Modern Finance

Modern Finance Institute

Article

Gender differences in financial knowledge calibration among German high school students

Justus Blaschke [1,*], **and Marco Pleßner** [2]

[1] University of Münster, Institute of Medical Informatics, Albert-Schweitzer-Campus 1, Building A11, 48149 Münster; justus.blaschke@uni-muenster.de.
[2] Hochschule Hamm-Lippstadt; marco.plessner@hshl.de
* Correspondence: University of Münster, Institute of Medical Informatics, Albert-Schweitzer-Campus 1, Building A11, 48149 Münster; justus.blaschke@uni-muenster.de.

Abstract: Financial education should not only increase knowledge but also help students assess what they know. This study examines gender differences in the calibration between objective financial knowledge and subjective self-assessment among 609 final-year German high school students. We estimate financial knowledge using item response theory, and measure overconfidence as the residual component of confidence after accounting for knowledge. Female students are, on average, more likely to be underconfident, whereas male students are more likely to be overconfident. The findings highlight calibration as an educationally relevant dimension of financial literacy.

Keywords: financial literacy; financial knowledge calibration; gender differences; financial education; overconfidence

Citation: Blaschke, J., and Pleßner, M. (2026). Gender differences in financial knowledge calibration among German high school students. *Modern Finance*, *4*(3),

Accepting Editor: Adam Zaremba

Received: 7 July 2026
Accepted: 12 September 2026
Published: 18 September 2026

1. Introduction

Financial choices have become increasingly complex over the past few decades. Private pension planning, low interest-rate environments, and the low-threshold availability of online investment products mean that young people will encounter consequential financial decisions soon after leaving school. Educational policy often responds to these challenges by expanding financial knowledge instruction. Yet financial education is not only about what students know; it is also about whether they can recognize the limits of their knowledge. Overconfident students may act on insufficient knowledge, whereas underconfident students may avoid financial topics or fail to use knowledge they already possess. The calibration between objective financial knowledge and subjective self-assessment is therefore an educationally relevant dimension of financial literacy.

This paper investigates gender differences in financial knowledge calibration among German high school students. It uses the same survey and data preparation as Blaschke (2022), which examined whether confidence can help explain gender differences in financial literacy among teenagers. In contrast, the present paper asks whether students' subjective assessments are systematically biased relative to their measured financial knowledge. The analysis applies the 13-item financial literacy scale of Lusardi and Mitchell (2017), administered in the exact German translation used by Erner et al. (2016), to 609 final-year students from the three main German secondary-school tracks. Financial knowledge is estimated using item response theory (IRT), and overconfidence is measured as the residual component of subjective performance assessment after accounting for measured knowledge.

The results show that female students are, on average, underconfident relative to their IRT-estimated financial knowledge, whereas male students are, on average, more overconfident. Conditional on cognitive reflection, grades, school type, and socioeconomic indicators, female students have a substantially lower probability of being classified as overconfident. This classification result is robust to an alternative raw

overconfidence measure and to excluding items flagged in the differential-item-functioning screen. The continuous calibration coefficient is more sensitive to the item exclusion, which is discussed as a measurement limitation. Additional analyses show that female students select 'I don't know' responses more frequently, consistent with the interpretation that response behavior and calibration are related.

These findings contribute to the related literature in three ways. First, the paper adds to research distinguishing objective financial knowledge, subjective financial knowledge, and confidence (Allgood & Walstad, 2016; Bucher-Koenen et al., 2025; Lawrence et al., 2024) by focusing on the calibration between knowledge and self-assessment rather than on either construct separately. This distinction matters because lower confidence does not necessarily imply underconfidence; it becomes a biased self-assessment only when it is low relative to measured knowledge. Second, the paper extends evidence on gender differences in financial literacy and financial confidence from predominantly adult populations to adolescents (Bannier & Schwarz, 2018; Bucher-Koenen et al., 2025; Lawrence et al., 2024). Evidence from students is important because gendered self-assessments may emerge before many adult financial decisions are made (Driva et al., 2016; Erner et al., 2016; Panos & Wright, 2025). Third, the paper applies an IRT-based measurement strategy used in financial knowledge and overconfidence research (Ooi, 2020; Piehlmaier, 2022) to study gendered calibration among German adolescents. The contribution is therefore empirical and educational rather than methodological.

The remainder of the study is structured as follows. Section 2 summarizes literature on subjective financial knowledge, confidence, and overconfidence. Section 3 defines the constructs and describes the IRT-based measurement strategy. Section 4 explains the study design, sample, descriptive results, and 'I don't know' responses. Section 5 presents the main results and robustness checks. Section 6 concludes.

2. Subjective financial knowledge, calibration, and gender

In financial literacy research, confidence is usually conceptualized as subjective financial knowledge: the individual's assessment of his or her own knowledge about financial matters. Subjective financial knowledge is analytically distinct from objective knowledge because individuals may accurately assess their knowledge, underestimate it, or overestimate it. Recent work has renewed attention to this distinction by showing that confidence and 'I don't know' response behavior are not merely survey artifacts but are related to measured financial literacy and financial choices (Bucher-Koenen et al., 2025; Cziriak et al., 2024; Tranfaglia et al., 2024).

The central construct in this paper is financial overconfidence, defined as a calibration error between subjective confidence and objective knowledge. A positive calibration error indicates that a person reports higher confidence than would be expected based on objective performance; a negative calibration error indicates underconfidence. This definition follows the broader judgment and decision-making literature on the confidence-accuracy relationship (Moore & Healy, 2008; Oskamp, 1965; Lichtenstein & Fischhoff, 1977; Fischhoff et al., 1977). Importantly, the construct is not defined as confidence causing knowledge. Instead, it measures whether self-assessment is high or low relative to estimated knowledge.

