways including access to resources, parental involvement, and school quality (Galindo & Sonnenschein, 2015; Muñez et al., 2021). For the current study, these psychosocial traits are incorporated into the full model (RO3), offering a broader view of student-level variation in mathematics performance within and across countries. Formally, these psychosocial variables extend the predictor vector X_i in the regression model, allowing tests of whether dispositional factors contribute additively or interactively with cognitive predictors in shaping outcomes.

2.3. Creativity and problem-solving in mathematics

The role of creativity in mathematics has gained increasing attention, particularly with the inclusion of creative thinking module in PISA 2022 (OECD, 2023b). Creative problem-solving in mathematics involves flexible reasoning, divergent thinking, and the ability to construct multiple solution paths (de Vink et al., 2022; Hadar & Tirosh, 2019; Suherman & Vidákovich, 2022). However, the operationalization of creativity in standardized testing contexts is complex (Kaufman et al., 2023), and its direct predictive power on mathematics performance remains uncertain (Gajda et al., 2017). Some studies suggest creativity enhances mathematical reasoning (Lithner, 2017), while others caution against overgeneralizing its effects across cultural contexts (Jonsson et al., 2020). In the current study, creativity is operationalized through students' frequency of engagement in creative activities, both within and outside of school, as measured by the PISA student questionnaire. While this construct captures behavioral aspects of creativity exposure rather than cognitive problem-solving itself, it nonetheless provides insight into the extent to which creative environments may support mathematical learning (Davies et al., 2013; Niu et al., 2022). Creativity-related dispositions are tested as part of the extended model (RO3), enabling evaluation of their added value to predictive accuracy and their contextual variability in the succeeding multigroup modeling (RO5). In the mathematical framework, creativity indices are included as elements of X_i. Their coefficients are estimated in both pooled and group-specific models, and variation in their signs across clusters signals structural non-invariance.

2.4. Modeling educational heterogeneity: hierarchical and latent approaches

Mathematics achievement varies not only due to individual-level factors, but also due to systemic and cultural differences across countries (He et al., 2017; OECD, 2023b). HLM is suited for nested data structures, enabling partitioning of variance across levels (Raudenbush & Bryk, 2002). Additionally, latent class and mixture modeling techniques such as K-means and GMM allow for unsupervised grouping of systems based on multivariate patterns, revealing hidden structures in educational data (Alshabandar et al., 2018; Liu et al., 2022; Sideridis et al., 2021). Importantly, recent studies have used MG-SEM to explore whether the strength of predictors differs across clusters or regions (André et al., 2020; Byrne, 2012). This study combines these techniques (RO4–RO5) to model both fixed predictor effects and structural heterogeneity, contributing to methodological advances in comparative education analytics. Technically, K-means yields partitions by minimizing within-cluster variance in Euclidean space, whereas GMM yields partitions by maximizing the likelihood of a finite mixture density with BIC selecting the optimal K. These methods define a discrete structure over the set of countries, which then serves as the grouping basis in MG-SEM.

2.5. Synthesis and research gap

Prior studies have rarely integrated cognitive, psychosocial, and creativity-related factors into a unified predictive framework of mathematics achievement, particularly using large-scale international datasets like PISA. While many investigations focus on within-country predictors or rely on aggregated cross-national comparisons, few examine latent educational typologies or account for contextual variability in predictor effects across countries. Additionally, creativity is often conceptualized narrowly as problem-solving ability, with less attention paid to behavioral engagement in creative activities; a construct increasingly relevant in contemporary curricula. The current study addresses these gaps by: (1) constructing hierarchical models that incorporate cross-subject academic performance, psychosocial traits, creative activity engagement, and socioeconomic status; and (2) applying unsupervised clustering and MG-SEM to uncover and test cross-national variation in predictive relationships. In doing so, the study advances the use of scalable, statistically rigorous methods that support both theoretical insight and policy relevance in globally diverse educational systems. In discrete mathematical terms, the novelty of this study lies in formalizing the prediction task as a classification-and-regression pipeline: hierarchical modeling captures nested variance, mixture-based clustering generates partitions, and MG-SEM tests invariance across these partitions.

3. Methodology

3.1. Study design and data source

This study employed a cross-sectional, secondary data analysis design using publicly available data from the 2022 PISA (OECD, 2023a). The dataset comprises student-level responses from 38 countries that participated in the Creative Thinking assessment module. A total of 144,446 students were included after list wise deletion of cases with missing data on key predictors. The data structure is inherently hierarchical, with students (Level 1) nested within countries (Level 2), thus justifying the use of multilevel modeling techniques (Raudenbush & Bryk, 2002).

Table 1 presents the descriptive statistics of the 144,446 students from 38 countries included in the analytic sample. In addition, table 1 descriptive patterns provide the empirical basis for the subsequent clustering analysis by illustrating cross-national variation in both cognitive and psychosocial indicators. Country-level means and standard deviations (SD) are reported for the primary outcome variable, mathematics achievement, as well as for all cognitive, psychosocial, and contextual predictors. As shown, mean mathematics scores ranged from 400.3 (Brazil) to 562.6 (Hong Kong), with a grand mean of 481.7 (SD = 91.46). Corresponding average scores in reading and science were 487.3 and 494.8, respectively. With respect to psychosocial variables, country means for creativity engagement ranged from -0.300 (Belgium) to 0.567 (Peru), while perseverance and curiosity showed moderate variation, with grand means of 0.024 and 0.052, respectively. Notably, ESCS (Economic, Social, and Cultural Status) varied widely, ranging from -1.149 (Turkey) to 0.471 (Australia and Iceland), reflecting substantial socioeconomic diversity across educational systems. The inclusion of these standardized predictors, scaled with approximately zero mean and unit variance, supports robust multilevel and multigroup estimation. These descriptive patterns also foreshadow the clustering results and cross-country modeling strategies adopted in later sections.

Table 1: Descriptive Statistics

Country	n	Math	(SD)	Reading	(SD)	Science	(SD)	Creative	(SD)	Perseverance	(SD)	Curiosity	(SD)	ESCS	(SD)
Argentina	3716	419.98	(69.18)	453.45	(82.11)	454.27	(76.82)	0.245	(0.93)	-0.039	(0.93)	0.062	(1.00)	-0.367	(1.09)
Australia	9368	502.49	(91.60)	516.94	(97.38)	524.28	(97.79)	-0.145	(0.74)	-0.097	(0.90)	0.002	(0.93)	0.471	(0.81)
Belgium	3485	510.68	(84.65)	499.06	(86.73)	510.01	(86.18)	-0.300	(0.74)	-0.069	(0.88)	-0.187	(0.86)	0.202	(0.87)
Brazil	4081	400.35	(73.40)	437.74	(89.85)	431.94	(85.78)	0.508	(1.05)	0.027	(0.96)	0.150	(0.98)	-0.769	(1.09)
Bulgaria	2535	446.41	(90.53)	437.62	(99.13)	449.28	(87.64)	0.406	(1.06)	-0.024	(1.06)	-0.002	(0.99)	-0.120	(0.98)

