

Questioning “Boy Math” and “Girl Math”: A Measurement-Based Test of Gendered Financial Behaviour Profiles Among Indonesian Students

Asma Khoirunnisa ^{a,1,*}, Heri Retnawati ^{a,b,2}, Risky Setiawan ^{a,3}, Samsul Hadi ^{c,4},
Hanifah Mar’atush Shalihah ^{a,5}, Zhansaya K. Markhmadova ^{d,6}, Alaa Abdelkader ^{e,7},
Aisha Juma Mussa ^{f,8}

^a Educational Research and Evaluation, Graduate School, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

^b Department of Educational Mathematics, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

^c Department of Electrical Engineering, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

^d Department of Pedagogy and Psychology, Abai Kazakh National Pedagogical University, Almaty, 050010, Kazakhstan

^e Department of Art Education, Helwan University, Cairo, 4034572, Egypt

^f Department of Monitoring and Evaluation, The Open University of Tanzania, Dar es Salaam, 23409, Tanzania

¹ asmakhoirunnisa.2025@student.uny.ac.id; ² heri_retnawati@uny.ac.id; ³ riskysetiawan@uny.ac.id;

⁴ samsul_hd@uny.ac.id; ⁵ hanifahmaratush.2025@student.uny.ac.id; ⁶ jan_27_2001@mail.ru;

⁷ alaaeltlwany@gmail.com; ⁸ ayshamussa22@gmail.com

* Corresponding Author

ARTICLE INFO

Article history

Received June 02, 2026

Revised July 14, 2026

Accepted August 18, 2026

Keywords

Financial Literacy;

Gender Differences;

Measurement Invariance;

K-Means Clustering;

Explainable Machine Learning

ABSTRACT

The popular phrases “Boy Math” and “Girl Math” have transformed humorous spending explanations into a wider discussion about gender and money. This study uses the discourse as a cultural entry point rather than as a construct directly measured by PISA 2018, which predates the trend and does not assess meme-specific post-hoc spending rationalisations. Using complete-case data from 5,833 Indonesian students, the study examined financial autonomy, careful price behaviour, bank confidence, digital confidence, financial behaviour, and financial attitude through ordinal confirmatory factor analysis, gender measurement invariance, latent factor scores, k-means clustering, and a secondary Random Forest diagnostic. The measurement model showed acceptable fit (CFI = 0.992, TLI = 0.991, RMSEA = 0.063, SRMR = 0.059), and invariance was supported up to the strict level. The two-cluster solution separated students consistently across all six dimensions, indicating higher and lower financial engagement levels rather than qualitatively distinct reasoning styles. The unadjusted association between gender and cluster membership was not significant, $\chi^2(1) = 2.135$, $p = 0.144$, Cramer’s $V = 0.019$. In adjusted logistic regression using conventional standard errors, male students had slightly lower odds of belonging to the higher-engagement group after grade was controlled (OR = 0.864, $p = 0.006$). School-clustered standard errors were not reported in the supplied output, so the adjusted inferential result remains provisional. Given the gender-balanced profiles, negligible unadjusted effect size, and odds ratio close to one, the relationship is substantively small. Random Forest reconstructed the internally generated cluster labels with 97.27% out-of-bag accuracy; this represents classification fidelity rather than external predictive validity. The findings do not support treating “Boy Math” and “Girl Math” as clear empirical gender categories.

©2026 The Author. This is an open-access article under the CC-BY license.



1. Introduction

Financial literacy is an important competence for young people because financial decisions now begin before adulthood and often happen in digital settings. Students are already exposed to saving, spending, online transactions, mobile payments, family money rules, peer consumption, digital banking, and commercial persuasion before they enter work. International assessment reports also show that financial literacy remains a policy priority because 15-year-old students need to understand money, compare options, and manage financial risks in daily life [1][2][3][4][5]. In Indonesia, this issue is also important because the 2024 National Survey on Financial Literacy and Inclusion reported a financial literacy index of 65.43 percent and a financial inclusion index of 75.02 percent [6]. These figures show progress, but they also show that access to financial products does not always mean students have strong financial understanding or responsible financial behaviour. For adolescents, financial learning comes from many sources, including school, family, peers, social media, and digital platforms.

The PISA financial literacy framework is useful for studying adolescent financial behaviour because it does not treat financial literacy only as arithmetic or knowledge of financial terms. It includes knowledge and understanding, financial processes, money and transactions, planning and managing finances, risk and reward, and the financial landscape [1][2][7][8]. The questionnaire also measures financial independence, banking confidence, digital financial confidence, consumer behaviour, financial practices, and attitudes toward money [6]. This is important because students may understand financial concepts but still have low confidence, weak monitoring habits, or impulsive spending behaviour. Recent studies also show that financial literacy research has moved from knowledge-only models to broader models that include knowledge, confidence, behaviour, attitude, inclusion, and well-being [9][10][11][12][13][14][15][16][17][18]. Therefore, adolescent financial literacy should not only be studied by asking whether boys or girls score higher. It should also examine how different financial behaviour dimensions form meaningful profiles.

Contemporary online discourse has increasingly framed spending decisions through gendered narratives, including labels such as “Boy Math” and “Girl Math,” which humorously describe how individuals justify spending, saving, shopping, or perceiving certain costs as less financially burdensome. However, these labels do not constitute established scientific constructs and should not be treated as literal behavioural categories, as doing so may produce weak conclusions and reinforce gender stereotypes. Moreover, the PISA 2018 Financial Literacy Questionnaire predates the emergence of these narratives and was not designed to measure such context-specific financial rationalisations. Therefore, the present study uses these terms only as a cultural framing device to examine whether gender is associated with broader latent financial behaviour profiles. This approach is consistent with the wider literature, which shows that gender differences in financial literacy are complex and may appear in confidence, self-assessment, investment attitudes, and non-response patterns, while international school-based assessments often report small or inconsistent differences in average financial literacy performance [9][16][17][18][19][20][21][22][23][24]. More recent studies involving high-school students, Generation Z, and digital financial literacy further indicate that the relationship between gender and financial behaviour may interact with financial education, parental discussion, digital access, socio-economic background, and individual confidence [9][21][22][23][24][25][26][27][28][29][30]. Thus, culturally constructed gender narratives may shape public perceptions of financial behaviour without necessarily representing stable psychological or behavioural types.

This creates the main research problem. If gender is compared only using raw mean scores, the issue of “Boy Math” and “Girl Math” can appear too simple. In fact, financial behaviour is latent, multidimensional, and measured using ordinal and binary questionnaire items. Mean scores are easy to calculate, but they assume that all items have equal weight and that response categories have equal distances. For non-cognitive instruments, this assumption can be too strong. Psychometric studies recommend modelling ordinal indicators more carefully when researchers want to interpret latent traits rather than only summarize item responses [31][32][33][34][35][36][37][38][39]. Therefore,

this study treats selected questionnaire blocks as indicators of latent financial behaviour constructs and uses latent factor scores as the main input for machine learning profiling. This approach keeps the value of the questionnaire while reducing the risk of over-interpreting raw averages.

The proposed solution is to separate profile formation from gender interpretation. First, financial behaviour profiles are identified using latent factor scores estimated through ordinal confirmatory factor analysis. Second, measurement invariance is tested to check whether the latent structure works similarly for female and male students. Third, clustering is conducted without gender as an input variable, so the profiles are based on behavioural and attitudinal patterns rather than sex labels. Fourth, gender is examined after the profiles are created using profile distribution, chi-square association, effect size, and regression analysis. Fifth, Random Forest is used as an explainable machine learning method to identify which latent dimensions most strongly reproduce profile membership. This design follows recent recommendations in educational data science that machine learning should reveal patterns while still maintaining measurement validity and interpretability [40][41][42][43][44][45][46][47][48].

The current literature shows several related research streams. The first stream studies financial literacy among students using PISA, national surveys, or school-based questionnaires [3][7][8][27][29]. The second stream examines gender gaps, confidence gaps, stereotype threat, and differences between knowledge and self-perception [16][17][18][20][21]. The third stream focuses on digital financial literacy and behavioural finance in relation to mobile banking, online payments, and financial technology [10][12]. The fourth stream applies psychometric models such as confirmatory factor analysis, measurement invariance, item response theory, and cognitive diagnostic modelling in educational assessment [31][32][33]. The fifth stream uses machine learning, clustering, Random Forest, SHAP, and educational analytics to study patterns in student data [49][50][51][52].