Prior financial literacy studies indicate that women are more likely than men to choose 'I don't know' options and less likely to guess on knowledge questions (Baldiga, 2014; Bucher-Koenen et al., 2025; Lusardi & Mitchell, 2011). Recent work on question design confirms that the availability and use of 'don't know' responses affect measured financial literacy gaps (Cziriak et al., 2024; Tranfaglia et al., 2024). These findings are particularly relevant for calibration because 'don't know' responses can reflect a lack of knowledge, low confidence, risk aversion in test-taking, or some combination of these mechanisms.

Gender stereotypes may also shape self-assessments. Research in psychology and economics shows that beliefs about gender-typical skills influence confidence in domains

such as mathematics, investing, and financial decision-making (Beyer, 1990; Beyer & Bowden, 1997; Bordalo et al., 2019). Since financial literacy contains numerical and investment-related tasks, it may be perceived as a more male-typed domain, which could increase male confidence and reduce female confidence even among students with comparable school performance (Driva et al., 2016; Forgasz et al., 2004).

Numerical ability remains an important component of financial literacy, but it is not the only component. Prior studies link numeracy, cognitive reflection, financial education, and financial self-beliefs to financial literacy and financial behavior (Frederick, 2005; Skagerlund et al., 2018; Tang, 2021). Recent evidence from Scottish students also suggests that gender gaps in financial understanding can emerge already in school and may be related not only to exposure to financial education but also to differences in how effectively such education translates into measured understanding (Panos & Wright, 2025). The present study therefore focuses on adolescents and asks whether gendered calibration patterns are already observable at the end of secondary school.

Most evidence on subjective financial knowledge and overconfidence comes from adult samples, but there are reasons to expect related patterns among adolescents. Gendered self-assessments in mathematics and stereotypically masculine domains emerge before adulthood (Beyer, 1990; Bench et al., 2015; Forgasz et al., 2004). Financial literacy also appears to be gendered early: German and international evidence documents gender gaps or gendered stereotypes among students and teenagers (Blaschke, 2022; Bottazzi & Lusardi, 2021; Driva et al., 2016; Erner et al., 2016; Panos & Wright, 2025). At the same time, adolescents may differ from adults because they have less direct experience with investment, credit, and household financial decisions. This makes the adolescent setting substantively important rather than a simple replication of adult evidence: it allows us to test whether gendered calibration patterns are already visible before many consequential adult financial decisions are made.

Hypothesis 1 (H1): Female students are, on average, underconfident in their financial literacy, whereas male students are, on average, overconfident in their financial literacy.

3. Methods

3.1. Construct definition

This study defines financial overconfidence as biased self-assessment: the part of stated confidence that remains after accounting for objective financial literacy. The construct is therefore a measure of calibration, not a causal antecedent of financial literacy. A positive value indicates that a student is more confident than expected given his or her estimated financial literacy; a negative value indicates that a student is less confident than expected.

The original version of the analysis also examined whether overconfidence predicts financial literacy. That specification is not retained because the overconfidence variable is itself constructed from financial literacy and stated confidence. The revised empirical approach treats overconfidence as the outcome to be explained. This is both theoretically clearer and aligned with the residual construction of the measure.

3.2. Financial literacy as an IRT latent trait

Financial literacy is a latent variable and cannot be observed directly. The questionnaire applied in this study is the 13-item financial literacy scale of Lusardi and Mitchell (2017), administered in the exact German translation used by Erner et al. (2016). The assessment includes five basic literacy questions on everyday financial topics and eight more sophisticated questions on investment-related topics. The survey also included the three-item Cognitive Reflection Test (CRT) by Frederick (2005), which measures the tendency to override intuitive but incorrect answers in simple reasoning tasks; a typical CRT item presents an arithmetic problem for which the immediately intuitive answer is

wrong. The CRT score is included because cognitive reflection has been linked to financial literacy and numeracy-related financial decision-making (Skagerlund et al., 2018).

A simple sum score treats all items as equally informative and does not account for differences in item difficulty or item discrimination. To address this limitation, financial literacy is estimated as a latent trait using a two-parameter logistic item response theory model. The probability that a student v answers item i correctly is modeled as a function of the student's latent financial literacy θ_v, the item difficulty b_i, and the item discrimination a_i:

$$P(x_{vi} = 1) = \frac{e^{a_i(\theta_v - b_i)}}{1+e^{a_i(\theta_v - b_i)}} \quad (1)$$

Item responses were coded as 1 for correct answers and 0 for incorrect answers. 'I don't know' responses were also coded as incorrect answers and were not treated as missing values. This follows the standard scoring logic of many financial literacy studies and reflects that respondents did not provide a correct answer. Nevertheless, this coding choice is not innocuous: if 'I don't know' responses partly reflect low confidence or unwillingness to guess rather than absence of knowledge, the resulting objective knowledge measure may also contain a response-style component. Because this concern is especially relevant when 'I don't know' responses differ by gender, the analysis reports these responses descriptively and includes an item-level assessment in Section 4.3.