Canada	12503	504.11	(85.75)	518.56	(93.46)	523.98	(87.52)	-0.047	(0.87)	0.009	(1.02)	0.050	(1.01)	0.407	(0.75)
Chile	2482	447.15	(75.58)	482.82	(87.02)	482.80	(85.33)	0.113	(0.94)	0.228	(1.12)	0.256	(1.06)	-0.092	(0.98)
Colombia	3451	411.00	(68.21)	442.34	(83.54)	442.70	(79.72)	0.537	(1.15)	0.305	(0.98)	0.315	(1.06)	-0.768	(1.20)
Croatia	3120	476.80	(81.19)	491.37	(77.46)	497.17	(84.51)	-0.172	(1.04)	-0.024	(0.99)	-0.045	(0.95)	-0.091	(0.82)
Finland	4942	496.85	(82.36)	507.23	(90.18)	524.95	(96.26)	-0.115	(0.86)	0.035	(1.00)	-0.151	(0.88)	0.292	(0.80)
France	2615	496.28	(80.92)	501.20	(89.65)	512.35	(89.49)	-0.305	(0.78)	0.027	(1.03)	0.137	(1.03)	0.128	(0.90)
Greece	2730	437.71	(77.47)	445.28	(83.27)	449.15	(81.62)	0.291	(0.98)	0.062	(0.95)	0.209	(0.93)	-0.078	(0.91)
Hong Kong	3305	562.65	(96.53)	520.96	(85.24)	539.38	(82.35)	-0.030	(0.81)	-0.214	(0.79)	-0.072	(0.91)	-0.367	(0.99)
Hungary	2931	495.24	(86.75)	496.48	(91.14)	508.73	(87.81)	-0.129	(0.87)	0.026	(1.00)	0.025	(0.98)	0.185	(0.91)
Iceland	1290	474.68	(77.71)	456.46	(91.26)	458.51	(83.19)	0.041	(0.91)	0.227	(1.13)	0.010	(1.00)	0.471	(0.72)
Ireland	4104	501.59	(73.50)	528.06	(78.46)	515.30	(82.98)	-0.038	(0.67)	0.002	(0.92)	0.012	(0.89)	0.381	(0.79)
Jamaica	1104	406.70	(68.83)	453.15	(93.71)	441.56	(90.42)	0.234	(1.00)	0.121	(1.06)	0.417	(1.14)	-0.357	(0.93)
Korea	3698	538.41	(99.94)	524.26	(90.48)	538.57	(94.10)	0.191	(0.68)	-0.033	(1.01)	0.141	(1.10)	0.275	(0.80)
Lithuania	3715	480.60	(81.19)	478.99	(84.08)	489.51	(84.67)	-0.246	(0.91)	-0.150	(0.84)	-0.235	(0.88)	0.063	(0.86)
Macao	3115	558.26	(86.29)	517.26	(79.31)	550.47	(79.63)	-0.009	(0.69)	-0.110	(0.83)	-0.076	(0.91)	-0.406	(0.92)
Malta	1675	486.09	(87.46)	470.00	(93.40)	486.70	(89.23)	-0.021	(0.87)	-0.020	(1.04)	0.116	(0.98)	0.072	(0.94)
Mexico	2708	408.22	(63.26)	428.81	(75.97)	421.65	(67.50)	0.110	(1.03)	0.258	(1.03)	0.297	(1.10)	-0.737	(1.13)
Moldova	3571	430.97	(72.44)	431.19	(77.02)	434.13	(74.83)	0.247	(0.96)	0.048	(0.91)	0.114	(0.94)	-0.422	(0.93)
Montenegro	2268	429.75	(76.56)	434.74	(80.91)	427.56	(75.82)	0.310	(1.07)	0.076	(1.07)	-0.006	(0.95)	-0.078	(0.81)
New Zealand	2831	504.52	(88.57)	531.91	(92.94)	532.76	(93.19)	-0.168	(0.70)	-0.138	(0.89)	0.014	(0.95)	0.335	(0.87)
Peru	2154	411.92	(70.16)	432.93	(78.27)	430.65	(76.50)	0.567	(1.05)	0.303	(1.03)	0.328	(1.01)	-0.785	(1.18)
Portugal	3218	491.78	(80.95)	496.18	(79.83)	504.67	(81.22)	-0.297	(0.85)	0.285	(0.96)	0.304	(0.96)	-0.098	(1.12)
Romania	3686	459.77	(88.26)	460.97	(85.88)	459.55	(85.46)	0.380	(0.99)	0.171	(1.02)	0.236	(0.97)	-0.111	(0.95)
Serbia	2866	456.20	(80.99)	460.16	(79.57)	465.43	(80.48)	0.064	(1.10)	0.011	(0.97)	-0.072	(0.93)	-0.150	(0.80)
Slovak	3028	488.92	(88.94)	472.19	(89.52)	486.67	(89.13)	0.093	(1.05)	-0.205	(0.94)	-0.126	(0.89)	-0.151	(0.88)
Slovenia	3802	488.39	(81.18)	475.99	(84.58)	505.13	(84.82)	-0.025	(0.93)	-0.191	(0.98)	-0.237	(0.94)	0.232	(0.83)
Spain	15516	491.90	(77.13)	494.15	(82.02)	501.04	(79.69)	-0.070	(0.79)	0.147	(0.98)	0.113	(0.98)	0.069	(0.95)
Switzerland	2495	527.49	(89.10)	505.34	(96.89)	520.61	(91.90)	0.003	(0.80)	0.087	(1.01)	0.036	(0.99)	0.301	(0.89)
Taiwan	2178	542.81	(109.06)	511.13	(97.57)	536.77	(98.96)	0.211	(0.82)	0.036	(0.99)	-0.007	(0.98)	-0.209	(0.91)
Turkey	4261	463.20	(86.51)	464.97	(80.77)	486.25	(84.04)	0.170	(0.85)	0.133	(1.06)	0.324	(1.10)	-1.149	(1.16)
UK	6438	501.76	(89.05)	514.68	(92.20)	514.07	(93.91)	-0.236	(0.71)	-0.136	(0.93)	-0.076	(0.93)	0.186	(0.87)
Ukrainian	1731	454.55	(81.75)	443.95	(82.22)	465.01	(80.06)	0.225	(0.98)	-0.066	(0.89)	-0.191	(0.90)	-0.213	(0.81)
Uruguay	1730	442.20	(72.15)	475.00	(79.81)	473.17	(78.03)	0.261	(1.12)	0.090	(0.97)	0.164	(1.02)	-0.541	(1.12)
TOTAL	144446	481.73	(91.46)	487.33	(92.27)	494.76	(92.35)	0.030	(0.91)	0.024	(0.98)	0.052	(0.98)	-0.026	(1.00)

Notes. SD = standard deviation. Countries are arranged alphabetically.

3.2. Study variables

Outcome Variable

Mathematics Achievement (MATH_MEAN): The primary outcome of interest was students' mathematical proficiency, operationalized using plausible values derived from the PISA 2022 mathematics assessment. Specifically, ten plausible values (PV1MATH–PV10MATH) were provided for each student by the OECD to account for measurement error inherent in large-scale testing (OECD, 2024). In this study, a simple average of these ten plausible values was computed for each student to form a composite indicator (MATH_MEAN) of overall mathematics achievement. This approach is appropriate for multilevel and exploratory modeling contexts where the focus is on prediction rather than population-level inference.