The novelty of this study is that it connects those research streams through a gender-sensitive and measurement-based design. Rather than claiming to measure the “Boy Math” and “Girl Math” phenomenon directly, this study evaluates whether the broader financial behaviour dimensions associated with money confidence, autonomy, attitudes, and consumer practices form identifiable profiles and whether those profiles differ by gender. This distinction positions the study as an empirical examination of financial behaviour patterns inspired by a cultural discourse, rather than a validation of the discourse itself.

2. Method

2.1. Research Design

This study used a quantitative secondary data design based on the PISA 2018 Financial Literacy Questionnaire. The unit of analysis was the student, and the analytic focus was the Indonesian sample. The design was non-experimental because the study did not manipulate instructional conditions or assign students to treatments. Instead, it used existing large-scale assessment data to examine latent financial behaviour dimensions and to identify machine learning-based student profiles. This design is appropriate because the purpose of the study was not to estimate a causal treatment effect, but to produce a measurement-informed profile of students' financial behaviour and to examine whether gender was associated with that profile membership after the profiles had been created.

The methodological design was built around a key principle: gender should not be used to create the profiles. This principle was adopted to avoid circular interpretation. If gender were used as an input variable during clustering, any later statement about gender differences across clusters would be partly built into the cluster solution itself. Therefore, the clustering stage used only latent financial behaviour scores. Gender was introduced afterward as an external grouping variable. This allows the study to ask whether female and male students are distributed differently across profiles without defining the profiles as male or female categories.

2.2. Data & Variables

The original PISA financial literacy questionnaire dataset contained 107,174 students and 1,058 variables. The analysis selected Indonesian students using the country code IDN. The selected Indonesian dataset contained 7,133 students and 37 variables before complete-case filtering. The variables included student identifiers, school identifiers, stratum, grade, gender, and selected financial literacy questionnaire items. After removing cases with missing values on the variables used in the modelling pipeline, the final complete-case analytic sample contained 5,833 students, consisting of 3,032 female students and 2,801 male students. Because the present article focuses on an analytic modelling pipeline, official PISA student weights and replicate weights were not applied. The results should therefore be interpreted as analytic evidence from the selected complete cases, not as official population estimates for Indonesia.

The selected questionnaire blocks were FL159, FL160, FL162, FL163, FL168, and FL169 (*see Table 1*). These blocks were chosen because they are directly related to financial independence, consumer behaviour, bank confidence, digital financial confidence, actual financial behaviour, and financial attitude. FL159 items were used to represent financial autonomy. FL160 items were used to represent careful price behaviour, with one item also indicating impulsive spending. FL162 measured confidence in bank-related activities, such as understanding bank statements and managing bank accounts. FL163 measured confidence in digital financial activities. FL168 contained binary behavioural indicators, such as checking change, using digital payments, and monitoring money. FL169 measured attitudes toward money matters. This selection allowed the study to focus on non-cognitive financial behaviour profiles rather than cognitive financial literacy test performance.

Table 1. Selected Constructs and PISA Financial Questionnaire Blocks

Construct	Code	Substantive meaning	Response format
Financial autonomy	FL159Q01HA- FL159Q04HA	Independence and permission in money decisions	Ordinal
Careful price behaviour	FL160Q01HA- FL160Q04HA	Comparison, waiting, and cautious consumption	Ordinal
Bank confidence	FL162Q01HA- FL162Q06HA	Confidence in bank-related tasks	Ordinal
Digital confidence	FL163Q01HA- FL163Q05HA	Confidence in digital financial tasks	Ordinal
Financial behaviour	FL168Q01HA- FL168Q08HA	Observed financial practices and monitoring	Binary
Financial attitude	FL169Q01HA- FL169Q04HA	Perceived relevance and value of money matters	Ordinal

2.3. Coding & Measurement Model

Several items were reverse-coded so that higher values reflected a theoretically consistent direction. FL159Q02HA and FL159Q03HA were reversed because needing permission for spending reflects lower autonomy. FL160Q03HA was reversed when it was used as part of careful price behaviour because buying without comparing prices reflects less careful behaviour. FL169Q03HA was reversed because the statement that money matters are not relevant now reflects a lower financial attitude orientation. The FL168 yes-or-no items were recoded as binary variables, where one represented yes and zero represented no. This coding step was necessary because the PISA questionnaire includes mixed response formats, and the direction of the indicators must be clear before measurement modelling.

Composite mean scores were computed only for preliminary descriptive analysis. The main analysis used latent factor scores estimated from an ordinal CFA model. This distinction is important because the selected PISA questionnaire items represent non-cognitive latent constructs, not directly observed continuous variables. The ordinal CFA model used WLSMV (Weighted Least Squares Mean and Variance Adjusted) estimation, ordered categorical indicators, theta parameterisation, and

standardised latent variances. The six latent constructs were financial autonomy, careful price behaviour, bank confidence, digital confidence, financial behaviour, and financial attitude. The use of WLSMV was chosen because it is appropriate for ordered categorical indicators and has been recommended in applied factor analysis when Likert-type and binary indicators are analysed as manifestations of latent variables [34][35].

The descriptive table (Table 2) is therefore useful as an entry point, but it is insufficient as the basis for the main conclusion. If the analysis stopped at mean scores, the article might conclude that male students were more confident and more impulsive while female students were more careful and slightly more attitude-oriented. Such a conclusion would be incomplete because it would ignore the latent nature of the constructs and the mixed response formats of the indicators. The more appropriate question is whether these dimensions combine into broader financial behaviour profiles and whether gender is meaningfully associated with those profiles after measurement issues have been addressed.

Table 2. Composite Mean Scores by Gender

Gender	n	Autonomy	Careful price	Impulsive spending	Bank confidence	Digital confidence	Financial behaviour	Financial attitude
Female	3,032	2.517	2.764	1.496	2.343	2.339	0.738	2.854
Male	2,801	2.559	2.712	1.577	2.467	2.447	0.760	2.826

2.4. Machine Learning Profiling & Gender Analysis

After the ordinal CFA model was estimated, latent factor scores were extracted using the CFA model and used as the input for machine learning profiling. The six latent score variables were standardised before clustering. The number of clusters was examined by comparing alternative solutions rather than being determined solely by theoretical expectations. Candidate solutions ($k = 2, 3$, and 4) were evaluated using cluster separation and interpretability criteria, including silhouette-based evaluation and substantive examination of cluster patterns. The two-cluster solution was retained because additional clusters did not produce sufficiently distinct behavioural configurations and mainly separated students along the same general financial engagement continuum. Therefore, the identified profiles are interpreted as levels of overall financial engagement rather than completely different behavioural types. This interpretation is important because all six latent dimensions showed a consistent direction, indicating that the clustering primarily captured differences in intensity across financial behaviour dimensions. The final labels were assigned after examining cluster characteristics rather than before the clustering process.

Gender analysis was conducted in three steps. First, the proportion of female and male students was calculated within each profile. Second, a chi-square test and Cramer's V were used to examine whether gender was associated with profile membership. Third, multinomial logistic regression was used to estimate profile membership from gender and grade level. Because PISA data are collected through a complex sampling design involving students nested within schools, regression estimates were interpreted cautiously. The available analytical output used conventional standard errors; therefore, the reported p-values may underestimate uncertainty if school-level clustering is considered. The adjusted gender effect is consequently treated as a robustness indicator rather than definitive evidence of a population-level association.

3. Results and Discussion

3.1. Descriptive Statistics

The complete-case sample contained 5,833 students, consisting of 3,032 female students and 2,801 male students. Female students represented 52.0 percent of the analytic sample, while male students represented 48.0 percent. This composition indicates that the sample was relatively balanced by gender, which allows an initial descriptive comparison between female and male students. Before conducting the main measurement analysis, this study first examined preliminary composite mean

scores across seven financial dimensions. The results are shown in Fig. 1. The scores represent raw composite means, not latent scores. Complete-case filtering retained 5,833 of 7,133 students (81.8%) and excluded 1,300 (18.2%). The supplied output did not report gender- or grade-specific retention tests; therefore, differential retention by gender and grade could not be evaluated.