3.3. IRT overconfidence

Overconfidence is operationalized as a residual calibration measure. After completing the financial literacy questions, participants stated how many of the 13 items they believed they had answered correctly. This ex post self-assessment is modeled as a flexible function of IRT-estimated financial literacy. The residual from this model captures whether a student is more or less confident than would be expected at his or her estimated knowledge level.

$$confidence_i = \beta_0 + \beta_1 IRT\ FinancialLiteracy_i + \beta_2 IRT\ FinancialLiteracy_i^2 + \beta_3 IRT\ FinancialLiteracy_i^3 + \beta_4 IRT\ FinancialLiteracy_i^4 + \varepsilon_i \quad (2)$$

The standardized residual ε_i is the IRT overconfidence measure. Values above zero indicate overconfidence; values below zero indicate underconfidence. The fourth-order polynomial is used to follow Piehlmaier (2022) and to avoid imposing a linear relationship between financial knowledge and stated confidence. A flexible polynomial is appropriate because calibration may be nonlinear: students at very low or very high knowledge levels may differ systematically in how they translate knowledge into self-assessed performance. The residual approach follows the logic of Grinblatt and Keloharju (2009) and Piehlmaier (2022), while the interpretation remains explicitly correlational and measurement-oriented rather than causal.

The main empirical model estimates gender differences in continuous IRT overconfidence using ordinary least squares with heteroskedasticity-robust standard errors. A binary probit model is retained as a complementary specification because it maps directly onto the substantive distinction between overconfidence and underconfidence. A raw difference score, defined as the self-assessed number of correct answers minus the actual number of correct answers, is used as an additional robustness check.

4. Data and descriptive results

4.1. Procedure and sample

Because the study surveyed minors through secondary education institutions, formal permission had to be obtained before data collection. The German school system is

governed by the federal states. Accordingly, formal requests to conduct the study were submitted to the responsible ministry of education or competent school authority in each state. Research in German schools may only be conducted after such approval. After approval, 25 secondary schools were randomly selected from the secondary schools in the respective state and invited to participate.

In total, 911 final-year high school students from eleven German states participated during a six-month survey period between January 2021 and July 2021. The participating states were Baden-Württemberg, Bavaria, Brandenburg, Bremen, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, Saxony-Anhalt, Schleswig-Holstein, and Thuringia. Participants came from graduating classes of the three main secondary-school tracks: Hauptschule, Realschule, and Gymnasium. The survey was administered in German, using the same German translation of the financial literacy questionnaire as Erner et al. (2016). No financial or performance-based incentives were provided.

Students received instructions before the start of the survey and had to confirm that they would not use calculators, mobile phones, or other forms of support. Due to the COVID-19 pandemic and partial school closures in Germany, some participants may have completed the online questionnaire at home rather than at school. To reduce misuse of the questionnaire, a general time limit was imposed, and teachers were asked to administer the survey during supervised lessons where possible. The timestamps of submitted questionnaires were largely clustered, suggesting that this procedure was mostly followed.

The data selection and preparation follow the same procedure as the published analysis based on this survey (Blaschke, 2022). The exclusion threshold was a complete-case rule: observations were removed if at least one variable required for the IRT estimation, confidence measure, or regression covariates was missing. Of the 911 total participants, 256 observations were excluded under this rule. A further 46 observations were excluded because responses were submitted outside typical school hours or on weekends. The final analytic sample therefore consists of 609 students. This is the same final analysis sample reported in Blaschke (2022), including the gender distribution of 313 female and 296 male students. Retaining the same data preparation avoids inconsistent sample definitions across publications using this survey. However, the large complete-case exclusion remains a limitation. Because the present revision intentionally retains the published analysis sample and the excluded incomplete records are not part of the anonymized analysis file used for this resubmission, missing-at-random patterns among the excluded observations cannot be tested directly here.

Table 1. Sample flow

Step	N
Original survey responses	911
Excluded because of missing values	256
Excluded because responses were outside typical school hours or on weekends	46
Final analytic sample	609

Note. The table reports the data selection used by Blaschke (2022) and retained here for consistency with the published data preparation. The final analytic sample is identical to that published sample (N = 609; female n = 313; male n = 296).

The distribution of students across school types is as follows: Gymnasium n = 323 (53%), Realschule n = 207 (34%), and Hauptschule or extended Hauptschule n = 79 (13%). This distribution is broadly consistent with the overall distribution of students across German secondary-school tracks reported by the German Ministry of Education and Research (BMBF 2019).

Table 2 shows that female and male students have very similar mathematics grades. Female students have higher German and English grades and are more likely to attend

Gymnasium, while male students have a higher mean CRT score. These descriptive differences motivate the inclusion of grades, cognitive reflection, and school type as controls in the regression models.

