

A Reversed Gender Pattern? Mathematical Wellbeing Among Primary Students in Iran

Julia L. Hill

julia.hill2@rmit.edu.au

RMIT University <https://orcid.org/0000-0002-8011-0319>

Ebrahim Mohammadpour

Farhangian University <https://orcid.org/0000-0002-9208-4015>

Veysel Akcakin

Usak University <https://orcid.org/0000-0002-7705-0722>

Research Article

Keywords: Affect, Elementary school, Gender-equality paradox, Middle East, Mixed methods, Values

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11535194/v1>

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

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Research on mathematical wellbeing (MWB) has typically found that girls report lower MWB than boys, yet almost all of this evidence comes from Western countries. Using the MWB framework, which conceptualises wellbeing in mathematics as the fulfilment of seven ultimate values within the learning process, we examined gender patterns in MWB among 472 students in Grades 4 to 6 in Iran, a context with no prior MWB research. A concurrent mixed-methods design combined a Persian adaptation of the MWB questionnaire with open-ended written responses. Confirmatory factor analysis largely supported the seven-factor structure, although the Relationships value was weak, and the scale showed full (scalar) measurement invariance across gender. Girls scored significantly higher than boys on all seven ultimate values, with the largest differences on positive emotions and accomplishment, and girls' scores were markedly more homogeneous. Open-ended responses showed a broadly consistent pattern: girls referred more often than boys to enjoyment, pride, and understanding, boys' accounts centred more narrowly on correct answers and marks, and both genders described relational wellbeing mainly through their teacher rather than their peers. We discuss these findings in relation to Iranian girls' mathematics achievement, the Iranian schooling context, and the gender-equality paradox, and consider what this exploratory account from a Middle Eastern context contributes to a field so far dominated by Western evidence.

1 Introduction

Student wellbeing has become an international education priority, yet within mathematics education the field has been slower to develop discipline-specific accounts of what it means for a student to thrive (White & Kern, 2018). The construct of mathematical wellbeing (MWB) was developed to address this gap and is grounded in value fulfilment theory, which holds that people fare well when they fulfil the values that matter to them over time (Tiberius, 2018). Applying this view to mathematics education, MWB conceptualises wellbeing in mathematics as the fulfilment of students' core values within the learning process (Hill et al., 2022). Empirical work on MWB has grown steadily but remains geographically concentrated, drawing mainly on samples from Australia, New Zealand, the United States, the United Kingdom, Türkiye, and China (Campbell & Bean, 2025; Hill & Hunter, 2024; Hill & Seah, 2023; Hunter & Hill, 2024). The Middle East is, to date, entirely absent from this literature, leaving unexamined whether the seven-factor MWB structure, and the gender patterns documented within it, generalise to a region with markedly different schooling structures and cultural norms around gender and mathematics.

Gender has been one of the most consistent lines of inquiry in MWB research, and the findings so far point in a fairly uniform direction. In an Australian sample of Year 8 students, girls reported significantly lower MWB than boys across all seven ultimate values, despite rating those same values as more important (Hill et al., 2024). In a larger and more culturally diverse New Zealand sample, gender differences were smaller and mostly non-significant, though girls again rated several values as more important than boys (Hill & Hunter, 2024). Across the small number of studies conducted to date, then, the direction of the gender gap in MWB ranges from a modest disadvantage for girls to approximate

parity; a broad, sustained advantage for girls has not previously been reported, and MWB has not yet been examined in the Middle East.

All of this evidence, however, comes largely from Western countries, and gender patterns in mathematics are known to vary across cultures and education systems (Else-Quest et al., 2010). Iran offers an important contrast to these contexts. Schooling in Iran has been organised by gender since 1979: from the first year of primary school onward, boys and girls attend separate schools; at the primary level, girls are mostly taught by women, whereas boys may be taught by men or women. Iran is also a context in which the well-documented gender gap in mathematics achievement favouring boys, present in the earliest Trends in International Mathematics and Science Study (TIMSS) cycles, reversed direction from 2003 onward, with Iranian girls since outperforming, or performing on par with, boys (Mullis et al., 2020).

This combination, low measured societal gender equality alongside an achievement gap favouring girls, situates Iran within the gender-equality paradox discussed in Section 3.2 (Breda et al., 2020; Stoet & Geary, 2018). Together with the intense investment in mathematics that Iran's national university entrance examination (the Konkur), which determines access to university places and fields of study, demands of both genders, this makes Iran a theoretically important and previously unexamined context in which to examine gender patterns in MWB.

This study provides the first empirical examination of MWB among students in Iran and, to our knowledge, in the Middle East more broadly. Because the sample is drawn from a single mid-sized city and restricted to the primary years, our aim is to characterise MWB within this context in its own right, rather than to draw firm comparisons with other countries or schooling systems, for which the present sample was not designed.

2 Mathematical Wellbeing

The MWB framework is grounded in Tiberius's (2018) value fulfilment theory, which proposes that a life goes well to the extent that a person pursues and fulfils their deeply held values over time. Tiberius distinguishes ultimate values, higher-order ends valued for their own sake and broadly consistent across cultures, from instrumental values, the culturally variable means through which people pursue those ends. Applied to mathematics, MWB is defined as the fulfilment of ultimate values within the mathematical learning process (Hill et al., 2022). A systematic review of 40 studies across six continents identified 90 instrumental values that were deductively categorised into seven ultimate values, summarised in Table 1 (Hill et al., 2022). The seven-factor structure has since been supported with students in Australia, New Zealand, Türkiye, and the United States (Akcakin & Hill, 2026; Campbell & Bean, 2025; Hill & Hunter, 2024; Hill et al., 2024), though never previously with an Iranian sample.

Table 1
The Mathematical Wellbeing Framework: Seven Ultimate Values

Ultimate value	Definition (valuing...)
Accomplishment	Achievement, setting and reaching goals, and mastery in mathematics
Cognitions	Mathematical knowledge, skills, and understanding
Engagement	Active involvement, interest, and absorption in mathematics
Meaning	Useful, purposeful, and relevant mathematical learning
Perseverance	Persistence, drive, grit through mathematical challenge
Positive emotions	Pleasurable emotional experiences (e.g., hope, enjoyment, pride) in mathematics
Relationships	Social connection and support in mathematics
<i>Note. Adapted from Hill et al. (2022). Definitions abbreviated.</i>	

3 Gender, the Gender-Equality Paradox, and Mathematics Education in Iran

3.1 Defining gender and gender differences in mathematical wellbeing

We position gender as a socially constructed characteristic rather than a purely biological one, consistent with the shift in mathematics education research from “sex differences” to “gender differences” (Leder, 1992). In this study, gender was recorded as girl or boy, and in Iran’s single-sex school system it also corresponds to the school each student attended. We acknowledge this binary measure as a limitation, noting that research beyond the binary remains challenging in large-scale studies (Becker & Hall, 2024; Forgasz, 2021). We use gender patterns for the overall configuration of differences between girls and boys across the seven ultimate values, gender difference for any specific observed difference, and gender gap for a statistically significant difference.

Beyond the MWB studies reviewed above, a broader literature documents gender differences across narrower facets of mathematical affect: girls consistently report higher mathematics anxiety, lower self-concept, and lower perceived usefulness of mathematics than boys, typically despite comparable achievement (Else-Quest et al., 2010; Goetz et al., 2013; Rodríguez et al., 2020; Skaalvik & Skaalvik, 2004; Watt et al., 2012). Girls also tend to espouse more socially oriented values, including peer collaboration and relational belonging, while boys more often prioritise individual achievement (Barkatsas et al., 2019). The MWB literature summarised above largely extends this pattern to a values-based construct, with girls reporting equal or lower MWB than boys wherever it has been examined.