To further clarify, PISA plausible values were handled by computing a simple average across the ten mathematics plausible values for each student. Although the OECD recommends the use of multiple plausible values in combination with replicate weights for population-level inference, the present study adopts an averaged plausible value approach for three reasons. First, the primary objective of the study is comparative modeling and structural pattern identification rather than national point estimation or policy benchmarking. Second, prior methodological work suggests that for regression-based and exploratory multilevel modeling, averaging plausible values yields stable parameter estimates that closely approximate results obtained using full plausible-value replication, particularly in large samples. Third, the integration of hierarchical modeling, mixture-based clustering, and multigroup SEM substantially increases computational complexity when full replication weights are applied. Accordingly, the averaged plausible value approach is adopted to facilitate model convergence and interpretability, while potential biases are acknowledged and addressed in the limitations section.

Predictor Variables

- Cognitive Predictors

Reading Proficiency (READ_MEAN) and Science Proficiency (SCIE_MEAN): Cross-subject competencies in reading and science were included as cognitive predictors of mathematical achievement. These were derived from the respective domains in the PISA 2022 assessment. As with mathematics, each domain provided ten plausible values (PV1READ–PV10READ for reading; PV1SCIE–PV10SCIE for science). The average of these plausible values was used to compute composite scores for reading (READ_MEAN) and science (SCIE_MEAN). These variables capture foundational academic skills that support mathematical reasoning and problem-solving, reflecting the theory of cross-domain transfer.

- Psychosocial Predictors

Perseverance (PERSEVAGR): Perseverance was measured through a PISA-derived index assessing students' grit and persistence in academic tasks. The index aggregates responses to items that assess students' tendency to maintain effort and interest over long periods, especially when faced with difficulties. The final index score was standardized by OECD procedures and is interpreted such that higher values indicate greater perseverance (OECD, 2024).

Curiosity (CURIOAGR): Curiosity was included as a motivational predictor, reflecting students' openness to new ideas and willingness to explore novel problems. This index aggregates responses to items related to intellectual curiosity and interest-driven exploration. Like perseverance, this construct is based on PISA student questionnaire items and standardized across the international sample.

Creativity Engagement (CREATIVITY_MEAN): Creativity-related engagement was assessed using two weighted likelihood estimates (WLEs) from the PISA 2022 student questionnaire. The first variable, *Creative Activities at School* (CREATAS), captures students' reported frequency of engaging in idea generation, design tasks, and problem-solving in formal classroom settings. The second variable, *Creative Activities Outside of School* (CREATOOS), reflects participation in similar creative endeavors pursued independently, such as artistic hobbies, maker projects, or community-based activities. Both indices were developed and scaled by the OECD, with higher values indicating greater involvement in creativity-supportive experiences. The two indicators were standardized and averaged to compute a composite score (CREATIVITY_MEAN). The internal consistency of the scale was high (Cronbach's α = .825) (Cohen et al., 2007), indicating that students' engagement in creative activities across school and informal settings is coherently aligned. In the present study, creativity is not operationalized as cognitive creative thinking ability (e.g., divergent thinking, ideational fluency, or originality), but rather as students' self-reported engagement in creativity-related activities within and outside school contexts, as captured by the PISA 2022 student questionnaire.

- Contextual Predictor

Economic, Social, and Cultural Status (ESCS): ESCS index is a composite measure developed by the OECD to capture students' socioeconomic background, widely used across PISA cycles. It combines information from three key dimensions: (1) highest parental education (measured in years of schooling based on International Standard Classification of Education; ISCED classification), (2) highest parental occupational status (coded using the International Socio-Economic Index of Occupational Status; ISEI), and (3) home possessions and resources, including access to books, educational materials, and household amenities (OECD, 2024). The resulting index is standardized across the international sample to have a mean of zero and a SD of one. Higher ESCS values indicate greater economic, social, and cultural capital, while lower values reflect structural disadvantage. This index is a robust and well-validated predictor of academic achievement and was included in the current study to account for background inequality in opportunity and access to learning resources.

3.3. Analytical procedures

Hierarchical Linear Modeling (RO1–RO3)

We used linear mixed-effects models to estimate the effects of student-level predictors while accounting for country-level random intercepts. Three models were sequentially estimated:

- Model 1 (RO1): Mathematics achievement predicted by reading proficiency. Model 1: $Y_i = \alpha_{g(i)} + \beta_1 READ_i + \epsilon_i$
- Model 2 (RO2): Model 1 + science proficiency. Model 2: $Y_i = \alpha_{g(i)} + \beta_1 READ_i + \beta_2 SCIE_i + \epsilon_i$
- Model 3 (RO3): Model 2 + creativity engagement, perseverance, curiosity, and ESCS. Model 3: $Y_i = \alpha_{g(i)} + X_i^T \beta + \epsilon_i$

where $\alpha_g \sim N(\alpha_0, \tau^2), \epsilon_i \sim N(0, \sigma^2)$. Models were compared using fixed-effect estimates, confidence intervals, and model diagnostics including intra-class correlation (ICC) and marginal/conditional R^2 (Raudenbush & Bryk, 2002). Akaike Information Criterion (AIC) was also used during preliminary testing to evaluate model fit across nested structures (Nakagawa et al., 2017).

Clustering Analysis (RO4)

To identify latent typologies of countries based on academic and psychosocial characteristics, we applied two unsupervised clustering methods:

- K-means clustering, using standardized country-level means of key predictors (reading, science, perseverance, curiosity, creativity engagement, and ESCS) (MacQueen, 1967).
- GMM using the mclust R package (Scrucca et al., 2016), with the optimal number of clusters selected using the Bayesian Information Criterion (BIC) (Schwarz, 1978).

Although both methods were compared, GMM was retained for subsequent analysis due to superior model fit and flexibility in capturing probabilistic group structure. Final cluster membership derived from GMM was used to group countries in the MG-SEM, consistent with established practices in integrating unsupervised clustering and group-based modeling frameworks (Lubke & Muthén, 2005; Pastor et al., 2007).

Multigroup Structural Equation Modeling (RO5)

We implemented MG-SEM using the lavaan package in R (Rosseel, 2012) to examine whether the strength and direction of predictors varied across GMM derived country clusters. Specifically, we:

- Estimate the structural path model (cognitive and psychosocial predictors → mathematics achievement) within each cluster.
- Compare an unconstrained model (all paths freely estimated across groups) to a constrained model (paths set equal across groups).
- Evaluated model fit using the χ^2 difference test, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) (Hu & Bentler, 1999).

Software and Weighting

All analyses were conducted in R using the following packages (Saqr & López-Pernas, 2024): lme4 (Bates et al., 2015) and performance (Lüdecke et al., 2021) for HLM, mclust for GMM (Scrucca et al., 2016), lavaan for SEM (Rosseel, 2012), and tidyverse (Wickham et al., 2019) and ggplot2 (Wickham, 2016) for data management and visualization. All algorithms (HLM estimation, clustering, SEM) can be viewed as optimization procedures: maximum likelihood for mixed models and GMM, least-squares for K-means, and covariance structure fitting for SEM. PISA's replicate weights were not applied in this version of the analysis due to the exploratory focus on model structure, but potential biases are addressed in the limitations.