4.2. Descriptive IRT results

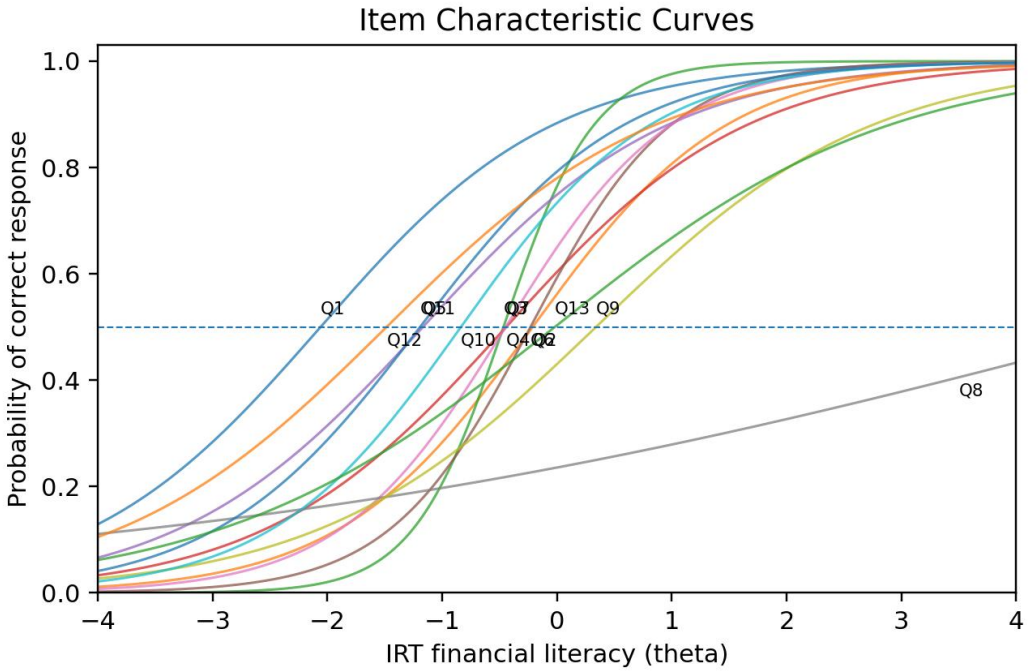
The item characteristic curves in Figure 1 illustrate the estimated 2PL model. Easier items cross the 0.5 probability line at lower values of theta, whereas more difficult items require higher values of estimated financial literacy. Q1, which measures numeracy, is the easiest item, while Q8, which asks about the link between interest rates and bond prices, is by far the most difficult item. Q3, on inflation, has the highest discrimination parameter and is therefore particularly informative for distinguishing students with lower and higher financial literacy. The full item parameters are reported in Appendix Table A1.

Table 2. Summary statistics by gender

Variable	Description	Female Mean	Female SD	Male Mean	Male SD
CRT score	Correct responses to the three CRT questions of Frederick (2005)	1.34*	1.08	1.55*	1.15
Under 16 years	Dummy variable (1 = under 16 years, 0 = above 16 years)	0.31		0.31	
Risk association	Dummy variable (1 = risk associated with threat, 0 = otherwise)	0.73***		0.50***	
Side job	Dummy variable (1 = job, 0 = no job)	0.38		0.41	
Saving	Dummy variable (1 = saving, 0 = no saving)	0.83		0.84	
Household income	Dummy variable (1 = own real estate, 0 = rent)	0.71		0.71	
Integration	Dummy variable (1 = parents born in Germany, 0 = parents born elsewhere)	0.80		0.77	
Club membership	Dummy variable (1 = membership, 0 = no membership)	0.59		0.62	
Mathematics grade	Grade in math from the last report card (1 = fail, 6 = excellent)	4.39	1.18	4.42	1.13
German grade	Grade in German from the last report card (1 = fail, 6 = excellent)	4.73***	0.91	4.23***	0.98
English grade	Grade in English from the last report card (1 = fail, 6 = excellent)	4.73***	1.02	4.38***	1.03
Gymnasium	Attended high school type (1 = Gymnasium, 0 = other)	0.58*		0.48*	
Realschule	Attended high school type (1 = Realschule, 0 = other)	0.32		0.36	
Hauptschule	Attended high school type (1 = Hauptschule, 0 = other)	0.10*		0.16*	

Note. Female n = 313 and male n = 296. The order of grade variables was reversed so that higher values indicate better grades. Differences in group means were assessed with Welch's unpaired t-

test for continuous variables and Fisher's exact test for dichotomous variables. *** p < .001, ** p < .01, * p < .05.



Note. Figure 1. Item characteristic curves of the 2PL model. The x-axis shows the latent financial literacy trait, and the y-axis shows the probability of a correct response.

Table 3. IRT financial literacy and IRT overconfidence by gender

Variable	Female Mean	Female SD	Female Min	Female Max	Male Mean	Male SD	Male Min	Male Max
IRT financial literacy	-0.221*** (-5.305)	0.955	-2.311	1.775	0.237	0.995	-2.311	1.775
IRT overconfidence	-0.186*** (-4.836)	0.940	-2.717	2.838	0.200	1.025	-3.403	3.384

Note. Both variables were standardized to increase interpretability. Differences in group means were assessed with Welch's unpaired t-test, with t-statistics shown in parentheses. *** p < .001, ** p < .01, * p < .05.

Table 3 shows two central descriptive patterns. First, female students have lower average IRT financial literacy than male students. Second, the mean IRT overconfidence value is negative for female students and positive for male students. In substantive terms, female students are underconfident relative to their estimated financial knowledge, whereas male students are overconfident on average. This pattern is consistent with H1.

4.3. 'I don't know' responses

Because 'I don't know' responses are theoretically linked to confidence and willingness to answer, they are examined separately. Female students choose 'I don't know' more often than male students: 1.85 times on average compared with 1.01 times for male students (p < .001). The difference remains positive in a control model that includes IRT financial literacy, cognitive reflection, grades, school type, and socioeconomic controls (b = 0.275, p = .034). This pattern supports the view that gender differences in measured financial literacy partly reflect response behavior in addition to knowledge.