Fig. 1 shows that gender differences in the preliminary financial measures were generally small and varied across constructs. Male students recorded slightly higher mean scores for financial autonomy with 2.559 compared with 2.517 among female students. Male students also showed higher scores in the descriptive single-item impulsive-spending indicator with 1.577 compared with 1.496, bank confidence with 2.467 compared with 2.343, digital confidence with 2.447 compared with 2.339, and financial behaviour with 0.760 compared with 0.738. In contrast, female students recorded slightly higher mean scores for careful price behaviour with 2.764 compared with 2.712 among male students. Female students also showed a slightly higher score in financial attitude with 2.854 compared with 2.826. The impulsive-spending value should not be interpreted as a separate latent dimension because it is the unreversed descriptive form of FL160Q03HA, which contributes in reverse direction to the careful price behaviour construct. All differences remain descriptive and should not be interpreted as evidence of statistical or causal differences.

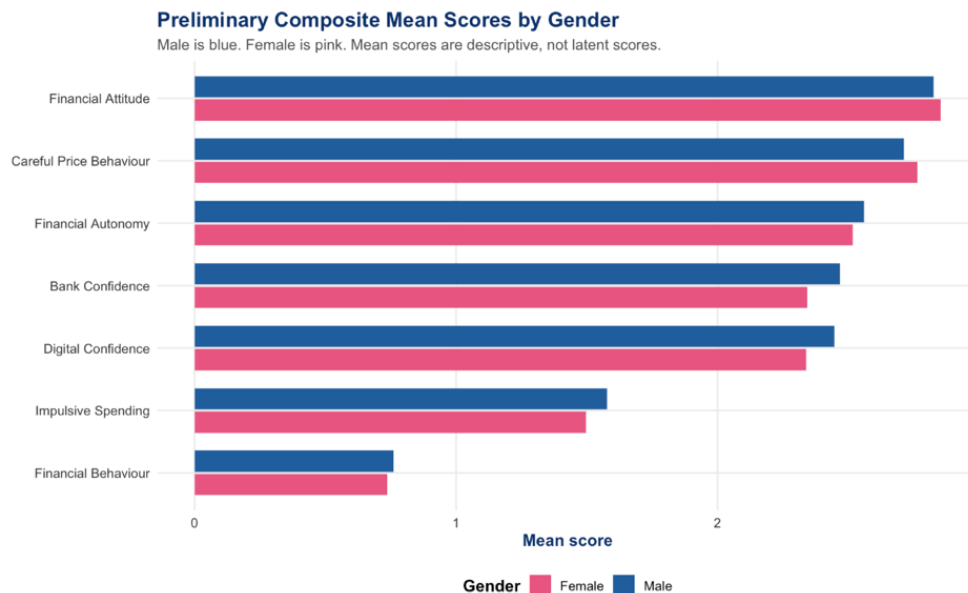


Fig. 1. Composite mean score by gender

3.2. Reliability, Factorability, & Ordinal CFA

The preliminary reliability results showed that bank confidence and digital confidence had very high internal consistency (see Table 3), with Cronbach's alpha values of 0.922 and 0.918, respectively. Other preliminary scales showed substantially weaker alpha values. Careful price behaviour had an alpha of 0.363, financial behaviour had an alpha of 0.549, and financial attitude had an alpha of 0.337. Most importantly, financial autonomy produced a negative alpha of -1.361. A negative alpha indicates negative average inter-item covariance and is a serious warning about coding direction, content coherence, or the suitability of a single-factor representation. Although alpha is not sufficient as a stand-alone indicator of scale quality, this extreme result cannot be dismissed merely because the indicators are ordinal or heterogeneous. The autonomy factor score should be interpreted only if the standardised loadings are directionally coherent and sufficiently strong and if a model-based reliability coefficient, such as McDonald's omega, is acceptable. The same requirement applies, though less severely, to careful price behaviour and financial attitude. These findings are consistent with the broader psychometric argument that alpha is not a sufficient indicator of scale quality, especially when constructs are short, multidimensional, or composed of behaviourally diverse indicators [38][39][53][54].

Table 3. Reliability Analysis of Preliminary Scales

Scale	Items	Raw alpha	Standardized alpha
Financial autonomy	4	-1.361	-1.375
Careful price behaviour	4	0.363	0.359
Bank confidence	6	0.922	0.922
Digital confidence	5	0.918	0.918
Financial behaviour	4	0.549	0.559
Financial attitude	4	0.337	0.347

The factorability diagnostics supported further latent modelling. The Kaiser-Meyer-Olkin value was 0.91, which indicates that the selected items were suitable for factor analysis. Bartlett's test of sphericity was significant, $\chi^2 = 78,410.36$, $p < 0.001$, indicating that the correlation matrix was not an identity matrix. Parallel analysis suggested up to eight factors. However, the final CFA model used six theoretically interpretable dimensions because the research objective was not to maximise the number of extracted factors, but to model meaningful financial behaviour constructs that could be used for profiling. This decision reflects a balance between empirical dimensionality and construct interpretability.

Based on Table 4, the six-factor ordinal CFA demonstrated acceptable overall model fit, with CFI = 0.992, TLI = 0.991, RMSEA = 0.063, and SRMR = 0.059. Although the CFI and TLI indicated strong incremental fit, these indices should not be interpreted independently of the absolute fit indices or item-level measurement evidence, particularly under WLSMV estimation with theta parameterisation. The RMSEA and SRMR indicated adequate rather than exceptional fit. Consequently, the standardised loadings, category thresholds, construct reliability, and latent-factor correlations were examined to determine whether the six constructs provided sufficiently coherent measurement scores for the subsequent clustering analysis.

Table 4. Ordinal CFA Model Fit

Fit index	Value
CFI	0.992
TLI	0.991
RMSEA	0.063
SRMR	0.059

Table 5 presents the standardised ordinal-CFA loadings and estimated category thresholds. The results showed consistently strong and positive loadings for Bank Confidence, Digital Confidence, and Financial Behaviour. However, several reverse-coded items produced negative loadings. Specifically, FL160Q03HA_R loaded negatively on Careful Price Behaviour at -0.436 , FL169Q03HA_R loaded negatively on Financial Attitude at -0.298 , and FL159Q02HA_R and FL159Q03HA_R loaded negatively on Financial Autonomy at -0.821 and -0.609 , respectively. These negative coefficients suggest that the coding direction, item wording, or factor specification of the affected items may not be aligned with the intended construct orientation.

Table 5. Standardised Ordinal-CFA Loadings and Estimated Thresholds

Construct	Item	Standardised loading	SE	p	Number of thresholds	Minimum threshold	Maximum threshold	Estimated thresholds
Bank Confidence	FL162Q01HA	0,58680556	0.004	<0.001	3	-1.881	2.790	$\tau_1 = -1.881$; $\tau_2 = 0.208$; $\tau_3 = 2.790$
Bank Confidence	FL162Q02HA	0,62083333	0.003	<0.001	3	-2.496	3.319	$\tau_1 = -2.496$; $\tau_2 = -0.169$; $\tau_3 = 3.319$
Bank Confidence	FL162Q03HA	0,63541667	0.002	<0.001	3	-2.694	3.910	$\tau_1 = -2.694$; $\tau_2 = 0.142$; $\tau_3 = 3.910$
Bank Confidence	FL162Q04HA	0,59444444	0.004	<0.001	3	-1.858	3.183	$\tau_1 = -1.858$; $\tau_2 = 0.535$; $\tau_3 = 3.183$