4. Results

RO1: Reading Proficiency as a Predictor of Mathematics Achievement

The first HLM (Model 1) estimated the effect of reading proficiency on mathematics achievement, accounting for country-level variance. The model showed a strong, statistically significant fixed effect of reading on mathematics scores (β = 0.799, $p < .001$), indicating that a one-point increase in reading score is associated with a 0.80-point increase in mathematics score. The model explained 67.4% of the student-level variance in mathematics scores (marginal R^2 = 0.674), with an ICC of 0.22, suggesting notable between-country differences (Raudenbush & Bryk, 2002).

Model 1 (HLM RO1): MATH_MEAN ~ READ_MEAN + (1 | Country)

$$Y_i = \alpha_{g(i)} + \beta_1 READ_i + \epsilon_i, \text{with } \alpha_g \sim N(\alpha_0, \tau^2)$$

Random intercept variance (τ_{00}) = 576.4; Residual variance (σ^2) = 2048.97

RO2: Added Contribution of Science Proficiency

Adding science proficiency to the model significantly improved predictive accuracy. In Model 2, both reading ($\beta = 0.241$) and science ($\beta = 0.665$) were statistically significant predictors ($p < .001$). Model fit improved considerably over Model 1 (ΔAIC = 83,825.5; $p < .001$), and the explained variance increased (marginal $R^2 = 0.821$; conditional $R^2 = 0.858$). This confirms that science contributes unique explanatory power beyond reading.

Model 2 (HLM RO2): MATH_MEAN ~ READ_MEAN + SCIE_MEAN + (1 | Country)

$$Y_i = \alpha_{g(i)} + \beta_1 \mathrm{READ}_i + \beta_2 \mathrm{SCIE}_i + \epsilon_i$$

Random intercept variance = 299.92; Residual variance = 1146.86

RO3: Effects of Creativity Engagement, Perseverance, Curiosity, and ESCS
Model 3 incorporated psychosocial and socioeconomic predictors. The full model revealed that perseverance ($\beta = 2.27$, $p < .001$) and ESCS ($\beta = 4.54$, $p < .001$) were strong positive predictors. Interestingly, curiosity had a negative effect ($\beta = -1.54$, $p < .001$), while creativity engagement was non-significant ($\beta = 0.11$, $p = 0.278$). Model fit slightly improved (AIC reduced by 2,534.5 vs. Model 2), and marginal R^2 reached 0.825, confirming additive explanatory value.

Model 3 (HLM RO3): MATH_MEAN ~ READ + SCIE + CREATIVITY ENGAGEMENT + PERSEVERANCE + CURIOSITY + ESCS + (1 | Country)

$Y_i = \alpha_{g(i)} + X_i^T \beta + \epsilon_i$, where Xi = (READ, SCIE, CRE, PER, CUR, ESCS)^T

Random intercept variance = 292.62; Residual variance = 1126.89

Table 2 presents the HLM estimates corresponding to RO1 to RO3. It also summarizes the hierarchical linear models corresponding to RO1–RO3, demonstrating how successive blocks of predictors incrementally improve the explanation of mathematics achievement. The results show that reading and science proficiency significantly predict mathematics achievement, while perseverance and socioeconomic status add further explanatory power in the full model (RO3). Creativity engagement was not a significant predictor, and curiosity had a negative association. Model fit improved with each step, as indicated by increases in marginal and conditional R^2.

Table 2: HLM Model Comparison

Predictors	RO1: Reading			RO2: + Science			RO3: + Creativity … ESCS		
	Estimate	CI	p	Estimate	CI	p	Estimate	CI	p
(Intercept)	92.03	84.29 - 99.78	<.001	35.56	29.96 - 41.17	<.001	44.90	39.34 - 50.47	<.001
Reading	0.80	0.80 - 0.80	<.001	0.24	0.24 - 0.25	<.001	0.24	0.23 - 0.24	<.001
Science				0.67	0.66 - 0.67	<.001	0.65	0.65 - 0.66	<.001
Creativity							0.11	-0.09 - 0.31	0.278
Perseverance							2.27	2.08 - 2.46	<.001
Curiosity							-1.54	-1.74 - -1.35	<.001
ESCS							4.54	4.33 – 4.74	<.001
Random Effects									
σ^2	2048.97			1146.86			1126.89		
τ^{00}	576.42			299.92			292.62		
ICC	0.22			0.21			0.21		
N	38			38			38		
Observations	144,446			144,446			144,446		
Marginal R^2	0.674			0.821			0.825		
Conditional R^2	0.746			0.858			0.861		

Notes. Unstandardized regression coefficients are reported. CI = 95% confidence interval. σ^2 = residual variance; τ_{00} = random intercept variance across countries; ICC = intraclass correlation coefficient. Marginal R^2 indicates variance explained by fixed effects; Conditional R^2 includes both fixed and random effects. N = number of countries. **$p < .001$.

Together, these results justify extending the analysis beyond pooled regression to examine cross-national heterogeneity through clustering and multigroup modeling.

RO4: Clustering Countries by Performance and Psychosocial Profiles
To begin, it is important to emphasize that the clustering results represent analytical typologies derived from country-level aggregates rather than definitive classifications of national education systems. Cluster membership reflects probabilistic similarity in observed indicators and should not be interpreted as implying homogeneity within countries or fixed national characteristics. To identify latent typologies of educational systems, both K-means clustering and GMM were applied to country-level means of cognitive and psychosocial predictors. K-means yielded a four-cluster solution (cluster sizes: 7, 4, 11, and 16), but it is a non-probabilistic, distance-based method and does not generate likelihood-based criteria such as the BIC (Fraley & Raftery, 2002).
In contrast, GMM uses a probabilistic framework and model-based approach. Using the mclust package in R (Scrucca et al., 2016), the optimal GMM solution was determined to be a three-component EVE model (equal volume, variable shape and orientation), which demonstrated superior model fit based on BIC = −553.45. As a result, GMM was selected for interpretation and served as the basis for multigroup analysis in RO5. The three GMM clusters revealed distinct national profiles:

- Cluster 1: High-performing and balanced systems with strong academic and psychosocial indicators (e.g., Argentina, Slovakia)
- Cluster 2: High mathematics and science performance, but low perseverance (e.g., Hong Kong, Korea)
- Cluster 3: Low ESCS and curiosity, but high perseverance (e.g., Colombia, Mexico)

These groupings were subsequently used as grouping variables for MG-SEM in RO5.
In addition, to assess the correspondence between K-means and GMM clustering solutions, a cross-classification heatmap was generated (see Figure 1) to evaluate the robustness and correspondence of the two clustering approaches, a cross-classification heatmap was constructed. The heatmap displays the frequency of countries assigned to each K-means (4 clusters) and GMM (3 clusters) cluster pairing.

