Working Paper No. 206

A, B, or C? Question Format and the Gender Gap in Financial Literacy

Maddalena Davoli

Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

UNIVERSITÄT BERN

Working Paper No. 206

A, B, or C? Question Format and the Gender Gap in Financial Literacy

Maddalena Davoli

May 2023

Please only cite after contacting the authors for a most recent version.

Please cite as:
"A, B, or C? Question Format and the Gender Gap in Financial Literacy." Swiss Leading House "Economics of Education" Working Paper No. 206, 2023. By Maddalena Davoli.

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

A, B, or C? Question Format and the Gender Gap in Financial Literacy *

Maddalena Davoli†

May 2023
Please only cite after contacting the authors for a most recent version.

Abstract

Financial literacy surveys, composed primarily of multiple-choice questions, consistently show women having lower financial knowledge than men. The education science literature finds that gender bias is inherent in multiple-choice testing. Using data from PISA 2015, this paper investigates the differential gender effect of question formats on students' financial literacy assessments. This paper, employing data from PISA 2015, analyzes the differential gender effect of question formats on students' financial literacy assessment. Having answers to both multiple-choice and open-response questions for each student, we employ a panel specification and use within-student variation while controlling for students' fixed characteristics. Findings show female students performing worse when answering multiple-choice questions, with no observable difference for the open-response format. Robustness tests indicate that the question characteristics underlying the multiple-choice format partly drive the results. I show how school policies aimed at training students for the multiple-choice format may help close the gender gap.

Keywords: *gender gaps, financial literacy, PISA, question format, within-student estimate*

JEL Classification: *I21, I24, J16, G53, D91*

*I am grateful to Michalis Haliassos, Olga Goldfayn, Simon Hess, Christin Siegfried and Jorge Quintana for valuable comments. The paper benefited from valuable feedbacks from participants at LEER 2022, LESE Conference 2022, IWAEE 2022, the Financial Literacy session of the 62^{st} Meeting of the Italian Economic Society, the Invalsi Annual Meeting and the Household Finance Reading Group (Goethe University). Any errors remain my own.

†University of Zuricht, Plattenstrasse 14, 8001, Zurich, Switzerland. maddalena.davoli@business.uzh.ch

1 Introduction

Despite women's and men's convergence in many economic outcomes, women consistently display lower levels of financial literacy than men, a finding widespread across many countries and contexts (Hasler and Lusardi (2017)). Given that financial literacy is a driver of financial inclusion (Grohmann et al. (2018)) and savvier financial behaviors[1] (Rooij et al. (2011); Lusardi and Mitchell (2014), among others), understanding both the actual magnitude and the cause of women's disadvantage in financial knowledge is essential for policymakers.

The household finance literature has explored a number of factors associated with, and possibly responsible for, the gender gap in financial literacy, from marital status to educational levels and from labor force participation to non-cognitive skills and expectations (Arellano et al. (2018); Driva et al. (2016)). Yet scholars have reached no consensus as to which factors are most important in causing these gaps in financial literacy (Lusardi and Mitchell (2014)).

Surprisingly, the issue of how financial knowledge is tested and how that testing strategy is associated with the observed differential patterns across genders remains under-researched. Ample evidence shows that dependencies exist between respondents' test results and test formats, with multiple-choice testing tending to favor males over females and with constructed-response questions showing the opposite (Lumsden and Scott (1987), Ferber et al. (1983)). However, as Wuttke et al. (2020) point out, research on the issues of test bias in financial literacy assessments is scarce. This research gap is even more puzzling given that the standard measure of financial literacy, first introduced in the 2004 Health and Retirement Study (Lusardi and Mitchell (2011)) consists of a set of 3 to 7 multiple-choice questions testing understanding of basic concepts such as interest, inflation, and risk diversification.