3.2 The gender-equality paradox

Large-scale international assessments have repeatedly shown that countries with greater overall gender equality do not consistently produce smaller gender gaps in mathematics and, in some domains, produce larger ones. Stoet and Geary (2018) termed this the gender-equality paradox: relative to boys, girls' representation among the highest achievers in science and mathematics, and their propensity to pursue STEM fields, is often greater in less gender-equal countries than in more egalitarian ones. Breda et al. (2020) linked this pattern to the strength of a country's internalised stereotype that mathematics is a male domain, which they found to be stronger, not weaker, in more egalitarian countries. The paradox has to date been documented mainly for achievement and participation outcomes; whether a similar dynamic extends to the more holistic, values-based construct of MWB has not previously been examined. This exploratory study, situated within a single Middle Eastern country, begins to address that question.

3.3 Mathematics education and gender in Iran

Iranian schooling has been single-sex from the first year of primary school since 1979, and this study's schools were no exception: boys and girls were necessarily drawn from different schools. Academic performance in mathematics carries substantial weight in Iran because of the Konkur, the national university entrance examination that channels access to high-status fields such as engineering and medicine and applies the same competitive pressure to boys and girls alike (Atari et al., 2020). Against this backdrop, Iranian girls' mathematics achievement has strengthened relative to boys' over time: having trailed boys in the earliest TIMSS cycles, Iranian girls have matched or exceeded boys' average achievement from 2003 onward (Mullis et al., 2020).

Iran does not participate in the Programme for International Student Assessment (PISA), run by the Organisation for Economic Co-operation and Development (OECD), so a directly comparable measure of Iran's internalised mathematics-gender stereotype is not available. Research on Iranian students' attitudes towards mathematics presents a mixed picture. Attitudes appear to be a strong predictor of mathematics achievement, particularly for girls (Kiamanesh & Pourasghar, 2009; Rafiepour, 2025). Several studies report no significant gender difference in overall attitudes (Kiamanesh & Pourasghar, 2009; Miri, 1981), although differences emerge on specific motivational beliefs, with boys reporting higher self-efficacy and test anxiety and girls higher intrinsic value (Kiamanesh & Pourasghar, 2009). Higher evaluation anxiety among Iranian males has also been reported in a university sample (Vahedi & Farrokhi, 2011), the reverse of the international pattern in which girls typically report higher mathematics and test anxiety (Devine et al., 2012; Goetz et al., 2013). At the same time, a questionnaire-based study of Iranian university students found that mathematics, along with medicine, was perceived as a masculine field, suggesting that a masculine stereotype of mathematics persists among Iranian students (Nejat et al., 2011). The relationship between gender and mathematical affect in Iran therefore appears more nuanced than achievement data alone would suggest.

4 The Current Study

Mathematical wellbeing has not previously been examined in Iran, or more broadly in the Middle East, and almost nothing is known about gender patterns in MWB outside Western contexts. This exploratory, concurrent mixed-methods study addresses that gap by examining MWB among Iranian primary students for the first time. Three research questions guided the study:

RQ1. Does the seven-factor structure of the MWB scale, and its measurement equivalence across gender, hold among Iranian primary students?

RQ2. Are there gender differences in MWB among Iranian primary students, and if so, on which ultimate values and in which direction?

RQ3. How do Iranian boys and girls describe, in their own words, what makes them feel good and function well in mathematics, and how do these accounts help explain the quantitative gender pattern?

5 Method

5.1 Design

This study employed a concurrent mixed-methods design (Creswell & Plano Clark, 2018), in which quantitative and qualitative data were collected within the same questionnaire administration rather than in sequential phases. Quantitative data, comprising the MWB scale, were used to examine the factor structure of the MWB scale, its measurement invariance across gender, and gender differences in latent MWB. Qualitative data, comprising students' open-ended written responses to the same instrument, were analysed thematically with the specific aim of helping to explain the study's central quantitative finding, the pronounced gender differences in MWB, consistent with an embedded design in which qualitative data are nested within, and interpreted in relation to, a primarily quantitative study.

5.2 Participants

Participants were 472 students (227 boys, 48.1%; 245 girls, 51.9%) enrolled in six public schools in a mid-sized Iranian city. Students were drawn from Grades 4 ($n = 163$, 34.5%), 5 ($n = 151$, 32.0%), and 6 ($n = 158$, 33.5%), with a comparable grade composition in the two gender groups (Table 2). Age was not recorded in the questionnaire; grade level therefore serves as the indicator of developmental stage, and students enrolled in Grades 4 to 6 in the Iranian system are typically aged between 9 and 12 years.

Schooling in Iran is single-sex from the primary level onward; boys and girls attend separate schools. In this sample, all girls' classes were taught by women, while five of the nine boys' classes were taught by men and four by women. The boys and girls in this sample were therefore necessarily recruited from different schools, which has implications for the interpretation of the gender comparisons presented in this study.

The schools were not randomly sampled. They were nominated by the provincial educational office to represent a range of student backgrounds. Although all six are public schools, two of the three boys'

schools are in the city centre, serve predominantly high-socioeconomic-status (SES) families, and may collect registration fees; the third is in a low-SES district. The three girls' schools were drawn from different locations across the city. The boys' subsample is thus, if anything, weighted towards more advantaged catchments.

Table 2
Sample Composition by Gender and Grade

Gender	Grade 4	Grade 5	Grade 6	Total
Boys	85	61	81	227
Girls	78	90	77	245
Total	163	151	158	472
Note. N = 472.				

5.3 Measures

Mathematical Wellbeing Scale. Mathematical wellbeing was assessed with the Mathematical Wellbeing (MWB) Scale, developed by Hill et al. (2021, 2024) based on students' own conceptions of what it means to thrive in mathematics. The instrument operationalises MWB across seven ultimate values: Accomplishment, Cognitions, Engagement, Meaning, Perseverance, Positive Emotions, and Relationships, each represented by three items with the exception of Accomplishment, which was represented by four, giving 22 items in total. Every item was rated on an 11-point scale ranging from 0 to 10, with higher scores indicating greater MWB. The instrument has demonstrated strong psychometric properties in Australian (Hill et al., 2024) and New Zealand (Hill & Hunter, 2024) samples but had not previously been adapted for use in Iran. All 22 items were administered in the present study; based on the preliminary analyses reported in the Results, one Accomplishment item was not retained, yielding a final 21-item instrument with three items per ultimate value.

The scale was adapted into Persian following the ITC Guidelines for Translating and Adapting Tests (International Test Commission, 2018). Two independent bilingual experts produced forward translations, which they then reconciled to create a preliminary Persian version. A blind back-translation was then conducted by an independent translator and compared with the original to verify semantic and conceptual equivalence. The Persian version was pilot-tested with 30 students from one boys' and one girls' school (five students from each of Grades 4, 5, and 6 per school) to assess item clarity and comprehensibility, and two Engagement items (Items 8 and 9) were revised to improve clarity and cultural appropriateness. Internal consistency for the full scale was high in the pilot sample (Cronbach's $\alpha = .86$). The adaptation procedure and item wording are reported in Supplementary Material S1.

Open-ended written responses. A single open-response item asked students what makes them feel really good and function well when learning mathematics, and why, worded to parallel the open-response

item used in prior MWB studies (Hill et al., 2021; Hunter & Hill, 2024). Participants wrote responses in Persian, which were then translated into English for analysis.

5.4 Procedure

Data were collected in April 2025 during regular class time. Ethical approval and permission were obtained from the General Department of Education of Kurdistan Province (approval number 1403/83763/5800 – 2701, 4 February 2025) and from school principals. Passive parental consent and active student assent were obtained. Students completed the paper-and-pencil questionnaire in approximately 45 minutes with a researcher present. To minimise socially desirable responding, students were told that there were no right or wrong answers, that the questionnaire would not affect their grades, and that responses would not be shared with their teachers, school, or parents; questionnaires were anonymous, collected by the researcher, and completed without teachers present.