While some overlap exists, the mapping between the two clustering methods reveals both agreement and divergence. For instance, K-means Cluster 3 aligns most strongly with GMM Cluster 1, with 10 countries shared between them. Likewise, K-means Cluster 4 is highly concentrated in GMM Cluster 3, comprising 13 countries, suggesting stable classification under both methods. In contrast, K-means Cluster 2 shows no overlap with GMM Clusters 1 or 3, indicating that its groupings diverge from the probabilistic structure identified by GMM. Additionally, GMM Cluster 2 receives contributions from three different K-means clusters, highlighting its broader, more diverse latent profile. Overall, these patterns support the decision to adopt the three-cluster GMM solution for further modeling. Compared to K-means, GMM provides greater flexibility in capturing uncertainty and variation in country profiles, making it more suitable for use in the MG-SEM (RO5). The cross-classification matrix $M \in \mathbb{N}^{K_{km} \times K_{gmm}}$ quantifies overlaps between K-means and GMM partitions, with stability indicated by concentration along the diagonal.
Overall, the GMM procedure classified countries into three distinct clusters based on their cognitive and psychosocial profiles. Cluster 1 included Argentina, Slovakia, Serbia, the Ukrainian regions, Montenegro, Iceland, and Bulgaria. These countries demonstrated relatively balanced academic and psychosocial characteristics, with most indicators falling near the global average. Cluster 2 was characterized by countries that, on average, exhibited high mathematics and science performance alongside lower reported perseverance; this cluster included several East Asian education systems, such as Taiwan, Macao, Korea, and Hong Kong. These countries exhibited exceptionally strong mathematics and science scores, but comparatively lower levels of reported perseverance. In contrast, Cluster 3 comprised Mexico, Moldova, Peru, Jamaica, Turkey, Uruguay, Romania, Greece, Colombia, Chile, and Brazil. This group was characterized by lower SES (ESCS) and curiosity levels, but notably high perseverance. These latent typologies informed the subsequent MG-SEM (RO5), which revealed meaningful variation in predictor effects across these contrasting educational profiles.

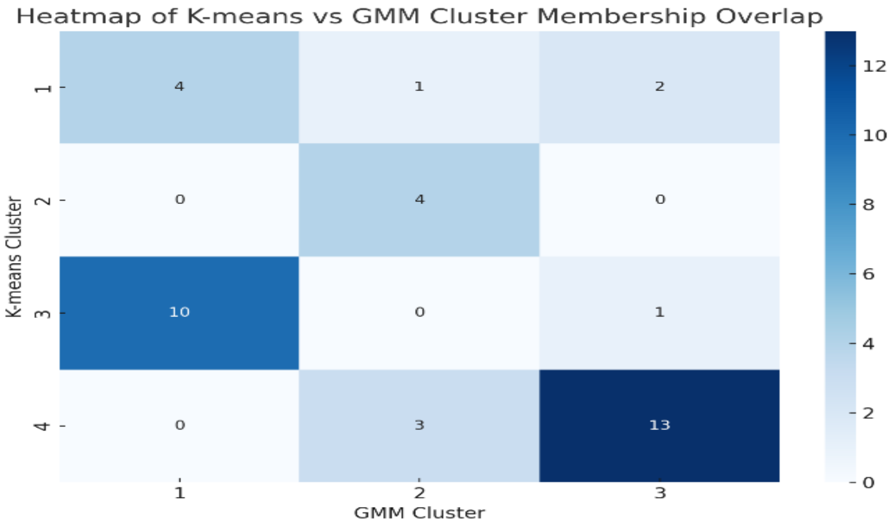

Fig. 1: Heatmap of K-Means VS GMM Cluster Membership Overlap.

Figure Notes: Cross-classification heatmap demonstrating the correspondence between the four-cluster K-means solution and the three-cluster GMM solution. Values represent the number of countries jointly assigned to each cluster pairing. Strong alignment is observed between K-means Cluster 3 and GMM Cluster 1, and between K-means Cluster 4 and GMM Cluster 3. In contrast, the dispersion of K-means Cluster 2 across GMM clusters suggests weaker structural coherence. These results highlight the superior probabilistic fit and greater classification stability of the GMM solution, which was subsequently adopted for multigroup SEM.
This comparison supports the selection of the GMM solution as the grouping structure for subsequent MG-SEM analyses (RO5).

RO5: Cross-Cluster Variation using MG-SEM
The following multigroup analyses should be interpreted as comparisons across analytically derived clusters of country-level profiles, rather than as direct comparisons between individual countries or national education systems. MG-SEM was conducted to examine whether the strength and direction of predictor effects varied across the three GMM derived country clusters. The unconstrained model, in which path coefficients were freely estimated across groups, showed perfect fit (CFI = 1.00, RMSEA = 0), while the constrained model, with equal paths across clusters, fit significantly worse ($\Delta\chi^2 = 1796.89$, $p < .001$; RMSEA = 0.052). This confirms structural heterogeneity and supports the need for cluster-specific interpretation of the predictive pathways. Key cross-cluster differences emerged:

- Perseverance had the strongest effect in Cluster 2 ($\beta = 3.26$), with smaller, but still significant effects in Clusters 1 and Cluster 3.
- Curiosity was a significant negative predictor in all clusters, although the strength of its effect varied.
- Creativity engagement showed divergent effects: significantly positive in Cluster 3 ($\beta = 0.79$), significantly negative in Cluster 2, and non-significant in Cluster 1.
- ESCS or SES was a stable and significant positive predictor across all clusters.

These findings underscore that the predictive value of cognitive and psychosocial traits is not uniform across contexts. While reading and science proficiency consistently predicted mathematics achievement across all groups, the influence of non-cognitive traits, particularly perseverance, curiosity, and creativity engagement, was contextually dependent. Such variation highlights the importance of tailoring educational strategies and interventions to the distinct psychological and socioeconomic profiles of student populations. Table 3 presents the multigroup SEM results, enabling direct comparison of predictor effects across the three GMM-derived country clusters.

Table 3: Cluster (GMM) Comparison Using MG-SEM

Predictor	Cluster 1		Cluster 2		Cluster 3	
	Estimates	β	Estimates	β	Estimates	β
Reading	0.229***	0.237	0.212***	0.226	0.237***	0.254
Science	0.670***	0.666	0.650***	0.685	0.631***	0.669
Creativity	-0.524	-0.006	-1.243***	-0.012	0.785***	0.010

Perseverance	2.367***	0.028	3.256***	0.037	0.637**	0.008
Curiosity	-3.219***	-0.037	-1.403***	-0.016	-1.792***	-0.023
ESCS	6.282***	0.069	5.149***	0.054	3.791***	0.054

Notes. Values represent unstandardized regression coefficients with standardized beta coefficients (Std. β) in adjacent columns. Significance levels are denoted as follows: ** p < .01 and *** p < .001. Non-significant estimates are unmarked.

5. Discussions and Conclusions

This study examined the predictive relationships among cognitive, psychosocial, and contextual variables and students' mathematics achievement using a robust, multi-method analytic framework. Drawing from a large-scale international dataset, we modeled student and country-level variation using HLM, identified latent typologies through clustering, and tested structural heterogeneity using MG-SEM. The findings offer insights using HLM, clustering, and MG-SEM. Formally, the study framed mathematics achievement as a mapping $X \mapsto Y$ with nested dependence and discrete partitions of countries, enabling both regression-based inference and classification of systemic typologies.