In this paper, we investigate whether the tools used for measuring financial literacy amplify the gender gap in test performance, thereby biasing scholarly understanding of the actual knowledge gap in the financial domain. More specifically, we analyze the differential gender effect of question formats on students' financial literacy assessments by comparing the test results from multiple-choice testing formats and open-response ones. To do so, we use the financial literacy module of the PISA 2015 assessment, which allows me to study gender differences among 13 country-representative samples of roughly thirty thousand 15-year-old students.

Unlike most available financial literacy surveys, PISA consists of a weighted mix of open-response (writing two or three sentences) and multiple-choice test items (selecting one of several answers).

[1] On average, those who are more financially knowledgeable borrow at low costs, diversify risk and accumulate wealth for old age (Rooij et al. (2011); Lusardi and Mitchell (2014), among others)

In contrast with the wide persistent gap reported by most of the empirical household finance literature (Lusardi and Mitchell (2014), Hasler and Lusardi (2017)), PISA is one of the few financial literacy tests that does not show a gender differential across survey waves (OECD (2020)), as opposed to the wide and persistent gap reported by almost any other empirical evidence on the issue (Lusardi and Mitchell (2014), Hasler and Lusardi (2017)). To identify the effect of question formats on gender gaps in financial literacy, we exploit the unique characteristics of PISA data. The availability of answers to both multiple-choice and open-response questions for each student allows us to estimate the within-student effect of question format on the probability of correctly answering financial literacy test items for both male and female 15-year-olds. This panel-like specification also allows us to control for student-invariant characteristics, such as unobserved cognitive skills, family background, and underlying non-cognitive skills.

As a first result, we find that while on average 15-years-old girls do not have lower overall financial literacy than boys (in line with OECD (2017b)), a gap appears when I differentiate between multiple-choice and open-response questions. More specifically, girls' likelihood of correctly answering financial literacy test items is about 2.5% percentage points lower when the question is formulated in a multiple-choice format. In contrast, I find no gap for open-response ones. After I control for a wide array of individual and contextual factors, such as students' direct financial experience, cognitive abilities, and non-cognitive skills, the result remains robust, in contradiction to most of the findings in the educational literature (Baldiga (2014), Riener and Wagner (2017), Akyol et al. (2016)). Our results suggest the mental strategies (cognitive processes, as defined by the OECD) that test takers need for answering a given assessment item might explain the gender-by-format effect. Boys perform better using the multiple-choice format, which more commonly calls for skills in analyzing and identifying financial information—a cognitive process that males tend to perform better. In contrast, girls perform better when answering questions that require them to evaluate or explain an issue, a cognitive process more often assessed in open-response questions.

The PISA population is very different from the one tested in typical household financial surveys. Thus, because the construction of the PISA financial literacy test is not comparable to that of household surveys, generalizing our results outside the school context is difficult. Nonetheless, we provide evidence that the results from the student sample are robust even when the test questions cover topics such as inflation and investing—questions typical in household finance surveys. Therefore, the results are critical for scholarly understanding of how to correctly assess financial literacy levels in the general population.

This paper contributes to both the financial and education literature in substantive ways. First, it provides an alternative explanation to a well-consolidated finding in the household finance literature that raises awareness on the general way financial literacy is tested in household surveys. Indeed, despite studies examining the persistent gender divide in financial knowledge, across surveys and countries—whether considering traditional socio-economic factors, such as educational level or marital status (Grohmann et al. (2018), Alessie et al. (2021), Mahdavi and Horton (2014) among others) or within household specialization (Fonseca et al. (2012))—have reached no conclusive answers. Likewise, studies on the importance of personality and non-cognitive traits (Arellano et al. (2018)),, such as differences in individual disposition and interest in financial issues (Lührmann et al. (2015)) or in expectations (Driva et al. (2016))—have inconclusive findings. HHowever, a definitive answer has not yet been found, and a considerable portion of the differences in performances across genders remains unexplained (Fonseca et al. (2012), Lusardi and Mitchell (2014), Alessie et al. (2021)).