5.5 Data Analysis

Quantitative analysis. Missing data were minimal (item-level $n = 458$ to 469 ; covariance coverage $\geq .95$), so all analyses used the full sample ($N = 472$) with full information maximum likelihood (Enders & Bandalos, 2001). Given the marked negative skew of the items, confirmatory factor analyses were estimated in Mplus 8.10 (Muthén & Muthén, 1998–2017) with robust maximum likelihood (MLR) and evaluated against conventional criteria (CFI and $TLI \geq .90$, $RMSEA \leq .06$, $SRMR \leq .08$; Hu & Bentler, 1999). Reliability and convergent and discriminant validity were assessed through composite reliability (CR), average variance extracted (AVE), and the Fornell–Larcker criterion (Fornell & Larcker, 1981). Measurement invariance across gender was tested through configural, metric, and scalar models (Meredith, 1993; Vandenberg & Lance, 2000), with invariance taken to hold where $\Delta CFI \leq .010$ and $\Delta RMSEA \leq .015$ (Chen, 2007). Latent means were then compared with boys as the reference group, and observed standardised differences (Cohen's d) were computed using the pooled standard deviation. Because gender was constant within school, the six school means were also compared using an exact permutation test as a sensitivity check (Ernst, 2004).

Qualitative analysis. Students' Persian responses were translated into English by the Iranian author, and a second bilingual colleague checked a subset against the original Persian. Responses were analysed in NVivo 15 using combined deductive and inductive thematic analysis (Braun & Clarke, 2006), applying the MWB codebook from prior studies (Hill et al., 2021; Hunter & Hill, 2024) and adding instrumental codes inductively where needed (Supplementary Material S2). Coding was conducted by an author experienced in MWB coding who had established inter-rater reliability in previous studies. Formal inter-rater reliability was not possible because internet restrictions in Iran prevented sustained shared access to the data; instead, 100 responses were coded jointly with the Iranian author in online meetings to ensure that codes reflected students' intended cultural meaning. The frequency of each instrumental value was compared between boys and girls using continuity-corrected chi-square tests, or Fisher's exact test where an expected cell count fell below 5, with Holm-Bonferroni correction within each ultimate value

(Holm, 1979). Because a response could be coded to more than one instrumental value, ultimate-value totals reflect coding references rather than unique respondents.

For the three instrumental values that differed significantly by gender, a further pass identified the contexts students associated with each reported feeling, that is, any content mentioned alongside it (e.g., the teacher, a grade, a specific ability); a response could mention more than one associated context.

6 Results

6.1 Preliminary Analyses

An initial confirmatory factor analysis showed that one Accomplishment item (Accomp4, “I feel like I can handle my maths assignments”) performed markedly worse than all others ($\lambda = .50$), depressing the AVE for Accomplishment below .50. The item was removed, raising the AVE to .56 without materially changing other loadings, and all subsequent results are based on the resulting 21-item model with three items per ultimate value.

6.2 Descriptive Statistics and Correlations

Item means ranged from 6.75 to 9.22, and between 46% and 80% of students selected the scale maximum on every item except Relation2 (“I have friends that support me with my maths learning when I need it”). All items were negatively skewed, indicating uniformly high MWB but also a pronounced ceiling effect. Relation2 differed from every other item ($M = 6.75$; 14.7% of students at the floor and 39.8% at the ceiling) and displayed a bimodal distribution.

Inspection of the group descriptives is instructive (Table 3). Girls reported higher mean scores than boys on every ultimate value, and their responses were consistently less dispersed (e.g., Accomplishment: $SD = 1.12$ for girls vs. 2.14 for boys). These group means are reported descriptively; the gender comparison was conducted at the latent level after measurement invariance had been established (Section 6.4).

Table 3
Descriptive Statistics for the Seven MWB Ultimate Value Scores, Overall and by Gender

Ultimate value	n	Boys M (SD)	Girls M (SD)	Total M (SD)	Skewness	Kurtosis
Accomplishment	470	8.23 (2.14)	9.14 (1.12)	8.70 (1.75)	-2.26	6.01
Cognitions	471	8.26 (2.08)	9.16 (1.16)	8.73 (1.72)	-2.17	5.87
Engagement	470	8.00 (2.14)	8.85 (1.58)	8.44 (1.92)	-1.88	4.05
Meaning	472	8.58 (2.06)	9.28 (1.33)	8.94 (1.75)	-2.70	8.39
Perseverance	471	8.53 (2.05)	9.28 (1.18)	8.92 (1.69)	-2.34	6.13
Positive emotions	468	8.06 (2.46)	8.97 (1.56)	8.53 (2.09)	-1.95	3.61
Relationships	469	7.83 (2.41)	8.52 (1.86)	8.19 (2.16)	-1.43	1.83
<i>Note. MWB ultimate values were scored on a 0–10 metric. Skewness and kurtosis refer to the total sample (SE skewness = .11; SE kurtosis = .23). Boys, n = 227; girls, n = 245.</i>						

All ultimate values were positively and significantly intercorrelated ($r = .24$ to $.70$; Supplementary Table S4), with Relationships the weakest correlate of the others. Discriminant validity was supported for most pairs, with two exceptions: Accomplishment and Cognitions ($r = .70$), and marginally Engagement and Positive Emotions ($r = .67$).

6.3 Confirmatory Factor Analysis

The correlated seven-factor model fitted the data well, scaled $\chi^2(168) = 273.57$, $p < .001$; CFI = .945; TLI = .931; RMSEA = .036, 90% CI [.028, .044]; SRMR = .045. Internal consistency for the 21 retained items was high (Cronbach's $\alpha = .89$). Standardised loadings ranged from .50 to .89 (all $p < .001$; Supplementary Table S5). Positive Emotions was the strongest ultimate value (CR = .87; AVE = .70), whereas Relationships was clearly weak ($\lambda = .50$ to .58; CR = .54; AVE = .28), more so than in the Australian validation sample, where it was the least reliable value but remained acceptable (Hill et al., 2024). The AVE also fell below .50 for Cognitions (.47), Engagement (.45), and Perseverance (.43). The answer to RQ1 is therefore a qualified one: the seven-factor structure was largely supported, with the exception of Relationships.

6.4 Measurement Invariance Across Gender

Constraining factor loadings (metric model) and then item intercepts (scalar model) to equality across gender left model fit essentially unchanged, and all change-in-fit indices met Chen's (2007) criteria

(Table 4). The scale therefore demonstrated full scalar invariance, permitting comparison of latent means, although this conclusion rests on a configural baseline that fitted only marginally (CFI = .891).

Table 4
Tests of Measurement Invariance Across Gender

Model	χ^2 (df)	CFI	RMSEA	SRMR	$\Delta\chi^2$ (Δ df)	p	Δ CFI	Δ RMSEA
Configural	560.83 (336)	.891	.053	.059	–	–	–	–
Metric	567.53 (350)	.894	.051	.073	13.73 (14)	.470	+ .003	–.002
Scalar	588.08 (364)	.891	.051	.077	19.35 (14)	.152	–.003	.000

Note. $n = 227$ boys, 245 girls. Estimator = MLR; difference tests are Satorra–Bentler corrected (Satorra & Bentler, 2001), and each change refers to the preceding model. Δ SRMR was + .014 (metric) and + .004 (scalar). Scalar versus configural, $\Delta\chi^2(28) = 30.48$, $p = .341$, Δ CFI = .000.

6.5 Gender Differences in Latent Mathematical Wellbeing (RQ2)

Girls scored significantly higher than boys on all seven ultimate values (Table 5; Fig. 1). The largest latent differences were for Positive Emotions (+ 0.98) and Accomplishment (+ 0.93) and the smallest for Relationships (+ 0.46); in standardised terms the observed differences were small to moderate ($d = 0.32$ to 0.54). Latent variances were also substantially larger among boys (e.g., Accomplishment, 3.63 vs. 0.81), indicating that girls' MWB was not only higher but markedly more homogeneous. In answer to RQ2, gender differences were broad, consistent in direction, and favoured girls throughout.