Construct	Item	Standardised loading	SE	p	Number of thresholds	Minimum threshold	Maximum threshold	Estimated thresholds
Bank Confidence	FL162Q05HA	0,62361111	0.003	<0.001	3	-2.311	3.631	$\tau_1 = -2.311$; $\tau_2 = 0.385$; $\tau_3 = 3.631$
Bank Confidence	FL162Q06HA	0,54513889	0.005	<0.001	3	-1.919	1.978	$\tau_1 = -1.919$; $\tau_2 = -0.354$; $\tau_3 = 1.978$
Careful Price Behaviour	FL160Q01HA	0,59652778	0.007	<0.001	3	-2.959	1.045	$\tau_1 = -2.959$; $\tau_2 = -1.046$; $\tau_3 = 1.045$
Careful Price Behaviour	FL160Q02HA	0,63402778	0.007	<0.001	3	-3.433	1.366	$\tau_1 = -3.433$; $\tau_2 = -1.311$; $\tau_3 = 1.366$
Careful Price Behaviour	FL160Q03HA_R	-0.436	0.013	<0.001	3	-1.301	1.153	$\tau_1 = -1.301$; $\tau_2 = 0.004$; $\tau_3 = 1.153$
Careful Price Behaviour	FL160Q04HA	0,45069444	0.010	<0.001	3	-1.397	1.328	$\tau_1 = -1.397$; $\tau_2 = -0.232$; $\tau_3 = 1.328$
Digital Confidence	FL163Q01HA	0,60416667	0.003	<0.001	3	-2.170	2.805	$\tau_1 = -2.170$; $\tau_2 = -0.070$; $\tau_3 = 2.805$
Digital Confidence	FL163Q02HA	0,63125	0.003	<0.001	3	-2.493	3.816	$\tau_1 = -2.493$; $\tau_2 = 0.451$; $\tau_3 = 3.816$
Digital Confidence	FL163Q03HA	0,61527778	0.003	<0.001	3	-2.173	3.263	$\tau_1 = -2.173$; $\tau_2 = 0.247$; $\tau_3 = 3.263$
Digital Confidence	FL163Q04HA	0,59791667	0.004	<0.001	3	-1.928	2.958	$\tau_1 = -1.928$; $\tau_2 = 0.337$; $\tau_3 = 2.958$
Digital Confidence	FL163Q05HA	0,58958333	0.004	<0.001	3	-2.058	2.497	$\tau_1 = -2.058$; $\tau_2 = -0.061$; $\tau_3 = 2.497$
Financial Attitude	FL169Q01HA	0,49236111	0.011	<0.001	3	-2.442	1.751	$\tau_1 = -2.442$; $\tau_2 = -0.852$; $\tau_3 = 1.751$
Financial Attitude	FL169Q02HA	0,55763889	0.010	<0.001	3	-3.179	1.591	$\tau_1 = -3.179$; $\tau_2 = -1.627$; $\tau_3 = 1.591$
Financial Attitude	FL169Q03HA_R	-0.298	0.015	<0.001	3	-1.611	1.722	$\tau_1 = -1.611$; $\tau_2 = 0.183$; $\tau_3 = 1.722$
Financial Attitude	FL169Q04HA	0,46875	0.011	<0.001	3	-2.706	0,48194444	$\tau_1 = -2.706$; $\tau_2 = -1.693$; $\tau_3 = 0.694$
Financial Autonomy	FL159Q01HA	0,55555556	0.007	<0.001	3	-2.683	1.396	$\tau_1 = -2.683$; $\tau_2 = -1.825$; $\tau_3 = 1.396$
Financial Autonomy	FL159Q02HA_R	-0.821	0.006	<0.001	3	-1.242	3.210	$\tau_1 = -1.242$; $\tau_2 = 1.924$; $\tau_3 = 3.210$
Financial Autonomy	FL159Q03HA_R	-0.609	0.009	<0.001	3	-1.175	2.215	$\tau_1 = -1.175$; $\tau_2 = 1.069$; $\tau_3 = 2.215$
Financial Autonomy	FL159Q04HA	0,56875	0.007	<0.001	3	-3.199	1.042	$\tau_1 = -3.199$; $\tau_2 = -2.285$; $\tau_3 = 1.042$
Financial Behaviour	FL168Q01HA_bin	0,44166667	0.034	<0.001	1	-1.836	-1.836	$\tau_1 = -1.836$
Financial Behaviour	FL168Q04HA_bin	0,575	0.021	<0.001	1	-0.838	-0.838	$\tau_1 = -0.838$
Financial Behaviour	FL168Q06HA_bin	0,43611111	0.021	<0.001	1	-0.027	-0.027	$\tau_1 = -0.027$
Financial Behaviour	FL168Q08HA_bin	0,48263889	0.026	<0.001	1	-1.662	-1.662	$\tau_1 = -1.662$

Note. *_R* identifies items intended to be reverse-coded, whereas *_bin* identifies binary-coded items. The symbol τ represents an estimated ordinal-category threshold. Negative loadings among reverse-coded items indicate that item coding, factor orientation, or construct specification should be re-examined.

The construct-level measurement results are summarised in Table 6. Bank Confidence demonstrated strong loadings ranging from 0.785 to 0.915 and excellent model-based reliability, with McDonald's $\omega = 0.947$. Digital Confidence similarly showed consistently strong loadings ranging

from 0.849 to 0.909 and an omega coefficient of 0.942. Financial Behaviour also demonstrated acceptable measurement quality, with loadings ranging from 0.628 to 0.828 and $\omega = 0.792$.

In contrast, Careful Price Behaviour showed a loading range from -0.436 to 0.913 and a marginal omega coefficient of 0.684 . Financial Attitude showed a loading range from -0.298 to 0.803 and relatively low reliability, with $\omega = 0.608$. The most serious measurement weakness was observed for Financial Autonomy. Its loadings ranged from -0.821 to 0.819 , while its omega coefficient was only 0.021 . This near-zero reliability estimate indicates that the four items did not form a coherent unidimensional score under the current coding and factor specification.

Table 6. Construct-Level Measurement Summary

Construct	Number of items	Minimum loading	Maximum loading	Minimum threshold	Maximum threshold	McDonald's ω	Measurement assessment
Bank Confidence	6	0,54513889	0,63541667	-2.694	3.910	0,65763889	All loadings were positive, and model-based reliability was satisfactory.
Careful Price Behaviour	4	-0.436	0,63402778	-3.433	1.366	0,475	One item loaded negatively, and reliability was marginal. Item direction or factor specification should be rechecked.
Digital Confidence	5	0,58958333	0,63125	-2.493	3.816	0,65416667	All loadings were positive, and model-based reliability was satisfactory.
Financial Attitude	4	-0.298	0,55763889	-3.179	1.751	0,42222222	One item loaded negatively, and reliability was below the conventional 0.70 criterion.
Financial Autonomy	4	-0.821	0,56875	-3.199	3.210	0.021	Two reverse-coded items loaded negatively, and model-based reliability was unacceptable.
Financial Behaviour	4	0,43611111	0,575	-1.836	-0.027	0,55	All loadings were positive, and model-based reliability was satisfactory.

Note. McDonald's omega values of approximately 0.70 or higher are commonly regarded as acceptable for research purposes. However, reliability should also be evaluated alongside the number of items, loading pattern, construct breadth, and intended use of the resulting scores.

Table 7 presents the standardised correlations among the six latent factors. Most correlations were positive and low to moderate, ranging from 0.172 to 0.528. Financial Autonomy was most strongly associated with Financial Attitude, with a correlation of 0.528, while its weakest association was with Financial Behaviour, with a correlation of 0.172. Careful Price Behaviour showed moderate correlations with Financial Attitude, Financial Behaviour, Bank Confidence, and Digital Confidence.

The largest latent-factor correlation was observed between Bank Confidence and Digital Confidence, with $r = 0.817$. This result indicates substantial shared variance between the two constructs and may suggest limited discriminant validity. Although the correlation remained below unity, its magnitude indicates that confidence in conventional banking activities and confidence in digital financial activities may represent closely related dimensions in this sample.

Taken together, the findings indicate that acceptable global CFA fit did not correspond to uniformly adequate construct-level measurement. Bank Confidence, Digital Confidence, and Financial Behaviour produced coherent loading patterns and satisfactory reliability. By contrast, Careful Price Behaviour and Financial Attitude contained negatively loading reverse-coded items and produced marginal or weak reliability estimates. Financial Autonomy demonstrated particularly poor

internal coherence and should not be interpreted as a reliable latent score under the current specification.