Table 4. 'I don't know' responses by item and gender

Item	Topic	Female DK %	Male DK %	Difference pp	p
Q1	Numeracy	4.2	2.4	1.8	0.259
Q2	Compound interest	10.5	1.7	8.9	< .001***
Q3	Inflation	27.2	13.5	13.6	< .001***
Q4	Inflation/money illusion	6.7	4.1	2.7	0.157
Q5	Time value of money	3.2	2.4	0.8	0.626
Q6	Mutual funds	20.8	14.2	6.6	0.034*
Q7	Company stock vs. mutual funds	25.9	14.5	11.4	0.001***
Q8	Interest rate/bond price link	25.6	18.9	6.6	0.052
Q9	Long-period returns	23.0	9.5	13.5	< .001***
Q10	Stocks vs. bonds	6.4	2.0	4.4	0.009**
Q11	Fluctuation/volatility	11.2	4.7	6.5	0.004**
Q12	Stock market functioning	14.4	8.1	6.3	0.015*
Q13	Risk vs. return	5.8	4.7	1.0	0.591

Note. DK = 'I don't know'. Percentages refer to the share of students in each gender group choosing the DK option for the item. p-values are based on Fisher's exact tests. *** p < .001, ** p < .01, * p < .05.

The item-level pattern shows that the DK gap is not concentrated in a single question. Female students are more likely to select DK on most items, with especially large differences for compound interest, inflation, company stock versus mutual funds, and long-period returns. Since DK responses are coded as incorrect in the IRT model, the objective knowledge estimate should be understood as performance under a questionnaire design that permits DK responses, not as a pure measure of knowledge independent of confidence.

5. Results

5.1. Gender differences in financial knowledge calibration

The main empirical question is whether the descriptive gender difference in calibration remains after accounting for cognitive reflection, grades, school type, and socioeconomic indicators. Table 5 reports three specifications. Column (1) uses continuous IRT overconfidence as the dependent variable and is the preferred specification because it preserves the full variation in the calibration measure. Column (2) reports a complementary probit model in which the dependent variable equals 1 when IRT overconfidence is non-negative. Column (3) reports the robustness model using raw overconfidence.

The continuous specification shows a negative and statistically significant female coefficient (b = -0.348, SE = 0.089, p < .001). The magnitude corresponds to roughly one-third of a standard deviation of the IRT overconfidence measure after controlling for the included covariates. The probit model leads to the same conclusion: the female coefficient is negative and highly significant. The approximate average marginal effect indicates that female students have about a 17 percentage-point lower probability of being classified as overconfident. The raw overconfidence model also produces a negative coefficient. The explained variance in the continuous model is modest (R2 = .060), indicating that gender and the included controls explain only a limited share of individual heterogeneity in calibration. The results should therefore be interpreted as evidence of a systematic average gender difference, not as a comprehensive model of all determinants of financial overconfidence.

The control variables are less consistently related to calibration. Club membership is positively related to the probability of being overconfident in the binary specification,

while mathematics grade is positively related to continuous IRT overconfidence. These coefficients should be interpreted cautiously because the primary purpose of the control set is to adjust the gender comparison, not to identify causal determinants of overconfidence.

Table 5. Gender differences in financial overconfidence

Variable	(1) Continuous IRT overconfidence	(2) Binary IRT overconfidence (probit)	(3) Raw overconfidence
Constant	-0.108		0.216
	(0.258)		(0.745)
Female	-0.348***	-0.488***	-0.492*
	(0.089)	(0.110)	(0.251)
CRT score	0.019	-0.020	-0.160
	(0.041)	(0.053)	(0.118)
Under 16 years	0.044	0.046	0.222
	(0.114)	(0.131)	(0.329)
Risk association	-0.063	-0.085	-0.075
	(0.087)	(0.057)	(0.249)
Side job	-0.071	-0.135	-0.132
	(0.086)	(0.111)	(0.243)
Saving	0.088	0.045	-0.043
	(0.108)	(0.145)	(0.309)
Household income	0.012	0.008	-0.057
	(0.094)	(0.127)	(0.267)
Integration	-0.049	0.247	-0.283
	(0.110)	(0.142)	(0.316)
Club membership	0.121	0.258*	0.280
	(0.089)	(0.115)	(0.251)
Mathematics grade	0.088*	0.100	0.196
	(0.038)	(0.051)	(0.108)
German grade	0.034	-0.001	0.031
	(0.049)	(0.065)	(0.140)
English grade	-0.066	-0.028	-0.194
	(0.046)	(0.060)	(0.132)
Gymnasium	-0.062	-0.004	-0.846
	(0.176)	(0.208)	(0.520)
Realschule	-0.016	0.037	-0.365
	(0.166)	(0.192)	(0.494)
Observations	609	609	609
R-squared	0.060		0.050
Approx. marginal effect: Female		-0.173	

Note. Robust standard errors are shown in parentheses. Column (2) is a probit model and uses the binary indicator 1 = overconfident and 0 = underconfident. Hauptschule is the reference category for school type. The marginal effect is an approximate average discrete change in predicted probability for the female indicator based on the original data preparation. *** p < .001, ** p < .01, * p < .05.`

5.2. Robustness and measurement checks

Table 6 reports model variants that use continuous IRT overconfidence as the outcome and vary the control set. The female coefficient is stable across specifications: it is negative and statistically significant in a gender-only model, after adding school type,

in the full model without school type, and in the full model. This suggests that the main finding is not an artifact of a particular control set.

Table 6. Robustness across model specifications

Model	Female coefficient	Robust SE	N	R-squared
Gender only	-0.388***	(0.080)	609	0.038
Gender + school type	-0.386***	(0.080)	609	0.038
Full model without school type	-0.353***	(0.088)	609	0.059
Full model	-0.348***	(0.089)	609	0.060

Note. Outcome is continuous IRT overconfidence. Robust standard errors are reported. *** $p < .001$, ** $p < .01$, * $p < .05$.