Table 5
Latent Mean Differences Between Girls and Boys Across the Seven MWB Ultimate Values

Ultimate value	Difference	SE	95% CI	p	d
Positive Emotions	+ 0.98	0.21	[0.58, 1.39]	< .001	0.45
Accomplishment	+ 0.93	0.17	[0.60, 1.27]	< .001	0.54
Cognitions	+ 0.86	0.16	[0.55, 1.17]	< .001	0.54
Engagement	+ 0.84	0.18	[0.49, 1.20]	< .001	0.45
Meaning	+ 0.71	0.19	[0.34, 1.08]	< .001	0.41
Perseverance	+ 0.63	0.15	[0.33, 0.93]	< .001	0.45
Relationships	+ 0.46	0.16	[0.14, 0.78]	.005	0.32
<p><i>Note. Reference group = boys (latent means fixed at 0); positive values denote higher latent means for girls. Latent differences are in each value's own metric and are not comparable across values. 95% CI = estimate $\pm$ 1.96 $\times$ SE. d = Cohen's d for the observed difference (girls minus boys), using the pooled standard deviation from Table 3.</i></p>					

6.6 Qualitative Accounts of Mathematical Wellbeing (RQ3)

The qualitative pattern was broadly consistent with the quantitative one (Table 6). Girls' responses referred more often than boys' to most ultimate values, and the three instrumental values that differed significantly after Holm correction all favoured girls: Understanding (16.3% vs. 6.2%, $p(\text{Holm}) = .007$), Enjoyment, fun, and happiness (44.1% vs. 12.8%, $p(\text{Holm}) < .001$), and Pride (3.7% vs. 0.0%, $p(\text{Holm}) = .011$). Positive Emotions thus showed both the largest quantitative gap and the clearest qualitative one. Relationships, the value with the smallest quantitative gap, showed no significant difference in its dominant instrumental value, teacher support and care (23.3% vs. 18.5%, $p(\text{Holm}) = .865$).

Table 6
Frequency of Instrumental Values in Open-Ended Responses, by Gender

Value / Instrumental Value	Boys n (%)	Girls n (%)	p	p (Holm)
Accomplishment	195	208	–	–
Learn well / improvement	65 (28.6%)	90 (36.7%)	.076	.684
High marks / tests	84 (37.0%)	62 (25.3%)	.008	.081
Accuracy / correctness	27 (11.9%)	41 (16.7%)	.172	1.000
Smartnesst	6 (2.6%)	3 (1.2%)	.323	1.000
Independent / self accomplishment†	1 (0.4%)	5 (2.0%)	.218	1.000
Trying one's best / motivation†	3 (1.3%)	3 (1.2%)	1.000	1.000
Confidencet	3 (1.3%)	1 (0.4%)	.355	1.000
Competition†	2 (0.9%)	1 (0.4%)	.610	1.000
Goals†	2 (0.9%)	1 (0.4%)	.610	1.000
Preparation†	2 (0.9%)	1 (0.4%)	.610	1.000
Cognitions	56	89	–	–
Understanding	14 (6.2%)	40 (16.3%)	< .001	.007*
New learning / acquiring knowledge	18 (7.9%)	18 (7.3%)	.948	1.000
Solving / problem solving	10 (4.4%)	14 (5.7%)	.662	1.000
Recall / memorisation	5 (2.2%)	6 (2.4%)	1.000	1.000
Specific maths topics†	4 (1.8%)	6 (2.4%)	.753	1.000
Efficiency & fluency†	3 (1.3%)	5 (2.0%)	.726	1.000
Materials / instruments†	1 (0.4%)	0 (0.0%)	.481	1.000
Multiple solutions†	1 (0.4%)	0 (0.0%)	.481	1.000
Engagement	126	143	–	–
Listening	60 (26.4%)	73 (29.8%)	.478	1.000
Attention / focus / silence	42 (18.5%)	46 (18.8%)	1.000	1.000
Challenging maths	12 (5.3%)	10 (4.1%)	.688	1.000

*Note. Boys, n = 227; girls, n = 245. Bold rows sum coding references across instrumental values and were not tested (–). † Fisher's exact test; other tests are continuity-corrected chi-square tests. p (Holm) is adjusted within each ultimate value. * p (Holm) < .05. All codes, with illustrative quotations, are listed in Supplementary Table S3.*

Value / Instrumental Value	Boys n (%)	Girls n (%)	p	p (Holm)
Accomplishment	195	208	–	–
Show/have interest	7 (3.1%)	9 (3.7%)	.921	1.000
Behavioural compliancet†	4 (1.8%)	4 (1.6%)	1.000	1.000
Maths games†	1 (0.4%)	1 (0.4%)	1.000	1.000
Meaning	53	80	–	–
Links to future / jobs	43 (18.9%)	55 (22.4%)	.409	1.000
Personally meaningful / utility	7 (3.1%)	22 (9.0%)	.013	.054
Connects to other subjects†	1 (0.4%)	2 (0.8%)	1.000	1.000
Real world links†	2 (0.9%)	1 (0.4%)	.610	1.000
Perseverance	29	36	–	–
Practice / revision / homework	19 (8.4%)	25 (10.2%)	.599	1.000
Hard work & effort	10 (4.4%)	11 (4.5%)	1.000	1.000
Positive Emotions	33	122	–	–
Enjoyment / fun / happiness	29 (12.8%)	108 (44.1%)	< .001	< .001*
Pridet†	0 (0.0%)	9 (3.7%)	.004	.011*
No anxiety / stresst†	2 (0.9%)	2 (0.8%)	1.000	1.000
Safe & relaxing†	2 (0.9%)	2 (0.8%)	1.000	1.000
Relationships	61	87	–	–
Teacher support & care	42 (18.5%)	57 (23.3%)	.247	.865
Rewards / praise	12 (5.3%)	5 (2.0%)	.100	.602
Family support	6 (2.6%)	13 (5.3%)	.216	.865
Sharing ideas & explanation†	0 (0.0%)	5 (2.0%)	.062	.435
Reciprocity / helping others†	0 (0.0%)	4 (1.6%)	.125	.623
Peer support & collaboration†	1 (0.4%)	2 (0.8%)	1.000	1.000
General / external support†	0 (0.0%)	1 (0.4%)	1.000	1.000
<p><i>Note. Boys, n = 227; girls, n = 245. Bold rows sum coding references across instrumental values and were not tested (–). † Fisher's exact test; other tests are continuity-corrected chi-square tests. p (Holm) is adjusted within each ultimate value. * p (Holm) < .05. All codes, with illustrative quotations, are listed in Supplementary Table S3.</i></p>				

6.6.1 Associated Contexts and Illustrative Responses

Table 7 reports the contexts students associated with the three instrumental values that differed significantly by gender.