5.1. Cognitive predictors: consistent cross-domain effects

Findings from RO1 and RO2 reaffirmed the importance of cross-subject academic indicators in predicting mathematics achievement. Reading and science proficiencies were both strong and significant predictors, supporting prior work on cross-domain transfer (Korpershoek et al., 2015; Zhu, 2022). These results align with cognitive theories suggesting that shared skills such as comprehension, abstraction, and reasoning underpin performance across science, technology, engineering, and mathematics (STEM) domains (Hayes & Kraemer, 2017; Lamb et al., 2015). Notably, science exerted a larger effect than reading in the full models, suggesting it may serve as a more proximal domain for predicting mathematical problem-solving, likely due to its conceptual and methodological overlap with mathematics (Jonsson et al., 2020). In the hierarchical formulation, the coefficient of science (β_{scie}) exceeded that of reading (β_{read}), indicating stronger transferability from scientific reasoning to mathematic.

5.2. Psychosocial predictors: uneven but informative patterns

RO3 extended the model by incorporating psychosocial predictors; perseverance, curiosity, and creativity engagement, along with SES status (ESCS). As expected, perseverance demonstrated a robust positive association with mathematics achievement, supporting studies that link grit and sustained effort to academic outcomes (DiNapoli, 2023; Yu et al., 2021). ESCS also remained a significant predictor across all models, reaffirming its role as a structural factor influencing educational opportunity (Galindo & Sonnenschein, 2015). However, curiosity displayed a counterintuitive negative association with mathematics achievement. This aligns with recent critiques that curiosity, while motivational, may not always translate into academic performance, particularly in rigidly structured test environments (Peterson & Cohen, 2019). Creativity, operationalized through frequency of engagement in creative activities, was not a significant predictor in the pooled model, highlighting the challenge of capturing the influence of creativity in standardized assessments (Kaufman et al., 2023). Formally, the coefficient of creativity (β_{cre}) was statistically indistinguishable from zero in pooled models, but its variability across groups confirmed non-invariance. Its context-dependent effects were further unpacked through the multigroup analysis.

5.3. Cross-national variation: latent typologies and structural differences

RO4 identified three latent country clusters through K-means and GMM clustering, each reflecting different combinations of cognitive and psychosocial traits. These included: (a) a high-performing cluster with strong academic scores, but low perseverance (e.g., Hong Kong, Korea); (b) a resilience-oriented cluster with high perseverance despite low ESCS and curiosity (e.g., Colombia, Mexico); and (c) a balanced cluster with mid-to-high outcomes across all traits (e.g., Canada, Switzerland). These typologies echo the ecological systems perspective (He et al., 2017), suggesting that educational outcomes are shaped by individual, institutional, and cultural interactions. Note that these latent typologies are not intended to essentialize countries or imply uniform educational practices within national borders; rather, they serve as heuristic groupings that summarize broad patterns in the data. From a classification perspective, the Gaussian Mixture Model produced a partition, where each cluster reflected a distinct joint distribution of cognitive and psychosocial covariates. Put simply, students across the world achieve mathematics success through diverse pathways. In high-performing systems like Hong Kong and Korea, academic achievement coexists with low perseverance. Notably important is with Cluster 2 (high-achieving East Asian systems), which exhibited lower self-reported perseverance despite top academic performance. This may reflect cultural differences in questionnaire response styles (e.g., modesty bias) rather than actual behavioral persistence, a common observation in cross-cultural PISA analyses of dispositional indices. In contrast, countries like Colombia and Mexico demonstrate that strong perseverance can mitigate the effects of lower ESCS and curiosity. Meanwhile, balanced systems like Canada and Switzerland benefit from consistently high academic and psychosocial indicators. These distinctions suggest that policy efforts should go beyond raising test scores and instead consider the psychological and structural profiles of students. One-size-fits-all interventions may overlook the diverse ways motivation, resilience, or systemic inequality shape achievement.

5.4. Structural heterogeneity in predictive pathways

Multigroup SEM in RO5 confirmed significant structural heterogeneity across clusters. For instance, perseverance had the largest effect in Cluster 2 (the high-achieving group), while creativity showed a significant positive association with mathematics achievement only in Cluster 3 (lower-SES, high-perseverance systems). Curiosity remained a negative predictor across all groups, though effect sizes varied. These results suggest that the same trait can operate differently depending on national context, emphasizing that pooled models may obscure important structural differences. This was formally confirmed by rejecting H_0: $\gamma_1 = \gamma_2 = \gamma_3$ through $\Delta\chi^2$, establishing that regression pathways differ significantly across clusters. In summary, MG-SEM revealed that predictors such as perseverance, ESCS, curiosity, and creativity engagement did not exert uniform effects across student populations. For example, in high-performing East Asian systems, perseverance played a stronger role in shaping outcomes, while in lower-resource settings, creativity became a more salient factor. This reinforces the idea that effective education strategies must be context-sensitive and responsive to national or local student profiles.

5.5. Theoretical and methodological contributions

This study supports the relevance of cross-domain transfer theory and metacognitive engagement frameworks by showing that competencies developed in one domain can significantly influence another. However, the variability in predictor strength across countries also highlights the limitations of assuming universal transferability. Methodologically, the integration of HLM, unsupervised clustering, and MG-SEM offers a replicable, rigorous framework for analyzing nested educational data and latent systemic variation. The novelty lies in combining regression, clustering, and SEM into a discrete classification–regression pipeline, where nested dependence is captured by HLM, latent typologies by mixture models, and invariance by MG-SEM. This contributes to ongoing efforts to advance quantitative modeling in comparative education research (Toker & Green, 2021).

5.6. Limitations and future directions

Despite its contributions, the study has limitations. First, plausible values were averaged rather than fully integrated using multiple imputation and sampling weights, which may slightly reduce the precision of estimates. Second, while clustering revealed meaningful typologies, results are sensitive to input scaling and cluster number selection. Importantly, because clustering was conducted using country-level means, the resulting typologies may obscure substantial within-country variation and should be interpreted as descriptive patterns rather than definitive national classifications. Third, creativity was measured behaviorally (via activity frequency), not cognitively (e.g., ideation or divergent thinking), possibly limiting its explanatory power. Fourth, the analytic sample was restricted to 38 countries due to listwise deletion of cases with missing data on key variables, reducing the scope from the full 64 countries/economies that participated in the PISA 2022 Creative Thinking assessment. This may limit the generalizability of the findings, particularly regarding the identified country clusters and cross-national variations, as certain education systems or regional profiles are underrepresented. Future studies could incorporate multiple imputation techniques or focus on the complete dataset to capture broader global heterogeneity. Finally, the absence of school or classroom-level variables means that meso-level factors remain unmodeled. Future research should explore longitudinal or multilevel time-series models, incorporate instructional and curricular variables, and refine creativity measures to better capture its role in mathematics achievement. Moreover, clustering solutions are sensitive to initialization, and future work could explore stability via resampling or consensus clustering.

5.7. Implications for policy and practice

The findings suggest that improving mathematics achievement requires multi-layered strategies: reinforcing cross-disciplinary learning, supporting persistence, and mitigating systemic inequality. Crucially, education policy must consider the distinct profiles of student populations, not just their average scores, when designing interventions. Cluster-based and multigroup approaches offer a promising route to crafting evidence-based, contextually responsive solutions rather than generic reforms. Beyond education, the framework demonstrates how discrete mathematical models: hierarchical regression, finite mixture clustering, and multigroup invariance testing, which can be synthesized to analyze nested, heterogeneous systems.