A further measurement check compares the IRT-based residual measure with a raw difference score based on the self-assessed and actual number of correct answers. The two measures are strongly correlated ($r = 0.721$), indicating convergent validity. Because a correlation is not a full robustness analysis, the raw overconfidence model in Table 5 re-estimates the gender effect with the alternative measure. The gender coefficient remains negative and statistically significant.

Appendix Table A2 provides a compact differential-item-functioning screen by estimating item-level logit models that include theta, female, and their interaction. This diagnostic flags two items, Q5 (time value of money) and Q6 (mutual funds), through significant female-by-theta interactions. These items may function differently across gender at different levels of estimated financial literacy, possibly because they combine financial concepts with contexts that differ in familiarity among adolescents. To assess whether the main finding depends on these items, Table 7 re-estimates the IRT literacy measure after excluding Q5 and Q6 and then reconstructs the calibration measure. The binary overconfidence result remains strong: female students are again significantly less likely to be classified as overconfident. The continuous residual coefficient is still negative but smaller and no longer statistically significant. The results therefore support the qualitative classification result, while also showing that the magnitude of the continuous calibration gap is sensitive to the two DIF-flagged items. For this reason, the conclusions are phrased conservatively, and the DIF screen is treated as an important measurement limitation.

6. Conclusion

This study examines gender differences in the calibration of subjective and objective financial knowledge among German high school students. Building on the same survey and data preparation used in Blaschke (2022), the paper asks a distinct question: whether female and male students differ in the accuracy of their self-assessments after measured financial knowledge is taken into account. Overconfidence is defined as a residual-based calibration measure: stated confidence that is higher or lower than expected from IRT-estimated financial literacy. This definition follows the logic of confidence-accuracy research and avoids interpreting confidence as a direct cause of knowledge.

The results show that female students are, on average, underconfident relative to their estimated financial knowledge, whereas male students are, on average, more overconfident. The gender difference is statistically significant and moderate in size. It is present in descriptive comparisons, in the main continuous model, in a binary probit model, and in a raw difference-score robustness check. The findings therefore add evidence from German adolescents to the broader literature on gendered self-assessments in financial domains.

Table 7. Robustness to excluding the DIF-flagged items Q5 and Q6

Specification	Outcome	Female coeff.	SE	p	N
Full 13-item IRT	Continuous IRT overconfidence	-0.348***	(0.089)	< .001	609
11-item IRT excluding Q5/Q6	Continuous IRT overconfidence	-0.110	(0.079)	.164	609
Full 13-item IRT	Binary IRT overconfidence (probit)	-0.488***	(0.110)	< .001	609
11-item IRT excluding Q5/Q6	Binary IRT overconfidence (probit)	-0.591***	(0.117)	< .001	609

Note. The 11-item specification re-estimates the IRT financial literacy score without Q5 and Q6 and reconstructs the residual overconfidence measure using the same fourth-order polynomial logic. Robust standard errors are shown for OLS models. The probit coefficient is shown for the binary overconfidence indicator. *** p < .001, ** p < .01, * p < .05.

The practical implication is not that confidence alone produces financial literacy. Rather, financial education may benefit from calibrated feedback. Students should learn financial concepts and should also receive opportunities to compare confidence with actual performance. Such feedback can be integrated into classroom-based financial education through low-stakes quizzes, prediction-before-feedback tasks, and reflection exercises after problem-solving. These approaches are educationally relevant because underconfidence may reduce engagement with financial topics, while overconfidence may lead students to overestimate their readiness for financial decisions.

Several limitations remain. First, the sample is restricted to German final-year secondary-school students and should not be generalized without replication. Second, the survey was conducted during the COVID-19 pandemic, and some students may have completed the questionnaire at home despite the supervision instructions. Third, the use of the same complete-case dataset as Blaschke (2022) ensures consistency with the published N = 609 analysis sample and gender distribution, but it limits the ability to examine the 256 observations excluded for missing values. Fourth, 'I don't know' responses were coded as incorrect answers; the additional analyses show that these responses are gendered and should be interpreted as part of the measurement context. Fifth, the DIF screen indicates that Q5 and Q6 may function differently by gender. The robustness check excluding these items supports the binary classification result but attenuates the continuous coefficient, suggesting that future studies should conduct fuller measurement-invariance analyses and, ideally, elicit item-level confidence.

Overall, the study suggests that gender differences in financial literacy are linked not only to knowledge levels but also to biased self-assessments. By focusing on financial knowledge calibration in a German adolescent sample, and by clearly distinguishing this question from prior work on objective financial literacy levels in the same survey, the paper makes a focused empirical contribution to research on subjective and objective financial knowledge.

Supplementary Materials: No supplementary materials are available.

Author Contributions: Conceptualization, J.B..; methodology, J.B. and M.P.; software, J.B.; validation, J.B. and M.P..; formal analysis, J.B.; investigation, J.B.; resources, J.B.; data curation, J.B.; writing—original draft preparation, J.B and M.P..; writing—review and editing, J.B. and M.P,.;

visualization, J.B.; supervision, J.B. and M.P..; project administration, J.B.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data are not publicly available due to the involvement of minors and restrictions related to school-based data collection. The analysis code can be made available from the corresponding author upon reasonable request.

Acknowledgments: We thank the participating schools, teachers, and students for supporting the data collection. We also thank the competent ministries of education and school authorities in the participating German federal states for granting permission to conduct the study.

Conflicts of Interest: The authors declare no conflict of interest

AI Use Statement: The authors used ChatGPT (OpenAI) for grammar and language refinement. All content was carefully reviewed and verified by the authors.