Table 7. Standardised Latent-Factor Correlation

Factor 1	Factor 2	Correlation	SE	p
Financial Autonomy	Careful Price Behaviour	0,27708333	0.014	<0.001
	Bank Confidence	0,19027778	0.014	<0.001
	Digital Confidence	0,15208333	0.014	<0.001
	Financial Behaviour	0,11944444	0.020	<0.001
	Financial Attitude	0,36666667	0.014	<0.001
Careful Price Behaviour	Bank Confidence	0,21180556	0.013	<0.001
	Digital Confidence	0,21666667	0.013	<0.001
	Financial Behaviour	0,24652778	0.018	<0.001
	Financial Attitude	0,28263889	0.015	<0.001
Bank Confidence	Digital Confidence	0,56736111	0.005	<0.001
	Financial Behaviour	0,20416667	0.018	<0.001
	Financial Attitude	0,22708333	0.015	<0.001
Digital Confidence	Financial Behaviour	0,24027778	0.018	<0.001
	Financial Attitude	0,22361111	0.014	<0.001
Financial Behaviour	Financial Attitude	0,18680556	0.021	<0.001

Note. All reported correlations were statistically significant at $p < 0.001$. The comparatively high correlation between Bank Confidence and Digital Confidence suggests substantial conceptual overlap and should be considered when evaluating discriminant validity.

Accordingly, the reverse coding of FL160Q03HA_R, FL169Q03HA_R, FL159Q02HA_R, and FL159Q03HA_R should be verified against the original PISA response categories and questionnaire wording. If the negative loadings remain after coding verification, the affected constructs should be respecified by correcting their factor orientation, removing incompatible items, or excluding the construct when adequate model-based reliability cannot be established. The factor scores and clustering results should therefore be regarded as provisional until the affected measurement models have been re-estimated and the clustering analysis has been recomputed using verified construct scores.

3.3. Measurement Invariance by Gender

Measurement invariance testing was conducted to examine whether the latent financial behaviour constructs could be compared across female and male students. Table 8 describes the configural, metric, scalar or threshold, and strict models all showed stable fit indices. The changes in CFI were 0.000 across the successive models, and the changes in RMSEA were very small. This result supports the interpretation that the measurement structure was broadly comparable across gender groups. In other words, gender comparisons and gender-profile analyses are less likely to be confounded by major differences in the way the latent constructs functioned across female and male students.

Table 8. Measurement Invariance by Gender

Model	CFI	TLI	RMSEA	SRMR	Delta CFI	Delta RMSEA
Configural	0.992	0.991	0.062	0.059	-	-
Metric	0.992	0.991	0.063	0.060	0.000	0.001
Scalar/Threshold	0.992	0.991	0.061	0.060	0.000	-0.002
Strict	0.991	0.991	0.061	0.060	0.000	0.000

This finding is important for the conceptual aim of the article. If the measures were not comparable across gender, then any discussion of gendered profiles would be questionable. The invariance evidence does not mean that female and male students have identical financial behaviour. It means that the measurement model provides a reasonably stable basis for comparing latent scores and examining profile membership. This strengthens the argument that the study can discuss gender representation in profiles without treating gender as the construct itself.

3.4. Two Financial Behaviour Profiles

K-means clustering of latent factor scores produced two financial behaviour profiles. From Table 9 and Fig. 2, profile 1 contained 2,821 students, or 48.36 percent of the complete-case sample. Profile 2 contained 3,012 students, or 51.64 percent. The two clusters were not labelled by gender. They were labelled by their latent financial behaviour characteristics. Profile 1 showed higher mean latent factor scores across all six dimensions: financial autonomy, careful price behaviour, bank confidence, digital confidence, financial behaviour, and financial attitude. Profile 2 showed lower mean latent factor scores across the same dimensions. Therefore, Profile 1 was labelled the Higher Financial Engagement Profile and Profile 2 was labelled the Lower Financial Engagement Profile.

Table 9. Latent Profile Characteristics

Profile	n	Autonomy	Careful price	Bank confidence	Digital confidence	Financial behaviour	Financial attitude	Suggested name
1	2,821	0.431	0.485	0.584	0.596	0.347	0.477	Higher Financial Engagement Profile
2	3,012	-0.414	-0.487	-0.561	-0.573	-0.433	-0.468	Lower Financial Engagement Profile

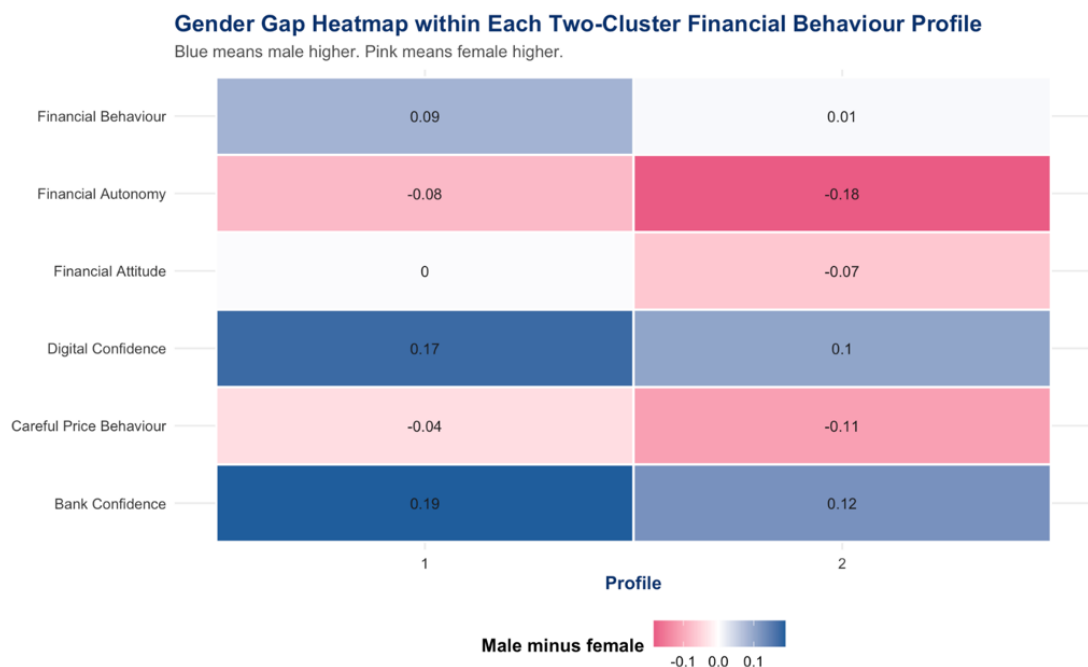


Fig. 2. Heatmap of latent financial behaviour characteristics by profile

Table 6 shows that the largest between-group differences occurred in digital confidence and bank confidence, followed by careful price behaviour and financial attitude. However, the common direction of all six centroids is more important than their rank order: the solution does not identify a high-confidence/low-care configuration or another clearly differentiated behavioural pattern. It separates students who are generally higher across the measured dimensions from those who are generally lower. Fig. 2, which displays male-minus-female differences within each engagement group, likewise shows relatively modest within-profile gender gaps. Male students were somewhat higher in bank and digital confidence, while female students were somewhat higher in autonomy or careful price behaviour in parts of the solution. These small within-profile contrasts do not create distinct male and female reasoning types.

3.5. Gender Distribution Across Profiles

Gender was examined after clustering. The distribution of gender across the two profiles was relatively balanced (see Fig. 3). From Table 10, female students in profile 1 represented 50.97 percent and male students in profile 1 represented 49.03 percent. In Profile 2, female students represented 52.92 percent and male students represented 47.08 percent. This pattern indicates that both profiles contained substantial numbers of female and male students. Although female students were slightly more represented in both profiles, the differences were small. The chi-square test confirmed that gender and profile membership were not meaningfully associated, $\chi^2(1) = 2.1346$, $p = 0.144$, Cramer's $V = 0.0191$ (see Table 11).