Table 7
Contexts Associated With the Three Significant Instrumental Values, by Gender

Associated context	Boys n (%)	Girls n (%)
Enjoyment, fun, and happiness (n = 29 boys, 108 girls)	–	–
Teacher (explanation quality, kindness, praise)	4 (13.8%)	30 (27.8%)
Grade / exam / test success	3 (10.3%)	22 (20.4%)
Getting the correct answer	6 (20.7%)	15 (13.9%)
Intrinsic liking / love of maths (general)	5 (17.2%)	18 (16.7%)
New learning / growing knowledge	1 (3.4%)	11 (10.2%)
Helping or sharing with peers	2 (6.9%)	8 (7.4%)
Moment of understanding	0 (0.0%)	7 (6.5%)
Family / parents proud or pleased	0 (0.0%)	3 (2.8%)
Pride (girls only, n = 9)	–	–
Self-pride (“proud of myself”)	–	5 (55.6%)
Tied specifically to solving a hard problem	–	3 (33.3%)
Teacher’s or friends’ social validation	–	2 (22.2%)
Understanding (n = 14 boys, 40 girls)	–	–
Teacher’s explanation	3 (21.4%)	12 (30.0%)
Framed around future usefulness	1 (7.1%)	7 (17.5%)
Grade / score / performance link	2 (14.3%)	6 (15.0%)
Stress or difficulty easing once understood	0 (0.0%)	7 (17.5%)
Praise / social recognition	2 (14.3%)	2 (5.0%)
Quiet / focus / concentration as a condition	0 (0.0%)	2 (5.0%)
Explicit understanding-vs-memorising contrast	1 (7.1%)	1 (2.5%)
<i>Note. Percentages are of coding references for each instrumental value and do not sum to 100, as a response could mention more than one associated context.</i>		

Enjoyment, fun, and happiness. Girls associated enjoyment with a wider range of contexts than boys did. Girls most often mentioned the teacher alongside enjoyment (27.8% vs. 13.8%), for example, “My teacher is very kind, and whenever we have maths, I get a good feeling and wish the whole day was just maths” (Girl), and mentioned grades more often than boys (20.4% vs. 10.3%): “Today, because I had practised well, I got a good grade in maths, and this made me happy” (Girl). Boys more often associated enjoyment with getting the correct answer (20.7% vs. 13.9%), for example, “When I know the correct answer to problems, I feel happy” (Boy), or with a general liking of the subject: “I just naturally like maths, so I’m happy” (Boy).

Pride. Pride was reported by girls only ($n = 9$) and was mostly self-referential (55.6%), often mentioned alongside solving a difficult problem: “What makes me very happy is being able to solve hard questions. I feel proud of myself for solving them” (Girl).

Understanding. For both genders, the context most often associated with understanding was the teacher’s explanation (30.0% of girls’ references, 21.4% of boys’): “When the teacher explains the lesson well and I understand, I feel good” (Girl). Only girls described understanding as relieving stress (17.5% vs. 0.0%): “Maths is stressful for me, but when I listen well to learn well and reduce my stress, it makes me feel good” (Girl).

7 Discussion

This study examined MWB for the first time among primary students in Iran. The seven-factor structure was largely supported and invariant across gender, although Relationships performed weakly. Girls scored higher than boys on all seven ultimate values, and their open-ended accounts referred more often to understanding, enjoyment, and pride, while both genders described relational wellbeing mainly through their teacher. We consider each research question in turn, and then what the Iranian case may add to the gender-equality paradox literature.

7.1 The Scale’s Structure, and the Weak Relationships Value (RQ1)

The MWB scale functioned largely as intended in its first application in Iran. The seven-factor structure was supported and invariant across gender, although Accomplishment and Cognitions were not clearly discriminated ($r = .70$). This overlap may reflect the age of the sample, as students aged 9 to 12 may not yet experience feeling successful and feeling that they understand as distinct, and in the qualitative data understanding was at times linked directly to a good mark. The broader result adds Iran to the Australian, New Zealand, Turkish, and US samples in which the seven-factor structure has been supported (Akcakin & Hill, 2026; Campbell & Bean, 2025; Hill & Hunter, 2024; Hill et al., 2024), and is consistent with the claim that ultimate values are broadly shared across cultures while instrumental values vary (Hill et al., 2022; Tiberius, 2018).

Relationships was the clear exception, and its weakness appears to reflect how relationships are experienced in Iranian mathematics classrooms as much as the items themselves. Relationships was weak on every criterion ($CR = .54$, $AVE = .28$), and the friends item (Relation2) was bimodal, with 14.7% of students choosing the lowest possible score. In their written accounts, students described the relational side of their MWB almost entirely through their teacher (18.5% of boys' and 23.3% of girls' responses), while peers were mentioned by fewer than 1% of either gender. Where mathematics is organised around listening to the teacher and working quietly, the two most frequently coded forms of Engagement in this study, items spanning teacher, peer, and other support may not cohere into a single experience. This interpretation is consistent with earlier MWB research showing that relationships are fulfilled largely through peers among Australian students but through teacher praise and family encouragement among Chinese students (Hill et al., 2021; Hill & Seah, 2023), and with Relationships being the least reliable value even in the Australian validation sample (Hill et al., 2024). Future adaptations could model teacher, peer, and family support separately.

7.2 The Size of the Gender Gap, and Possible Explanations (RQ2)

The gender difference in this sample was broad, consistent, and opposite in direction to most MWB research to date. Girls reported higher MWB on every ultimate value, with small to moderate observed effects ($d = 0.32$ to 0.54), similar in size to the gap favouring boys recently reported in England but reversed (Hill et al., 2026). Girls' largest advantages, on Positive Emotions, Accomplishment, and Cognitions, also fell on the values on which boys have tended to report higher MWB elsewhere, whereas Relationships, the one value favouring girls in recent cross-national data, showed the smallest gap here (Hill et al., 2024; Hill et al., 2026). This suggests that the direction of gender differences in MWB is shaped by context rather than fixed. The Iranian pattern is not without precedent, as Turkish girls in Year 6 also reported higher MWB than boys before the pattern reversed by Year 9 (Hill et al., 2026), and gender differences in New Zealand were small and non-significant (Hill & Hunter, 2024).

Girls' MWB was also markedly more homogeneous than boys'. Latent variances were substantially larger among boys on every value (e.g., Accomplishment, 3.63 vs. 0.81). Some of this difference reflects compressed variance, because girls' means sat closer to the scale maximum, but the dispersion among boys also suggests that their average conceals a group of boys with considerably lower MWB, a group that may warrant particular attention from teachers.

The direction of the MWB gap is consistent with Iranian girls' standing in mathematics achievement. Having trailed boys in the earliest TIMSS cycles, Iranian girls have performed on par with or above boys since 2003 (Mullis et al., 2020), and in TIMSS 2023 Grade 4 girls averaged 425 and boys 414, a non-significant difference (von Davier et al., 2024). Boys' Grade 8 achievement also declined across the 1999 to 2007 cycles while girls' improved (Kiamanesh & Mohsenpour, 2010), consistent with the absence of a gap favouring boys in several Middle Eastern countries (Fryer & Levitt, 2010). Yet whereas the achievement gap in TIMSS 2023 was small and non-significant, the MWB differences in this study were

small to moderate, suggesting that girls' wellbeing in mathematics reflects more than their performance alone.

The organisation of Iranian schooling by gender offers one possible contextual explanation for girls' higher MWB, although its effect cannot be isolated with this design. Girls in this study learned mathematics in all-girls schools, usually taught by women, and their accounts centred on enjoyment, understanding, and pride. Single-sex classrooms may reduce stereotype threat and change the frame of reference against which students judge themselves, so that girls compare their mathematics with that of other girls rather than with boys who may be perceived as the more able group. Such conditions may allow girls to participate, take risks, and experience accomplishment and enjoyment. This is consistent with evidence that stereotype threat depresses women's mathematics performance (Nguyen & Ryan, 2008; Spencer et al., 1999), that academic self-concept depends on the comparison group (Marsh & Hau, 2003), and that girls often outperform boys where single-sex schooling prevails (Basharat & Steinmann, 2025). Single-sex schooling is not, however, a sufficient explanation, as controlled studies show little advantage for girls' mathematics attitudes or performance (Pahlke et al., 2014). Notably, the school catchment confound runs against the observed pattern, as two of the three boys' schools served more advantaged families.

Changing gender beliefs and aspirations may also contribute to girls' higher MWB. Girls in this study more often described mathematics as personally meaningful (9.0% vs. 3.1%, a difference that approached significance), and from a value-fulfilment perspective, mathematics that is bound to valued future goals should support wellbeing (Tiberius, 2018). This interpretation is consistent with evidence that Iranian girls in post-revolutionary cohorts have held positive attitudes towards education and roles beyond the family (Aghajanian et al., 2007), that their educational and career aspirations now match or exceed boys' (Mehranpour et al., 2024), and that their attitudes towards mathematics are more closely tied to achievement (Rafiepour, 2025) and more intrinsically valued (Kiamanesh & Pourasghar, 2009). This should not be overstated, however, as a masculine stereotype of mathematics has also been reported among Iranian students (Nejat et al., 2011).