5.8. Declaration of AI use in the preparation of the manuscript

The authors used Wordtune and ChatGPT to do language checks and improve the grammar and clarity of the manuscript. All content was carefully reviewed and edited by the authors, who take full responsibility for the final version.

References

[1] Akbasli, S., Sahin, M., & Yaykiran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice, 7*(16), 108-121.

[2] Alshabandar, R., Hussain, A., Keight, R., Laws, A., & Baker, T. (2018). The application of Gaussian Mixture Models for the identification of at-risk learners in Massive Open Online Courses. 2018 IEEE Congress on Evolutionary Computation, Rio de Janeiro, Brazil. https://doi.org/10.1109/CEC.2018.8477770.

[3] André, S., Maulana, R., Helms-Lorenz, M., Telli, S., Chun, S., Fernandez-García, C.-M., de Jager, T., Irnidayanti, Y., Inda-Caro, M., Lee, O., Safrina, R., Coetzee, T., & Jeon, M. (2020). Student perceptions in measuring teaching behavior across six countries: A multi-group confirmatory factor analysis approach to measurement invariance. *Frontiers in Psychology, 11*, 273. https://doi.org/10.3389/fpsyg.2020.00273.

[4] Awofala, A. O., Lawal, R. F., Arigbabu, A. A., & Fatade, A. O. (2022). Mathematics productive disposition as a correlate of senior secondary school students' achievement in mathematics in Nigeria. *International Journal of Mathematical Education in Science and Technology, 53*(6), 1326-1342. https://doi.org/10.1080/0020739X.2020.1815881

[5] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

[6] Bayirli, E. G., Kaygun, A., & Öz, E. (2023). An analysis of PISA 2018 mathematics assessment for Asia-Pacific countries using educational data mining. *Mathematics, 11*(6), 1318. https://doi.org/10.3390/math11061318.

[7] Belciu, A., Corbea, A. M. I., Diaconita, V., & Simonca, I. (2024). Exploring the relationship between income inequality and education: An analysis using PISA test results and the Gini coefficient. In C. Ciurea, P. Pocatilu, & F. G. Filip (Eds.), *Proceedings of 22nd International Conference on Informatics in Economy (IE 2023)* (Vol. 367, pp. 51-59). Springer. https://doi.org/10.1007/978-981-99-6529-8_5.

[8] Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Taylor & Francis Group. https://doi.org/10.4324/9780203807644.

[9] Caponera, E., Sestito, P., & Russo, P. M. (2016). The influence of reading literacy on mathematics and science achievement. *The Journal of Educational Research, 109*(2), 197-204. https://doi.org/10.1080/00220671.2014.936998.

[10] Chao, P.-C., & Ching, G. S. (2025). Uncovering the impact of COVID-19 disruptions on students mathematics achievement: A CART analysis of selected PISA 2022 data. *International Journal of Basic and Applied Sciences, 14*(6), 115-122. https://doi.org/10.14419/y9ynz887.

[11] Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge. https://doi.org/10.4324/9780203029053

[12] Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education: A systematic literature review. *Thinking Skills and Creativity, 8*, 80-91. https://doi.org/10.1016/j.tsc.2012.07.004

[13] de Vink, I. C., Willemsen, R. H., Lazonder, A. W., & Kroesbergen, E. H. (2022). Creativity in mathematics performance: The role of divergent and convergent thinking. *British Journal of Educational Psychology, 92*(2), 484-501. https://doi.org/10.1111/bjep.12459.

[14] DiNapoli, J. (2023). Distinguishing between grit, persistence, and perseverance for learning mathematics with understanding. *Education Sciences, 13*(4), 402. https://doi.org/10.3390/educsci13040402.

[15] Doumas, L. A. A., Puebla, G., Martin, A. E., & Hummel, J. E. (2022). A theory of relation learning and cross-domain generalization. *Psychological Review, 129*(5), 999–1041. https://doi.org/10.1037/rev0000346

[16] Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611-631. https://doi.org/10.1198/016214502760047131

[17] Gajda, A., Karwowski, M., & Beghetto, R. A. (2017). Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology, 109*(2), 269-299. https://doi.org/10.1037/edu0000133

[18] Galindo, C., & Sonnenschein, S. (2015). Decreasing the SES math achievement gap: Initial math proficiency and home learning environments. *Contemporary Educational Psychology, 43*, 25-38. https://doi.org/10.1016/j.cedpsych.2015.08.003.

[19] Gomez, A. L., Pecina, E. D., Villanueva, S. A., & Huber, T. (2020). The undeniable relationship between reading comprehension and mathematics performance. *Issues in Educational Research, 30*(4), 1329-1354.

[20] Hadar, L. L., & Tirosh, M. (2019). Creative thinking in mathematics curriculum: An analytic framework. *Thinking Skills and Creativity, 33*, 100585. https://doi.org/10.1016/j.tsc.2019.100585

[21] Hansen, M., & Gonzalez, T. (2014). Investigating the relationship between STEM learning principles and student achievement in math and science. *American Journal of Education, 120*(2), 139-171. https://doi.org/10.1086/674376

[22] Hayes, J. C., & Kraemer, D. J. M. (2017). Grounded understanding of abstract concepts: The case of STEM learning. *Cognitive Research: Principles and Implications, 2*, 7. https://doi.org/10.1186/s41235-016-0046-z

[23] He, J., Van de Vijver, F. J. R., & Kulikova, A. (2017). Country-level correlates of educational achievement: Evidence from large-scale surveys. *Educational Research and Evaluation, 23*(5-6), 163–179. https://doi.org/10.1080/13803611.2017.1455288.

[24] 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

[25] Jonsson, B., Granberg, C., & Lithner, J. (2020). Gaining mathematical understanding: The effects of creative mathematical reasoning and cognitive proficiency. *Frontiers in Psychology, 11*, 574366. https://doi.org/10.3389/fpsyg.2020.574366.

[26] Kamid, Huda, N., Syafmen, W., Sufri, & Sofnidar. (2021). The relationship between students' mathematical disposition and their learning outcomes. *Journal of Education and Learning, 15*(3), 376-382. https://doi.org/10.11591/edulearn.v15i3.17604.

[27] Kaufman, J. C., Kapoor, H., Patston, T., & Cropley, D. H. (2023). Explaining standardized educational test scores: The role of creativity above and beyond GPA and personality. *Psychology of Aesthetics, Creativity, and the Arts, 17*(6), 725-734. https://doi.org/10.1037/aca0000433

[28] Khine, M. S., Fraser, B. J., & Afari, E. (2020). Structural relationships between learning environments and students' non-cognitive outcomes: Secondary analysis of PISA data. *Learning Environments Research, 23*, 395–412. https://doi.org/10.1007/s10984-020-09313-2.

[29] Korpershoek, H., Kuyper, H., & van der Werf, G. (2015). The relation between students' math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education. *International Journal of Science and Mathematics Education, 13*, 1013–1037. https://doi.org/10.1007/s10763-014-9534-0

[30] Korres, K., & Tsami, E. (2010). Supporting the development of critical thinking skills in secondary education through the use of interdisciplinary statistics' and mathematics' problems. *Journal of Interdisciplinary Mathematics, 13*(5), 491-507. https://doi.org/10.1080/09720502.2010.10700716.