Table A1. Results of the 2PL model on IRT financial literacy

Item	Item topic	Difficulty	Discrimination
Q1	Numeracy	-2.060***	0.984***
Q2	Compound interest	-0.212***	1.180***
Q3	Inflation	-0.466***	2.527***
Q4	Inflation/money illusion	-0.443***	0.950***
Q5	Time value of money	-1.169***	0.935***
Q6	Mutual funds	-0.231***	1.628***
Q7	Company stock vs. mutual funds	-0.447***	1.384***
Q8	Interest rate/bond price link	5.190***	0.227***
Q9	Long-period returns	0.339***	0.827***
Q10	Stocks vs. bonds	-0.838***	1.210***
Q11	Fluctuation/volatility	-1.193***	1.123***
Q12	Stock market functioning	-1.485***	0.852***
Q13	Risk vs. return	-0.020***	0.684***

Note. Items with negative difficulty values are easier than items with positive difficulty values. Discrimination describes how well the item distinguishes students with lower and higher latent financial literacy. *** p < .001, ** p < .01, * p < .05.

Table A2. Compact item-level gender DIF screen

Item	Female coefficient	p	Female × theta coef.	p
Q1 Numeracy	-0.242	0.445	-0.231	0.450
Q2 CompoundInterest	-0.295	0.136	-0.389	0.114
Q3 Inflation	0.761	0.058	0.930	0.153
Q4 Inflation simp	0.125	0.515	0.061	0.780
Q5 TVoM	0.141	0.529	-0.789**	0.002
Q6 MutualFund	0.430	0.062	-0.791*	0.019
Q7 CompvsMut	-0.398	0.069	-0.238	0.395
Q8 IRBondprice	0.128	0.517	0.134	0.506
Q9 LTReturns	-0.313	0.088	0.096	0.643
Q10 Risk	-0.740**	0.001	0.352	0.185
Q11 Vola	-0.066	0.791	-0.187	0.489
Q12 StockMarketFunc	0.004	0.987	0.163	0.487
Q13 RiskvsReturn	-0.098	0.581	0.137	0.476

Note. Each row reports a compact item-level logit diagnostic with theta, female, and the theta-by-female interaction. This is a screening diagnostic rather than a full measurement-invariance test. *** p < .001, ** p < .01, * p < .05.

References

Allgood, S., & Walstad, W. B. (2016). The effects of perceived and actual financial literacy on financial behaviors. *Economic Inquiry, 54*(1), 675–697. https://doi.org/10.1111/ecin.12255

Balasubramnian, B., & Sargent, C. S. (2020). Impact of inflated perceptions of financial literacy on financial decision-making. *Journal of Economic Psychology, 80*, 102306. https://doi.org/10.1016/j.joep.2020.102306

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science, 60*(2), 434–448. https://doi.org/10.1287/mnsc.2013.1776

Bannier, C. E., & Schwarz, M. (2018). Gender- and education-related effects of financial literacy and confidence on financial wealth. *Journal of Economic Psychology, 67*, 66–86. https://doi.org/10.1016/j.joep.2018.05.005

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics, 116*(1), 261–292. https://doi.org/10.1162/003355301556400

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles, 72*(11–12), 536–546. https://doi.org/10.1007/s11199-015-0486-9

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960–970. https://doi.org/10.1037/0022-3514.59.5.960

Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin, 23*(2), 157–172. https://doi.org/10.1177/0146167297232005

Bhandari, G., & Deaves, R. (2006). The demographics of overconfidence. *Journal of Behavioral Finance, 7*(1), 5–11. https://doi.org/10.1207/s15427579jpfm0701_2

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Addison-Wesley.

Blaschke, J. (2022). Gender differences in financial literacy among teenagers—Can confidence bridge the gap? *Cogent Economics & Finance, 10*(1), 2144328. https://doi.org/10.1080/23322039.2022.2144328

Bundesministerium für Bildung und Forschung. (2019). *Bildung und Forschung in Zahlen 2021: Schülerinnen und Schüler im 8. Schuljahr nach Schularten (1952–2019).* https://www.datenportal.bmbf.de/portal/de/B23.html

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review, 109*(3), 739–773. https://doi.org/10.1257/aer.20170007

Bottazzi, L., & Lusardi, A. (2021). Stereotypes in financial literacy: Evidence from PISA. *Journal of Corporate Finance, 71*, 101831. https://doi.org/10.1016/j.jcorpfin.2020.101831

Bucher-Koenen, T., Alessie, R., Lusardi, A., & van Rooij, M. (2025). Fearless woman: Financial literacy, confidence, and stock market participation. *Management Science, 71*(9), 7414–7430. https://doi.org/10.1287/mnsc.2023.00425

Bucher-Koenen, T., & Lusardi, A. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance, 10*(4), 565–584. https://doi.org/10.1017/S1474747211000485

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6). https://doi.org/10.18637/jss.v048.i06

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cziriak, M., Bucher-Koenen, T., & Alessie, R. J. M. (2024). *Beyond knowledge: Confidence and the gender gap in financial literacy* (ZEW Discussion Paper No. 24-083). https://doi.org/10.2139/ssrn.5103917

Dahlbom, L., Jakobsson, A., Jakobsson, N., & Kotsadam, A. (2011). Gender and overconfidence: Are girls really overconfident? *Applied Economics Letters, 18*(4), 325–327. https://doi.org/10.1080/13504851003670668