Table 10. Gender Distribution by Financial Behaviour Profile

Profile	Gender	n	Percent within profile
1	Female	1,438	50.97
	Male	1,383	49.03
2	Female	1,594	52.92
	Male	1,418	47.08

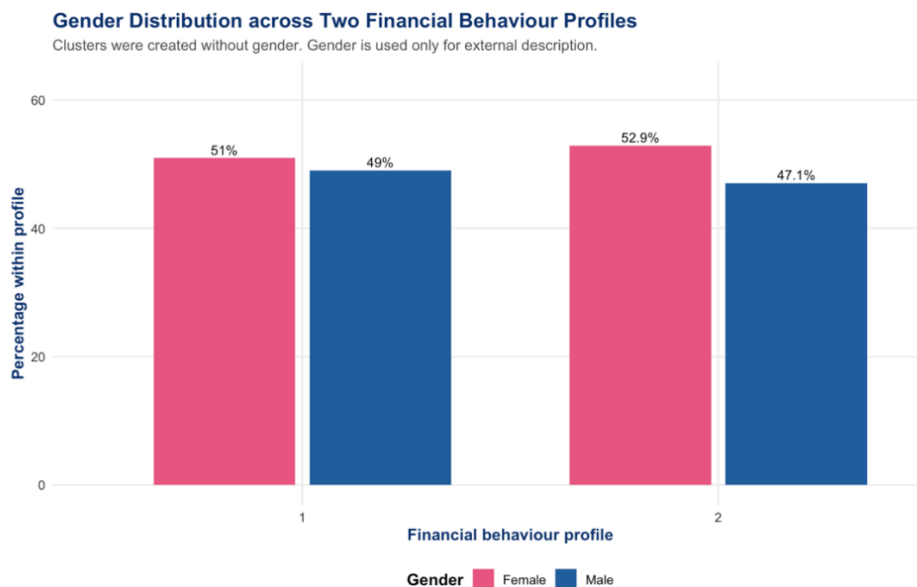


Fig. 3. Gender distribution across the two financial behaviour profiles

This result is central to the interpretation of “Boy Math” and “Girl Math”. If those terms represented strong empirical gender categories, one would expect a clear gender concentration in one profile or the other. That did not occur. The two profiles were primarily distinguished by latent financial behaviour dimensions rather than by gender composition. The finding does not deny that gender may shape financial socialisation, confidence, or opportunity. Instead, it shows that in this sample, the measured profiles were not reducible to female and male categories. Therefore, the safer interpretation is that gendered cultural narratives may point toward real questions about confidence, behaviour, and attitude, but they should not be treated as evidence that boys and girls have fixed financial reasoning styles.

Table 11. Gender and Profile Association Test

Statistic	df	p value	Cramer's V
2.1346	1	0.1440	0.0191

The regression result added a more nuanced finding. Male students had lower odds of belonging to Profile 1 than female students after grade level was controlled, $OR = 0.864$, $p = 0.006$. Grade level also predicted profile membership, $OR = 0.667$, $p < 0.001$. This adjusted result is not logically inconsistent with the non-significant chi-square test because the two analyses address different estimands: the chi-square test assesses the unadjusted marginal association, whereas the regression estimates the gender association conditional on grade. Adjustment can reveal a small conditional relationship that is not apparent in the unadjusted cross-tabulation. Nevertheless, the odds ratio is close to one, Cramer’s V is negligible, both profiles remain gender-balanced, and the reported regression used conventional rather than school-clustered standard errors. The adjusted gender coefficient is therefore statistically detectable but substantively small and inferentially provisional; it does not justify describing either engagement level as a female or male profile.

3.6. Internal Reclassification Diagnostic

Random Forest was applied to evaluate how well the six latent dimensions reproduced the two-profile solution. The model produced an out-of-bag error rate of 2.73 percent, corresponding to an out-of-bag accuracy of 97.27 percent. This high accuracy means that the two clusters were highly separable by the latent financial behaviour dimensions. The result should be interpreted as profile classification fidelity, not as external predictive validity. The same latent variables that generated the clusters were also used to classify the clusters. Therefore, the Random Forest model confirms that the cluster solution is internally coherent and interpretable, but it does not prove that the same model would predict independent real-world financial outcomes with equal accuracy.

The Random Forest variable-importance results are presented in Table 12. Financial Attitude showed the highest overall importance, with a Mean Decrease Accuracy of 102.283 and a Mean Decrease Gini of 710.175. Careful Price Behaviour ranked second according to Mean Decrease Accuracy at 97.369, whereas Digital Confidence ranked second according to Mean Decrease Gini at 591.620. These differences indicate that the relative ranking of the dimensions depended on the importance metric used.

The profile-specific importance values also showed different patterns across the two clusters. For Profile 1, Financial Attitude produced the highest importance value at 88.458, followed by Financial Behaviour at 84.462 and Financial Autonomy at 75.888. For Profile 2, Careful Price Behaviour ranked highest at 72.618, followed by Financial Attitude at 70.554 and Digital Confidence at 55.952. Bank Confidence had the lowest profile-specific importance value for Profile 2 and the lowest overall Mean Decrease Accuracy.

Table 12. Random Forest Variable Importance for Internal Profile Reclassification

Variable	Profile 1 importance	Profile 2 importance	Mean Decrease Accuracy	Mean Decrease Gini
Financial Attitude	88.458	70.554	102.283	710.175
Digital Confidence	56.817	55.952	75.228	591.620
Careful Price Behaviour	72.628	72.618	97.369	525.397
Financial Behaviour	84.462	52.700	75.302	419.385
Bank Confidence	58.267	48.503	69.652	413.473
Financial Autonomy	75.888	55.134	85.359	252.759

Note. Profile 1 and Profile 2 importance values represent class-specific contributions to the correct classification of each profile. Mean Decrease Accuracy indicates the reduction in classification accuracy after the values of a variable are permuted. Mean Decrease Gini indicates the accumulated reduction in node impurity attributable to a variable across the decision trees. Higher values indicate a stronger contribution to internal profile classification.

Financial Attitude made the strongest overall contribution to reconstructing the profile labels. However, the class-specific results showed that the dimensions did not contribute equally to both profiles. Financial Attitude and Financial Behaviour were especially influential in reconstructing Profile 1, whereas Careful Price Behaviour and Financial Attitude were the most influential dimensions for Profile 2. Digital Confidence had a comparatively high Mean Decrease Gini despite its lower class-specific and Mean Decrease Accuracy values, suggesting that it frequently contributed

to impurity reduction across the trees without producing the same degree of deterioration in predictive accuracy when permuted.

Financial Autonomy produced a relatively high Mean Decrease Accuracy of 85.359 but the lowest Mean Decrease Gini of 252.759. This divergence indicates that the variable affected classification accuracy when disrupted but was used less extensively or produced smaller impurity reductions across the tree-building process. Nevertheless, its importance should be interpreted with particular caution because the ordinal CFA results indicated severe measurement problems for the Financial Autonomy construct.

These importance values should not be interpreted as evidence that the dimensions independently predict real-world financial outcomes. The same six latent dimensions were used both to create the cluster labels and to reconstruct them. Consequently, the results primarily describe how the variables contributed to the internal geometric separation generated by the clustering procedure. The Random Forest analysis is therefore retained only as an internal reclassification and reproducibility diagnostic, while the cluster centroids provide the more direct description of the differences between the two profiles.

3.7. Discussion

The findings support a cautious reinterpretation of “Boy Math” and “Girl Math”. These terms may attract public attention, but they have limited scientific value when treated as fixed gender categories. The empirical results did not reveal distinctly male or female financial types. Instead, the analysis identified differences in levels of financial engagement. One profile showed stronger digital confidence, bank confidence, careful price behaviour, financial attitude, financial autonomy, and financial behaviour. The other profile showed weaker performance across these dimensions. Since both profiles included substantial numbers of female and male students, the results challenge gender-essentialist interpretations and support a needs-based approach to financial education. This interpretation aligns with prior studies showing that financial literacy gaps should not be reduced to simple gender stereotypes, because confidence, response patterns, and socialisation can shape observed differences [14][15][16][21].

The results also reinforce a broader point in financial literacy research. Financial literacy is not only a matter of knowledge. It also involves attitudes, confidence, decision-making, and behaviour [4][7][8]. This explains why a single gender label cannot capture students’ financial learning needs. The observed profiles are more useful because they identify combinations of strengths and weaknesses across several financial dimensions. This is consistent with OECD’s view that financial literacy should support informed decisions, responsible behaviour, and participation in increasingly digital financial environments [2][3][5].

Methodologically, this study shows the importance of placing measurement analysis before machine learning classification. Random Forest can classify students with high accuracy, but accuracy alone cannot establish construct validity [25][30][55]. For this reason, the study first used ordinal CFA and measurement invariance testing to strengthen the interpretation of the latent scores. This step matters because questionnaire items are ordinal and may not contribute equally to each construct [40][34][41]. Measurement invariance testing also helped assess whether constructs had comparable meaning across groups before the profiles were interpreted [31][32][46]. After this stage, clustering and Random Forest analysis served as tools for pattern detection and interpretation, not as substitutes for measurement validity.