7.3 Interpreting the Gender Difference Through Girls' and Boys' Accounts (RQ3)

Students' accounts suggest that girls' higher MWB was associated with mastery and meaning, whereas boys' MWB was associated more narrowly with performance. Three instrumental values differed significantly by gender: Understanding, Enjoyment/fun/happiness, and Pride, all favouring girls. Girls more often linked these values to understanding relieving stress and to the future usefulness of what they learned (Table 7), to personally meaningful content (9.0% vs. 3.1%, Holm-adjusted $p = .054$), and to learning and improvement (36.7% vs. 28.6%, Holm-adjusted $p = .684$), and Pride, reported by girls only, was framed almost entirely in self-referential terms. Boys, by contrast, most often associated enjoyment with getting the correct answer and referred to grades more often in their accounts of accomplishment (37.0% vs. 25.3%, Holm-adjusted $p = .081$).

This pattern suggests that girls' MWB rested on a broader base than boys'. Girls' positive feelings were linked to understanding and meaning, but also to the teacher (27.8% of girls' enjoyment references vs. 13.8% of boys') and to grades (20.4% vs. 10.3%), indicating that girls drew on both intrinsic and extrinsic sources of wellbeing, rather than one in place of the other. In self-determination theory terms, girls' accounts drew together competence, relatedness, and personal meaning, the conditions under which motivation becomes more autonomous (Deci & Ryan, 2000), whereas boys' accounts centred on competence demonstrated through performance. This runs counter to other primary-school samples, in which boys typically report higher intrinsic motivation for mathematics (Rodríguez et al., 2020), and to the broader pattern of girls reporting lower interest and confidence despite comparable achievement (Else-Quest et al., 2010; Goetz et al., 2013). By contrast, teacher support and care was reported at similar rates by both genders, suggesting a shared, teacher-centred relational culture rather than a gendered one.

While the quantitative and qualitative findings were largely consistent, they diverged for some ultimate values, and these differences are important to consider. Engagement and Perseverance showed quantitative gaps favouring girls, yet boys and girls referred to listening, attention, practice, and effort at almost identical rates. Conversely, the qualitative gap for enjoyment (44.1% vs. 12.8%) was far larger than the quantitative gap for Positive Emotions. One explanation is that girls were more inclined to put positive feelings into words, so part of the qualitative gap may reflect emotional expression as well as experience, consistent with meta-analytic evidence that girls express more positive emotion than boys from childhood (Chaplin & Aldao, 2013). The small number of boys' references to some values (e.g., 14 for Understanding) also calls for caution.

7.4 Mathematical Wellbeing and the Gender-Equality Paradox

The Iranian findings suggest that the gender-equality paradox may extend from achievement to wellbeing. Iran ranks among the lowest countries on the Global Gender Gap Index (World Economic Forum, 2024), yet girls in this sample reported higher MWB than boys on every ultimate value. The paradox would anticipate this pattern if, where the stereotype that mathematics is a male domain is less strongly internalised, girls' experience of mathematics is less diminished (Breda et al., 2020; Stoet & Geary, 2018). Our cross-national MWB research points in the same direction: the largest gap favouring boys was found in England, where the stereotype is strongly internalised, whereas Turkish girls, in the context with the weakest internalised stereotype of the four countries studied, reported higher MWB than boys in Year 6 (Hill et al., 2026).

Similar patterns have been documented among Palestinian/Arab Israeli students, another Middle Eastern context in which girls outperform boys in school mathematics and are more represented in advanced coursework. Palestinian/Arab Israeli high school students attributed negative relationships with mathematics mainly to boys, and associated girls with caring about success and finding mathematics interesting (Rubel et al., 2022), much as Iranian girls in this study described enjoyment,

pride, and understanding while boys' MWB was more dispersed. Yet teachers in Arabic-language schools in Israel disagreed less strongly than teachers in Hebrew-language schools with narratives of boys' superiority and innate mathematical ability, leading Rubel and Shahbari (2024) to conclude that girls' advantage is likely shaped by factors other than a societal rejection of these narratives. This parallels the Iranian case, in which a masculine stereotype of mathematics has been reported among students (Nejat et al., 2011) alongside girls' strong achievement and, in this study, their higher MWB.

These parallels should nevertheless be read cautiously. Iran lacks a comparable measure of stereotype internalisation, and greater gender equity has also been associated with smaller gender differences in mathematics attitudes (Else-Quest et al., 2010), so the cross-national evidence is mixed. Moreover, the Iranian contrast involves gender-segregated schooling, not only lower societal equality, and may operate through the frame-of-reference and stereotype-threat processes discussed above. The Iranian case is therefore best read as an invitation to test whether, and how, the paradox extends to wellbeing.

8 Limitations, Future Directions, and Implications

This study has several limitations. Because Iranian schooling is single-sex, gender was confounded with school, teacher, and catchment, and the six schools were purposively nominated; the design cannot separate gender from school-level differences or isolate an effect of single-sex instruction. Because students were nested in only six schools, neither multilevel modelling nor cluster-robust standard errors could be applied reliably with so few clusters (McNeish & Stapleton, 2016). The significance tests for the gender comparisons should therefore be read as approximate. As a sensitivity check, we compared the six school means directly: each girls' school had a higher mean MWB score than each boys' school, the most extreme of the 20 possible allocations of schools to gender (exact permutation test, two-tailed $p = .10$, the smallest value attainable with six schools; Ernst, 2004). The sample came from one mid-sized city and Grades 4 to 6, and was cross-sectional; as MWB changes across schooling (Akcakin & Hill, 2026), the findings should not be generalised to older students or to Iran as a whole. The configural invariance model fitted only marginally (CFI = .891), Relationships was weak, and the ceiling effect restricted variance and compressed girls' scores in particular. The qualitative strand relied on a single written item, translated from Persian into English for coding. Formal inter-rater reliability could not be established, although joint coding of 100 responses with the Iranian author supported culturally accurate interpretation.

These limitations suggest directions for future research. Longitudinal work could examine whether Iranian girls' higher MWB persists into secondary school and the high-stakes years leading to the Konkur, as the Turkish pattern suggests it may not. Interviews with students and teachers would allow the relational and frame-of-reference mechanisms proposed here to be examined directly. Studies comparing single-sex and coeducational settings within the same country, and comparing MWB gender gaps across countries that differ in gender equality and stereotype internalisation, would test the extension of the gender-equality paradox proposed in Section 7.4.

For Iranian teachers, the findings point less to girls than to boys. The wider dispersion of boys' MWB suggests that, within an average that remains high, some boys experience mathematics considerably less positively, and boys' accounts centred narrowly on correct answers and marks. Teaching that makes understanding, meaning, and the teacher relationship as visible to boys as they appear to be to girls may broaden the base of boys' MWB. For the international field, the Iranian case shows that girls' lower MWB, documented in several Western countries, is not inevitable. This is not an argument for single-sex schooling, which controlled evidence does not support (Pahlke et al., 2014), but for attending to the classroom conditions, such as the frames of reference students use and the relationships through which they experience mathematics, that may allow girls' MWB to flourish in coeducational settings.

9 Conclusion

This exploratory study offers the first account of MWB among students in Iran and, to our knowledge, in the Middle East. The MWB scale was largely supported in its Persian adaptation, and in contrast to the largely Western samples studied to date, girls reported higher MWB than boys on every ultimate value. Their accounts suggest that this advantage rested on enjoyment, understanding, and pride, often experienced through the relationship with their teacher. While the design cannot isolate why, the findings add to evidence that gender patterns in MWB are shaped by context, and that research from non-Western school systems is needed if the field is to understand, rather than assume, how mathematical wellbeing is gendered.

Declarations

Competing interests.

The authors have no competing interests to declare.