Driva, A., Luehrmann, M., & Winter, J. (2016). Gender differences and stereotypes in financial literacy: Off to an early start. *Economics Letters, 146*, 143–146. https://doi.org/10.1016/j.econlet.2016.07.029

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers, 56*(2), 209–230. https://doi.org/10.1093/oep/gpf048

Erner, C., Goedde-Menke, M., & Oberste, M. (2016). Financial literacy of high school students: Evidence from Germany. *Journal of Economic Education, 47*(2), 95–105. https://doi.org/10.1080/00220485.2016.1146102

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Forgasz, H. J., Leder, G. C., & Kloosterman, P. (2004). New perspectives on the gender stereotyping of mathematics. *Mathematical Thinking and Learning, 6*(4), 389–420. https://doi.org/10.1207/s15327833mtl0604_2

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Geiser, C., & Eid, M. (2010). Item-Response-Theorie. In C. Wolf et al. (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 311–332). VS Verlag.

Hadar, L., Sood, S., & Fox, C. R. (2013). Subjective knowledge in consumer financial decisions. *Journal of Marketing Research, 50*(3), 303–316. https://doi.org/10.1509/jmr.10.0518

Haag, L., & Brahm, T. (2025). The gender gap in economic and financial literacy: A review and research agenda. *International Journal of Consumer Studies, 49*(2), e70031. https://doi.org/10.1111/ijcs.70031

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Netherlands.

Jakobsson, N., Levin, M., & Kotsadam, A. (2013). Gender and overconfidence: Effects of context, gendered stereotypes, and peer group. *Advances in Applied Sociology, 3*(2), 137–141. https://doi.org/10.4236/aasoci.2013.32018

Kurz, C. J., Li, G., & Vine, D. J. (2018). Are millennials different? *Finance and Economics Discussion Series, 2018-080*. https://doi.org/10.1016/B978-0-12-813524-2.00008-1

Lawrence, E. R., Nguyen, T. D., & Wick, B. (2024). Gender differences in overconfidence and household financial literacy. *Journal of Banking & Finance, 166*, 107237. https://doi.org/10.1016/j.jbankfin.2024.107237

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

Luehrmann, M., Serra-Garcia, M., & Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking & Finance, 54*, 160–174. https://doi.org/10.1016/j.jbankfin.2014.11.009

Lusardi, A., & Mitchell, O. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics and Finance, 10*(4), 497–508. https://doi.org/10.1017/S1474747211000448

Lusardi, A., & Mitchell, O. (2017). How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness. *Quarterly Journal of Finance, 7*(3), 1750008. https://doi.org/10.1142/S2010139217500082

Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement, 13*(1), 57–75. https://doi.org/10.1177/014662168901300106

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Oberrauch, L. (2024). Assessing financial literacy among the young. *Journal of Financial Literacy and Wellbeing, 2*(1), 63–78. https://doi.org/10.1017/flw.2024.17

OECD. (2022). *Long-term interest rates 1981–2021*. OECD Data.

OECD. (2024). *PISA 2022 results (Volume IV): How financially smart are students?* https://doi.org/10.1787/5a849c2a-en

Ooi, E. (2020). Give mind to the gap: Measuring gender differences in financial knowledge. *Journal of Consumer Affairs, 54*(3), 931–950. https://doi.org/10.1111/joca.12310

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*(3), 261–265. https://doi.org/10.1037/h0022125

Panos, G. A., & Wright, R. E. (2025). Financial teaching and financial understanding amongst young people. *Economics Letters, 251*, 112309. https://doi.org/10.1016/j.econlet.2025.112309

Piehlmaier, D. M. (2022). Overconfidence and the adoption of robo-advice: Why overconfident investors drive the expansion of automated financial advice. *Financial Innovation, 8*(1). https://doi.org/10.1186/s40854-021-00324-3

Rowlingson, K. (2002). Private pension planning: The rhetoric of responsibility, the reality of insecurity. *Journal of Social Policy, 31*(4), 623–642. https://doi.org/10.1017/S0047279402006773

Serido, J., Shim, S., & Tang, C. (2013). A developmental model of financial capability: A framework for promoting a successful transition to adulthood. *International Journal of Behavioral Development, 37*(4), 287–297. https://doi.org/10.1177/0165025413479476

Skagerlund, K., Lind, T., Stroembeck, C., Tinghoeg, G., & Vaestfjaell, D. (2018). Financial literacy and the role of numeracy. *Journal of Behavioral and Experimental Economics, 74*, 18–25. https://doi.org/10.1016/j.socec.2018.03.004

Tang, N. (2021). Cognitive abilities, self-efficacy, and financial behavior. *Journal of Economic Psychology, 87*, 102447. https://doi.org/10.1016/j.joep.2021.102447

Tranfaglia, A., Lloro, A., & Merry, E. (2024). Question design and the gender gap in financial literacy. *FEDS Notes*. Board of Governors of the Federal Reserve System. https://doi.org/10.17016/2380-7172.3415

van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics, 101*(2), 449–472. https://doi.org/10.1016/j.jfineco.2011.03.006

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*(4), 817–838. https://doi.org/10.2307/1912934

Winkelmann, R. (1996). Employment prospects and skill acquisition of apprenticeship-trained workers in Germany. *Industrial and Labor Relations Review, 49*(4), 658–672. https://doi.org/10.1177/001979399604900405

Xiao, J. J., Ahn, S. Y., Serido, J., & Shim, S. (2014). Earlier financial literacy and later financial behavior of college students. *International Journal of Consumer Studies, 38*(6), 593–601. https://doi.org/10.1111/ijcs.12122