The Indonesian PISA context gives these findings practical relevance. Access to financial products and digital services does not automatically produce confidence, careful behaviour, or sound financial attitudes. Students may use digital tools without fully understanding how to compare prices, monitor spending, interpret bank information, or evaluate financial risks. Prior research also shows that financial education can improve financial knowledge and downstream behaviour, especially when learning is structured, relevant, and connected to real decisions [10][11][12][13]. Therefore, students in the lower engagement profile may need practical learning activities on price comparison, budgeting,

digital transaction safety, bank statement reading, and reflective money attitudes. Students in the higher engagement profile may need more advanced tasks on financial risk, entrepreneurship, saving strategies, and responsible digital finance.

The findings suggest that gender-sensitive financial education should not rely on gender-segregated assumptions. A gender-sensitive approach should recognise that students may face different expectations, opportunities, confidence levels, and financial socialisation. However, it should not assume that female students naturally think in one financial style and male students in another. The weak association between gender and profile membership supports interventions based on measured needs rather than stereotypes. In practice, teachers and policymakers should focus on confidence-building, behavioural monitoring, and digital financial literacy for all students, while still checking whether specific groups face barriers in access, encouragement, or self-efficacy.

This study has several limitations. First, the analysis used complete cases and did not apply official PISA sampling weights or replicate weights. Future research should use the full PISA complex survey design to produce population-level estimates. Second, the study relied on secondary questionnaire data, so it could not include variables such as social media exposure, parental allowance, school financial education content, or actual digital payment use. Third, the profiles came from cross-sectional data, so they should not be interpreted as developmental pathways. Fourth, Random Forest accuracy should be read as cluster classification fidelity, because the model predicted labels created from the same latent dimensions. Future studies should test whether these profiles predict external outcomes, such as cognitive financial literacy performance, saving behaviour, digital payment practices, or financial well-being.

4. Conclusion

This study examined whether the popular discourse of “Boy Math” and “Girl Math” could be reframed through measurable financial behaviour patterns among Indonesian students. Using PISA 2018 Financial Literacy Questionnaire data, the study estimated latent financial behaviour dimensions through ordinal CFA, tested measurement invariance by gender, clustered students using latent factor scores, examined gender after clustering, and used Random Forest as a secondary internal reclassification diagnostic. The final two-group solution identified a Higher Financial Engagement Profile and a Lower Financial Engagement Profile. Because all six centroids differed in the same direction, these groups are better understood as engagement levels rather than distinct financial reasoning styles. Their interpretation also remains conditional on fuller reporting of item loadings, thresholds, factor correlations, and model-based reliability, especially for the financial autonomy construct.

The results show that gender should not be used to define the engagement groups. Female and male students were distributed relatively evenly across both groups, and the unadjusted gender-profile association was weak and non-significant. The adjusted regression indicated a statistically significant but small gender coefficient after grade was controlled. These findings are not contradictory: the chi-square test estimates the marginal association, whereas the regression estimates the association conditional on grade. Given the odds ratio close to one, negligible Cramer’s V, balanced group composition, and the absence of school-clustered standard errors, the adjusted result should be treated as substantively minor and provisional. The main conclusion is therefore that “Boy Math” and “Girl Math” are cultural labels that can motivate research questions, not empirical categories of financial reasoning. Financial literacy interventions should be designed around measured needs and engagement levels rather than gender stereotypes. Future studies should strengthen measurement reporting, apply survey-design corrections, evaluate alternative cluster solutions, and test the groups against independent financial literacy or socioeconomic outcomes.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] J. Mo, "PISA 2018 results: Are students smart about money?," *PISA in Focus*, vol. 106, May 2020, <https://doi.org/10.1787/d540d9e9-en>.
- [2] OECD, "PISA 2018 Results (Volume IV): Are Students Smart about Money?," *PISA*, May 2020, <https://doi.org/10.1787/48ebd1ba-en>.
- [3] OECD, "PISA 2022 Results (Volume IV): How Financially Smart Are Students?," *PISA*, Jun. 2024, <https://doi.org/10.1787/5a849c2a-en>.
- [4] OECD, "OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion 2022," *OECD*, Mar. 2022, <https://doi.org/10.1787/cbc4114f-en>.
- [5] OECD, "OECD Legal Instruments: Recommendation of the Council on Financial Literacy," *OECD*, 2020, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0461>.
- [6] OECD, "Financial Literacy Questionnaire For Pisa 2018 (International Option) Main Survey Version," *OECD*, 2017,
- [7] A. Lusardi and O. S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *J. Econ. Lit.*, vol. 52, no. 1, pp. 5–44, 2014, <https://doi.org/10.1257/jel.52.1.5>.
- [8] A. Lusardi, "Financial literacy and the need for financial education: evidence and implications," *Swiss J. Econ. Stat.*, vol. 155, no. 1, 2019, <https://doi.org/10.1186/s41937-019-0027-5>.
- [9] J. M. Cordero and F. Pedraja, "The effect of financial education training on the financial literacy of Spanish students in PISA," *Appl. Econ.*, vol. 51, no. 16, pp. 1679–1693, Apr. 2019, <https://doi.org/10.1080/00036846.2018.1528336>.
- [10] T. Kaiser, A. Lusardi, L. Menkhoff, and C. Urban, "Financial education affects financial knowledge and downstream behaviors," *J. financ. econ.*, vol. 145, no. 2, pp. 255–272, Aug. 2022, <https://doi.org/10.1016/j.jfineco.2021.09.022>.
- [11] T. Kaiser and L. Menkhoff, "Financial education in schools: A meta-analysis of experimental studies," *Econ. Educ. Rev.*, vol. 78, p. 101930, Oct. 2020, <https://doi.org/10.1016/j.econedurev.2019.101930>.
- [12] T. Kaiser and L. Menkhoff, "Does financial education impact financial literacy and financial behavior, and if so, when?," *World Bank Economic Review*, vol. 31, no. 3, pp. 611–630, Oct. 2017, <https://doi.org/10.1093/wber/lhx018>.
- [13] D. Fernandes, J. G. Lynch, and R. G. Netemeyer, "Financial literacy, financial education, and downstream financial behaviors," *Manage. Sci.*, vol. 60, no. 8, pp. 1861–1883, 2014, <https://doi.org/10.1287/mnsc.2013.1849>.
- [14] L. Bottazzi and A. Lusardi, "Stereotypes in financial literacy: Evidence from PISA," *Journal of Corporate Finance*, vol. 71, p. 101831, Dec. 2021, <https://doi.org/10.1016/j.jcorpfin.2020.101831>.
- [15] T. West, L. de Zwaan, and D. Johnson, "Do women have lower levels of financial literacy, or are they opting out? A look at the non-response gender bias in financial literacy measurement," *Financial Services Review*, vol. 31, no. 1, pp. 55–71, Sep. 2023, <https://doi.org/10.61190/fsr.v31i1.3195>.
- [16] T. Bucher-Koenen, R. Alessie, A. Lusardi, and M. van Rooij, "Fearless Woman: Financial Literacy, Confidence, and Stock Market Participation," *Manage. Sci.*, vol. 71, no. 9, pp. 7414–7430, Dec. 2024, <https://doi.org/10.1287/mnsc.2023.00425>.
- [17] K. Goyal and S. Kumar, "Financial literacy: A systematic review and bibliometric analysis," *Int. J. Consum. Stud.*, vol. 45, no. 1, pp. 80–105, Jan. 2021, <https://doi.org/10.1111/ijcs.12605>.