Ethics approval.

The study was approved by the General Department of Education of Kurdistan Province, Iran (approval number 1403/83763/5800 – 2701, 4 February 2025), and by the principals of the participating schools.

Consent to participate.

Parents were informed about the study and could withdraw their child (passive consent), and all participating students gave their assent.

Funding.

No funding was received for this study.

Data availability.

The data are not publicly available because of the conditions of ethical approval but are available from the corresponding author on reasonable request.

References

1. Aghajanian A, Tashakkori A, Thompson VD, Mehryar AH, Kazemipour S (2007) Attitudes of Iranian female adolescents toward education and nonfamilial roles: A study of a postrevolutionary cohort. *Marriage Family Rev* 42(1):49–64. https://doi.org/10.1300/J002v42n01_04
2. Akcakin V, Hill JL (2026) *Mathematical wellbeing across the school years in a high-stakes examination system: Value fulfilment profiles from Grades 3 to 12* [Manuscript submitted for publication]. Faculty of Education, Usak University
3. Atari M, Graham J, Deghani M (2020) Foundations of morality in Iran. *Evol Hum Behav* 41(5):367–384. <https://doi.org/10.1016/j.evolhumbehav.2020.07.014>
4. Barkatsas T, Law HY, Seah WT, Wong NY (2019) The valuing of mathematics learning in schools: A gendered perspective. *Int J Emerg Math Educ* 3(1):41–56. <https://doi.org/10.12928/ijeme.v3i1.11648>
5. Basharat S, Steinmann I (2025) Divided desks, divided school climate conditions? Investigating the implications of single-sex schooling. *Equity in Education & Society*. Advance online publication. <https://doi.org/10.1177/27526461251348603>
6. Becker JR, Hall J (2024) Research on gender and mathematics: Exploring new and future directions. *ZDM – Math Educ* 56(1):141–151. <https://doi.org/10.1007/s11858-023-01510-6>
7. Braun V, Clarke V (2006) Using thematic analysis in psychology. *Qualitative Res Psychol* 3(2):77–101. <https://doi.org/10.1191/1478088706qp063oa>
8. Breda T, Jouini E, Napp C, Thebault G (2020) Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences*, 117(49), 31063–31069. <https://doi.org/10.1073/pnas.2008704117>
9. Campbell TG, Bean B (2025) Factors influencing young children’s mathematical wellbeing in the United States. *Soc Indic Res* 178(1):255–278. <https://doi.org/10.1007/s11205-025-03581-2>
10. Chaplin TM, Aldao A (2013) Gender differences in emotion expression in children: A meta-analytic review. *Psychol Bull* 139(4):735–765. <https://doi.org/10.1037/a0030737>
11. Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model* 14(3):464–504. <https://doi.org/10.1080/10705510701301834>
12. Creswell JW, Plano Clark VL (2018) *Designing and conducting mixed methods research*, 3rd edn. SAGE
13. Deci EL, Ryan RM (2000) The what and why of goal pursuits: Human needs and the self-determination of behavior. *Psychol Inq* 11(4):227–268.

https://doi.org/10.1207/S15327965PLI1104_01

14. Devine A, Fawcett K, Szűcs D, Dowker A (2012) Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behav Brain Funct* 8:33. <https://doi.org/10.1186/1744-9081-8-33>
15. Else-Quest NM, Hyde JS, Linn MC (2010) Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychol Bull* 136(1):103–127. <https://doi.org/10.1037/a0018053>
16. Enders CK, Bandalos DL (2001) The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Struct Equ Model* 8(3):430–457. https://doi.org/10.1207/S15328007SEM0803_5
17. Ernst MD (2004) Permutation methods: A basis for exact inference. *Stat Sci* 19(4):676–685. <https://doi.org/10.1214/088342304000000396>
18. Forgasz H (2021) Gender: A dilemma for large-scale studies in mathematics education. *Math Educ Res J* 33(4):631–640. <https://doi.org/10.1007/s13394-020-00353-8>
19. Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50. <https://doi.org/10.1177/002224378101800104>
20. Fryer RG Jr., Levitt SD (2010) An empirical analysis of the gender gap in mathematics. *Am Economic Journal: Appl Econ* 2(2):210–240. <https://doi.org/10.1257/app.2.2.210>
21. Goetz T, Bieg M, Lüdtke O, Pekrun R, Hall NC (2013) Do girls really experience more anxiety in mathematics? *Psychol Sci* 24(10):2079–2087. <https://doi.org/10.1177/0956797613486989>
22. Hill JL, Hunter J (2024) Diverse students' mathematical wellbeing. *New Z J Educational Stud* 59(1):211–234. <https://doi.org/10.1007/s40841-024-00318-6>
23. Hill JL, Hunter J, Akcakin V, Green A, Jones I (2026) *Cross-national gender patterns in students' mathematical wellbeing* [Manuscript submitted for publication]. School of Education, RMIT University
24. Hill JL, Kern ML, Seah WT, van Driel J (2021) Feeling good and functioning well in mathematics education: Exploring students' conceptions of mathematical wellbeing and values. *ECNU Rev Educ* 4(2):349–375. <https://doi.org/10.1177/2096531120928084>
25. Hill JL, Kern ML, Seah WT, van Driel J (2022) Developing a model of mathematical wellbeing through a thematic analysis of the literature. In C. Fernández, S. Llinares, Á. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 379–386). PME
26. Hill JL, Kern ML, Seah WT, van Driel J (2024) To what extent are students fulfilling their values and thriving in mathematics education? The case for Victoria, Australia. In Y. Dede, G. Marschall, & P. Clarkson (Eds.), *Values and valuing in mathematics education: Moving forward into practice* (pp. 259–283). Springer. https://doi.org/10.1007/978-981-99-9454-0_13
27. Hill JL, Seah WT (2023) Student values and wellbeing in mathematics education: Perspectives of Chinese primary students. *ZDM – Math Educ* 55(2):385–398. <https://doi.org/10.1007/s11858-022-01418-7>

28. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6(2):65–70
29. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55.
<https://doi.org/10.1080/10705519909540118>
30. Hunter J, Hill JL (2024) Exploring mathematical wellbeing across cultures: Insights from diverse students. *Math Educ Res J* 37(3):529–556. <https://doi.org/10.1007/s13394-024-00500-5>
31. International Test Commission (2018) ITC guidelines for translating and adapting tests. *Int J Test* 18(2):101–134. 2nd edn. <https://doi.org/10.1080/15305058.2017.1398166>
32. Kiamanesh AR, Mohsenpour M (2010) *Trends in factors affecting Iranian eighth graders' mathematics achievement by gender (TIMSS 1999, 2003 & 2007)* [Paper presentation]. 4th IEA International Research Conference, Gothenburg, Sweden
33. Kiamanesh A, Pourasghar N (2009) Gender differences in variables related to students' math performance and the role of these variables on math achievement. *J Mod Psychol Res* 4(13):161–190
34. Leder GC (1992) Mathematics and gender: Changing perspectives. In: Grouws DA (ed) *Handbook of research on mathematics teaching and learning*. Macmillan, pp 597–622
35. Marsh HW, Hau K-T (2003) Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *Am Psychol* 58(5):364–376. <https://doi.org/10.1037/0003-066X.58.5.364>
36. McNeish DM, Stapleton LM (2016) The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychol Rev* 28(2):295–314. <https://doi.org/10.1007/s10648-014-9287-x>
37. Mehranpour S, Hosseini F, Firoozabadi L (2024) The impact of gender stereotypes on the participation of women in STEM fields in Iran. *Res Adv Educ* 3(3):15–20
38. Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58(4):525–543. <https://doi.org/10.1007/BF02294825>
39. Miri E (1981) *The effect of educational practices on attitudes toward mathematics among two age groups of Iranian and American students* [Doctoral dissertation, Fordham University]
40. Mullis IVS, Martin MO, Foy P, Kelly DL, Fishbein B (2020) TIMSS 2019 international results in mathematics and science. TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/timss2019/international-results/>
41. Muthén LK, Muthén BO (1998–2017) *Mplus user's guide* (8th ed.). Muthén & Muthén
42. Nejat P, Pournaghash Tehrani SS, Hatami J (2011) Gender stereotypes related to different academic fields among Iranian university students. *Social Psychol Res* 1(3):1–18
43. Nguyen H-HD, Ryan AM (2008) Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *J Appl Psychol* 93(6):1314–1334. <https://doi.org/10.1037/a0012702>