[31] Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

[32] Lamb, R., Akmal, T., & Petrie, K. (2015). Development of a cognition-priming model describing learning in a STEM classroom. *Journal of Research in Science Teaching, 52*(3), 410-437. https://doi.org/10.1002/tea.21200

[33] Lee, J. L., & Stankov, L. (2023). Multilevel latent profile analysis of global competence in PISA 2018: Small country-level differences. *Journal of Cross-Cultural Psychology, 54*(6-7), 658-683. https://doi.org/10.1177/00220221231183150.

[34] Lezhnina, O., & Kismihók, G. (2022). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. *International Journal of Research & Method in Education, 45*(2), 180-199. https://doi.org/10.1080/1743727X.2021.1963226.

[35] Li, S., & Lajoie, S. P. (2022). Cognitive engagement in self-regulated learning: An integrative model. *European Journal of Psychology of Education, 37*, 833–852. https://doi.org/10.1007/s10212-021-00565-x

[36] Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM Mathematics Education, 49*, 937–949. https://doi.org/10.1007/s11858-017-0867-3.

[37] Liu, F., Yang, D., Liu, Y., Zhang, Q., Chen, S., Li, W., Ren, J., Tian, X., & Wang, X. (2022). Use of latent profile analysis and k-means clustering to identify student anxiety profiles. *BMC Psychiatry, 22*, 12. https://doi.org/10.1186/s12888-021-03648-7.

[38] Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21-39. https://doi.org/10.1037/1082-989X.10.1.21.

[39] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.

[40] Muñez, D., Bull, R., & Lee, K. (2021). Socioeconomic status, home mathematics environment and math achievement in kindergarten: A mediation analysis. *Developmental Science, 24*(6), e13135. https://doi.org/10.1111/desc.13135.

[41] Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 2017021. https://doi.org/10.1098/rsif.2017.0213.

[42] Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences, 11*(1), 237. https://doi.org/10.3390/app11010237.

[43] Niu, W., Cheng, L., Duan, D., & Zhang, Q. (2022). Impact of perceived supportive learning environment on mathematical achievement: The mediating roles of autonomous self-regulation and creative thinking. *Frontiers in Psychology, 12*, 781594. https://doi.org/10.3389/fpsyg.2021.781594.

[44] OECD. (2023a). *PISA 2022 database* OECD. https://www.oecd.org/en/data/datasets/pisa-2022-database.html.

[45] OECD. (2023b). *PISA 2022 results (Volume I): The state of learning and equity in education*. PISA, OECD Publishing. https://doi.org/10.1787/53f23881-en.

[46] OECD. (2024). *PISA 2022 technical report*. PISA, OECD Publishing. https://doi.org/10.1787/01820d6d-en

[47] Pasigon, C. P. (2024). Mathematical proficiency, scientific reasoning, metacognitive skills, and performance of learners in physics: A mathematical model. *International Journal of Learning, Teaching and Educational Research, 23*(4), 252-278. https://doi.org/10.26803/ijlter.23.4.14

[48] Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*(1), 8-47. https://doi.org/10.1016/j.cedpsych.2006.10.003.

[49] Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin, 146*(7), 595–634. https://doi.org/10.1037/bul0000231

[50] Peterson, E. G., & Cohen, J. (2019). A case for domain-specific curiosity in mathematics. *Educational Psychology Review, 31*, 807–832. https://doi.org/10.1007/s10648-019-09501-4

[51] Rajak, A., Shrivastava, A. K., & Vidushi. (2020). Applying and comparing machine learning classification algorithms for predicting the results of students. *Journal of Discrete Mathematical Sciences and Cryptography, 23*(2), 419-427. https://doi.org/10.1080/09720529.2020.1728895.

[52] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

[53] Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

[54] Saqr, M., & López-Pernas, S. (Eds.). (2024). *Learning analytics methods and tutorials: A practical guide using R*. Springer. https://doi.org/10.1007/978-3-031-54464-4

[55] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464. https://doi.org/10.1214/aos/1176344136.

[56] Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021

[57] Sideridis, G. D., TsaousisIoannis, I., & Al-Harbi, K. (2021). Identifying student subgroups as a function of school level attributes: A multilevel latent class analysis. *Frontiers in Psychology, 12*, 624221. https://doi.org/10.3389/fpsyg.2021.624221.

[58] Singh, H., & Kathuria, A. (2023). Heterogeneity in passenger satisfaction of bus rapid transit system among age and gender groups: A PLS-SEM Multi-group analysis. *Transport Policy, 141*, 27-41. https://doi.org/10.1016/j.tranpol.2023.07.009.

[59] Spielberger, C. D., & Starr, L. M. (1994). Curiosity and exploratory behavior. In H. F. O'Neil & M. Drillings (Eds.), *Motivation: Theory and research* (pp. 221-243). Routledge.

[60] Suherman, S., & Vidákovich, T. (2022). Assessment of mathematical creative thinking: A systematic review. *Thinking Skills and Creativity, 44*, 101019. https://doi.org/10.1016/j.tsc.2022.101019.

[61] Sun, L., Bradley, K. D., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Educatio, 34*(14), 2107-2125. https://doi.org/10.1080/09500693.2012.708063.

[62] Tang, J., Gibson, A., & Bruza, P. (2025). Analysis of exploratory behaviour: A step towards modelling of curiosity. LAK '25: Proceedings of the 15th International Learning Analytics and Knowledge Conference, Dublin, Ireland. https://doi.org/10.1145/3706468.3706561

[63] Toker, T., & Green, K. (2021). A comparison of latent class analysis and the mixture Rasch model using 8th grade mathematics data in the fourth international mathematics and science study. *International Journal of Assessment Tools in Education, 8*(4), 959-974. https://doi.org/10.21449/ijate.1024251.

[64] Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28*(4), 409-426. https://doi.org/10.1080/01443410701708228

[65] Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://doi.org/10.1007/978-3-319-24277-4_9.

[66] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686.

[67] Wu, X., Wu, R., Chang, H.-H., Kong, Q., & Zhang, Y. (2020). International comparative study on PISA mathematics achievement test based on cognitive diagnostic models. *Frontiers in Psychology, 11*, 2230. https://doi.org/10.3389/fpsyg.2020.02230

[68] You, H. S., Park, S., & Delgado, C. (2021). A closer look at US schools: What characteristics are associated with scientific literacy? A multivariate multilevel analysis using PISA 2015. *Science Education, 105*(2), 406-437. https://doi.org/10.1002/sce.21609

[69] Yu, Y., Hua, L., Feng, X., Wang, Y., Yu, Z., Zi, T., Zhao, Y., & Li, J. (2021). True grit in learning math: The math anxiety-achievement link Is mediated by math-specific grit. *Frontiers in Psychology, 12*, 645793. https://doi.org/10.3389/fpsyg.2021.645793.

[70] Zhu, Y. (2022). Reading matters more than mathematics in science learning: an analysis of the relationship between student achievement in reading, mathematics, and science. *International Journal of Science Education, 44*(1), 1-17. https://doi.org/10.1080/09500693.2021.2007552.