-
- [18] S. M. M. Prado, K. Chiluiza, P. Everaert, and M. Valcke, "Design and Evaluation among Young Adults of a Financial Literacy Scale Focused on Key Financial Decisions," *Educ. Sci. (Basel)*, vol. 12, no. 7, p. 460, Jul. 2022, <https://doi.org/10.3390/educsci12070460>.
- [19] M. Setiawan, N. Effendi, T. Santoso, V. I. Dewi, and M. S. Sapulette, "Digital financial literacy, current behavior of saving and spending and its future foresight," *Economics of Innovation and New Technology*, vol. 31, no. 4, pp. 320–338, 2022, <https://doi.org/10.1080/10438599.2020.1799142>.
- [20] K. M. Vieira, T. K. Matheis, E. dos R. Lehnhart, and F. O. Tavares, "Digital Financial Knowledge Scale (DFKS): Insights from a Developing Economy," *International Journal of Financial Studies*, vol. 12, no. 4, p. 120, Dec. 2024, <https://doi.org/10.3390/ijfs12040120>.
- [21] A. Hasler and A. Lusardi, "The Gender Gap in Financial Literacy: A Global Perspective," *Global Financial Literacy Excellence Center, The George Washington University School of Business*, pp. 2-16, 2017, <https://gflec.org/wp-content/uploads/2017/05/The-Gender-Gap-in-Financial-Literacy-A-Global-Perspective-Report.pdf>.
- [22] C. Strömbäck, T. Lind, K. Skagerlund, D. Västfjäll, and G. Tinghög, "Does self-control predict financial behavior and financial well-being?," *J. Behav. Exp. Finance*, vol. 14, pp. 30–38, Jun. 2017, <https://doi.org/10.1016/j.jbef.2017.04.002>.
- [23] D. Moreno-Herrero, M. Salas-Velasco, and J. Sánchez-Campillo, "Factors that influence the level of financial literacy among young people: The role of parental engagement and students' experiences with money matters," *Child. Youth Serv. Rev.*, vol. 95, pp. 334–351, Dec. 2018, <https://doi.org/10.1016/j.childyouth.2018.10.042>.
- [24] G. Silinskas, A. K. Ahonen, and T. A. Wilska, "Financial literacy among Finnish adolescents in PISA 2018: the role of financial learning and dispositional factors," *Large. Scale. Assess. Educ.*, vol. 9, no. 1, Nov. 2021, <https://doi.org/10.1186/s40536-021-00118-0>.
- [25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, <https://doi.org/10.1023/A:1010933404324>.
- [26] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Appl. Stat.*, vol. 28, no. 1, p. 100, 1979, <https://doi.org/10.2307/2346830>.
- [27] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [28] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data Set Via the Gap Statistic," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 63, no. 2, pp. 411–423, Jul. 2001, <https://doi.org/10.1111/1467-9868.00293>.
- [29] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in neural information processing systems*, vol. 30, 2017, https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.
- [30] C. Molnar, *Interpretable Machine Learning*. Leanpub, 2022. <https://christophm.github.io/interpretable-ml-book/>.
- [31] F. F. Chen, "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance," *Struct. Equ. Modeling*, vol. 14, no. 3, pp. 464–504, 2007, <https://doi.org/10.1080/10705510701301834>.
- [32] W. Meredith, "Measurement invariance, factor analysis and factorial invariance," *Psychometrika*, vol. 58, no. 4, pp. 525–543, Dec. 1993, <https://doi.org/10.1007/BF02294825>.
- [33] P. M. Bentler, "Comparative fit indexes in structural models," *Psychol. Bull.*, vol. 107, no. 2, pp. 238–246, 1990, <https://doi.org/10.1037//0033-2909.107.2.238>.
- [34] D. B. Flora and P. J. Curran, "An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data," *Psychol. Methods*, vol. 9, no. 4, p. 466, Dec. 2004, <https://doi.org/10.1037/1082-989X.9.4.466>.
- [35] M. Rhemtulla, P. É. Brosseau-Liard, and V. Savalei, "When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under
-

- suboptimal conditions,” *Psychol. Methods*, vol. 17, no. 3, pp. 354–373, Sep. 2012, <https://doi.org/10.1037/a0029315>.
- [36] C. H. Li, “Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares,” *Behav. Res. Methods*, vol. 48, no. 3, pp. 936–949, Sep. 2016, <https://doi.org/10.3758/s13428-015-0619-7>.
- [37] B. Muthén, “A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators,” *Psychometrika*, vol. 49, no. 1, pp. 115–132, Mar. 1984, <https://doi.org/10.1007/BF02294210>.
- [38] K. Sijtsma, “On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha,” *Psychometrika*, vol. 74, no. 1, p. 107, Mar. 2008, <https://doi.org/10.1007/s11336-008-9101-0>.
- [39] R. P. McDonald, “Test theory: A unified treatment,” *Test Theory: A Unified Treatment*, pp. 1–485, Jun. 2013, <https://doi.org/10.4324/9781410601087>.
- [40] W. Revelle, “(PDF) psych: Procedures for Psychological, Psychometric, and Personality Research. R Package Version 1.0–95,” *R Statistical Package*, 2013, https://www.researchgate.net/publication/281345624_psych_Procedures_for_Psychological_Psychometric_and_Personality_Research_R_Package_Version_1_0-95.
- [41] R. B. Kline, *Principles and Practice of Structural Equation Modeling*. The Guilford Press, 2023, <https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462551910/print>.
- [42] K. A. Bollen, “Structural equations with latent variables,” *Structural Equations with Latent Variables*, pp. 1–514, Jan. 2014, <https://doi.org/10.1002/9781118619179>.
- [43] J. L. Horn, “A rationale and test for the number of factors in factor analysis,” *Psychometrika*, vol. 30, no. 2, pp. 179–185, Jun. 1965, <https://doi.org/10.1007/BF02289447>.
- [44] R. B. Cattell, “The Scree Test For The Number Of Factors,” *Multivariate Behav. Res.*, vol. 1, no. 2, pp. 245–276, Apr. 1966, https://doi.org/10.1207/s15327906mbr0102_10.
- [45] L. T. Hu and P. M. Bentler, “Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives,” *Struct. Equ. Modeling*, vol. 6, no. 1, pp. 1–55, 1999, <https://doi.org/10.1080/10705519909540118>.
- [46] G. W. Cheung and R. B. Rensvold, “Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance,” *Structural Equation Modeling*, vol. 9, no. 2, pp. 233–255, 2002, https://doi.org/10.1207/S15328007SEM0902_5.
- [47] T. Le Quy, G. Friege, and E. Ntoutsu, “A Review of Clustering Models in Educational Data Science Toward Fairness-Aware Learning,” in *Educational Data Science: Essentials, Approaches, and Tendencies*, pp. 43–94, 2023, https://doi.org/10.1007/978-981-99-0026-8_2.
- [48] M. Du, R. Amor, K.-L. Ma, and B. C. Wünsche, “Data visualization for improving financial literacy: A systematic review,” *Visual Informatics*, vol. 10, no. 1, p. 100272, Mar. 2026, <https://doi.org/10.1016/j.visinf.2025.100272>.
- [49] F. A. S. Wibowo *et al.*, “Impact Of Feature Selection On Decision Tree And Random Forest For Classifying Student Study Success,” *Barekeng Jurnal Ilmu Matematika Dan Terapan*, vol. 19, no. 3, pp. 2083–2096, Jul. 2025, <https://doi.org/10.30598/barekengvol19iss3pp2083-2096>.
- [50] M. L. D. Sakti *et al.*, “The Effect Of Sample Size On The Stability Of Xgboost Model Performance In Predicting Student Study Period,” *Barekeng*, vol. 19, no. 4, pp. 2679–2692, Sep. 2025, <https://doi.org/10.30598/barekengvol19iss4pp2679-2692>.
- [51] Z. S. Ibrahim, H. Retnawati, A. Irambona, and B. E. O. Pérez, “Stability of estimation item parameter in IRT dichotomy considering the number of participants,” *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 114–127, Jun. 2024, <https://doi.org/10.21831/reid.v10i1.73055>.
- [52] Y. Huang, Y. Zhou, J. Chen, and D. Wu, “Applying Machine Learning and SHAP Method to Identify Key Influences on Middle-School Students’ Mathematics Literacy Performance,” *J. Intell.*, vol. 12, no. 10, Oct. 2024, <https://doi.org/10.3390/jintelligence12100093>.

-
- [53] R. M. Thorndike, "Book Review : Psychometric Theory," *Appl. Psychol. Meas.*, vol. 19, no. 3, pp. 303–305, Sep. 1995, <https://doi.org/10.1177/014662169501900308>.
- [54] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences," *Statistical Power Analysis for the Behavioral Sciences*, pp. 1–567, Jan. 2013, <https://doi.org/10.4324/9780203771587>.
- [55] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning*. in Springer Texts in Statistics. Cham: Springer International Publishing, 2023, <https://doi.org/10.1007/978-3-031-38747-0>.