44. Pahlke E, Hyde JS, Allison CM (2014) The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychol Bull* 140(4):1042–1072. <https://doi.org/10.1037/a0035740>
45. Rafiepour A (2025) Identify effective factors in eighth grade Iranian students' mathematical performance at TIMSS 2019. *J Educational Psychol Stud*, 22(1)
46. Rodríguez S, Regueiro B, Piñeiro I, Estévez I, Valle A (2020) Gender differences in mathematics motivation: Differential effects on performance in primary education. *Front Psychol* 10:3050. <https://doi.org/10.3389/fpsyg.2019.03050>
47. Rubel L, Ayalon M, Shahbari J, Awad F (2022) Mathematics as gendered? Views from Palestinian/Arab Israeli high schoolers. In C. Fernández, S. Llinares, Á. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 355–362). PME
48. Rubel L, Shahbari J (2024) Unpacking the gender equality paradox: Narratives about mathematics and gender among teachers. *Research in Mathematics Education*. Advance online publication. <https://doi.org/10.1080/14794802.2024.2435016>
49. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66(4):507–514. <https://doi.org/10.1007/BF02296192>
50. Skaalvik S, Skaalvik EM (2004) Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles* 50(3–4):241–252. <https://doi.org/10.1023/B:SERS.0000015555.40976.e6>
51. Spencer SJ, Steele CM, Quinn DM (1999) Stereotype threat and women's math performance. *J Exp Soc Psychol* 35(1):4–28. <https://doi.org/10.1006/jesp.1998.1373>
52. Stoet G, Geary DC (2018) The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychol Sci* 29(4):581–593. <https://doi.org/10.1177/0956797617741719>
53. Tiberius V (2018) Well-being as value fulfillment: How we can help each other to live well. Oxford University Press. <https://doi.org/10.1093/oso/9780198809494.001.0001>
54. Vahedi S, Farrokhi F (2011) A confirmatory factor analysis of the structure of Abbreviated Math Anxiety Scale. *Iran J Psychiatry* 6(2):47–53
55. Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Res Methods* 3(1):4–70. <https://doi.org/10.1177/109442810031002>
56. von Davier M, Kennedy A, Reynolds K, Fishbein B, Khorramdel L, Aldrich C, Bookbinder A, Bezirhan U, Yin L (2024) TIMSS 2023 international results in mathematics and science. TIMSS & PIRLS International Study Center, Boston College. <https://www.iea.nl/publications/timss-2023-international-report>
57. Watt HMG, Shapka JD, Morris ZA, Durik AM, Keating DP, Eccles JS (2012) Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career

plans: A comparison of samples from Australia, Canada, and the United States. *Dev Psychol* 48(6):1594–1611. <https://doi.org/10.1037/a0027838>

58. White MA, Kern ML (2018) Positive education: Learning and teaching for wellbeing and academic mastery. *Int J Wellbeing* 8(1):1–17. <https://doi.org/10.5502/ijw.v8i1.588>

59. World Economic Forum (2024) *Global gender gap report 2024*. <https://www.weforum.org/publications/global-gender-gap-report-2024/>

Figures

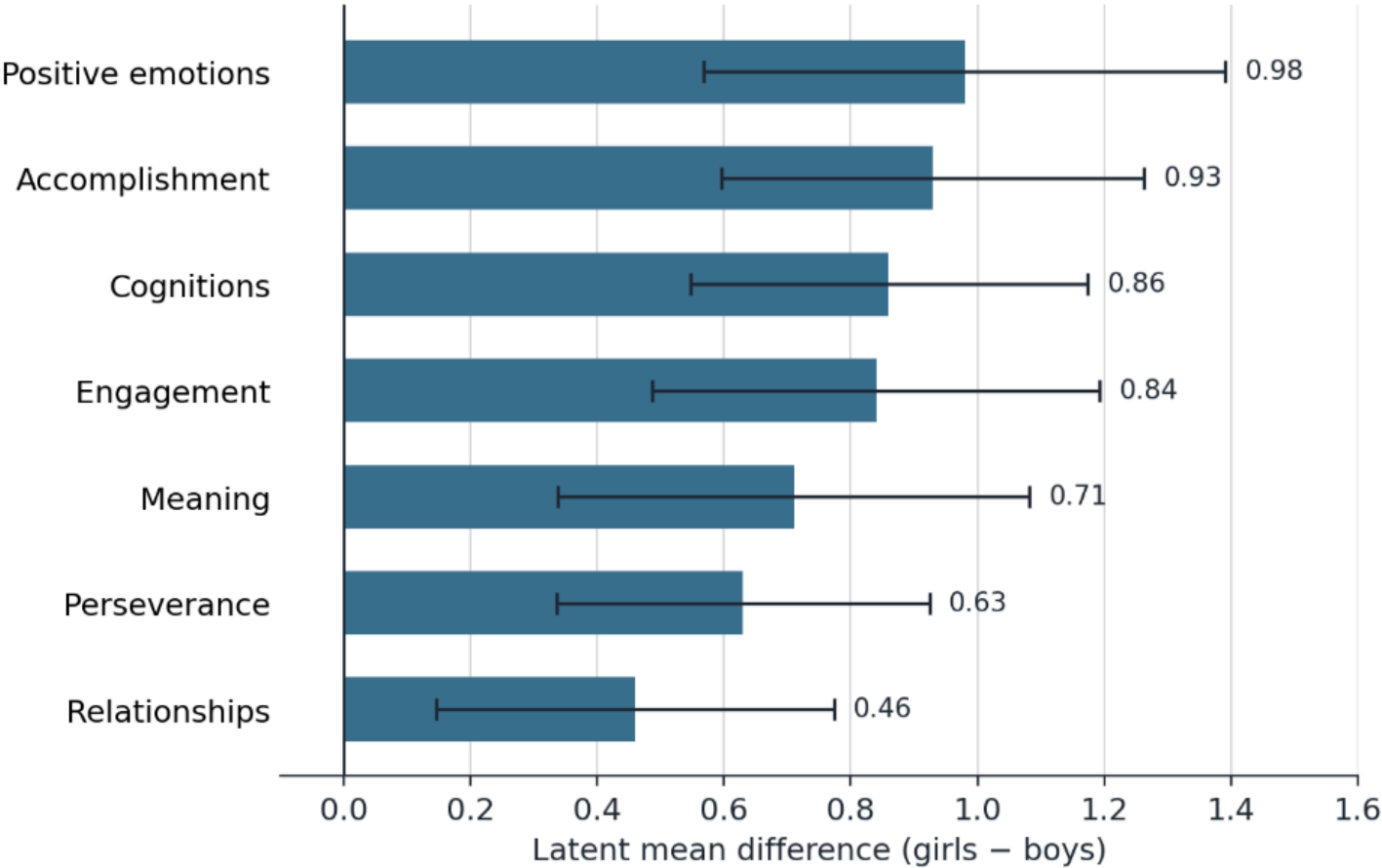


Figure 1

Latent Mean Differences Between Girls and Boys Across the Seven MWB Ultimate Values

Note. Bars represent the latent mean difference (girls minus boys), estimated with boys as the reference group. Error bars denote 95% confidence intervals. Every interval lies entirely above zero; these intervals do not account for the clustering of students within schools (see Section 8).

Supplementary Files

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

- [MWBIranSupplementaryMaterial26Oct26.docx](#)