Second, my results adds to and extends the growing literature questioning what standardized assessment precisely measures and whether differences across groups in personality traits and non-cognitive skills appear to directly affect test performance (Borghans et al. (2016), Borgonovi (2021), Brunello et al. (2021), Anaya et al. (2021)). A rich literature on the economics of education demonstrates the gender difference in performance on multiple-choice tests, explained partly by women's higher tendency to skip questions on m-c tests and by overall gender differences in non-cognitive traits such as willingness to guess, confidence, and risk aversion (Baldiga (2014), Alessie et al. (2021), Iriberri and Rey-Biel (2021), Brunello et al. (2021), Pekkarinen (2015)). This literature also examines how different types of penalties for incorrect answers can reduce the gender gap relative to unanswered questions (Riener and Wagner (2017), Coffman and Klinowski (2020), Saygin and Atwater (2021)).

Third, limited evidence for gender bias in texting formats is available for subjects other than mathematics. Mondak and Anderson (2004) provide evidence from political knowledge surveys, and Baldiga (2014) tests the willingness to guess in subjects answering multiple-choice questions from an SAT history test. Closely related to our work is Siegfried and Wuttke (2019), who test economic knowledge on a sample of roughly 200 German students using constructed- and selected-response formats for similar question contents. Although they find no evidence of item formats showing a gender differential effect, their analysis suggests that a mix of test formats can balance out gender differences.

Although these and all previous studies provide valuable insights into the interaction between gender and question formats, they often (a) perform their analysis on small samples and (b) do not have the data for directly comparing multiple-choice and open-response formats: Moreover, as they

often focus on very specific contexts (e.g., high-stake college entrance exams), their results are not largely generalizable. This present paper constitutes the first large-scale empirical analysis using a data set that allows for investigating the interaction between gender and test question formats in the financial literacy domain. Our results are quite relevant, especially in light of the uniform way in which financial literacy is tested (i.e., by means of multiple-choice questions), and suggest the need to ensure fairness in financial literacy assessments. In addition, because of the data I use, I can analyze the mechanism underlying the results that other studies were not able to investigate[2]—the multidimensionality of a test item and the way in which question characteristics other than formatting can also interact with the gender dimension and explain the gender-by-format effect. These characteristics include content, context, and cognitive processes. Understanding the financial knowledge dimensions that a test measures and the way that specific question formats can favor certain groups is critical for redressing gender and other inequalities.

The rest of the paper is structured as follows: Section 2 describes the dataset and outlines some descriptive results. Section 3 presents our baseline results and tests their robustness, and Section 4 nvestigates the mechanisms underlying those results. Section 5 concludes.

2 Data and Methodology

2.1 Financial Literacy in PISA Data

To investigate the research questions outlined above, I employ data from the Programme for International Student Assessment (PISA hereafter). The Organisation for Economic Co-operation and Development (OECD) has administered this international standardized test since the year 2000 on a three-year cycle, assessing achievement in math, science, and reading of representative random samples of 15-year-old from a wide range of different countries. The PISA dataset is the result of a two-stage stratified design, where, first, individual schools are sampled, and secondly, students are sampled within sampled schools.

In 2012 the OECD developed for the first time a framework to assess financial literacy in the student population, implemented as well in the 2015 and 2018 waves of PISA. The financial literacy assessment consists of 43 question items, divided into two clusters and tested in a one-hour-long testing session. At the same time, the students answering the financial literacy clusters also take part in the science, reading, and math assessment (OECD (2013)).

As some of the PISA questions are re-administered across waves, the exact wording and content

[2]See for example Baldiga (2014), underlying the need to analyze other factors besides confidence and other non-cognitive traits

of the items are not publicly available. However, the OECD shares other details about the assessment questions used. First, in all PISA assessments, items are divided into two broad format categories. On the one hand, we observe *constructed-response items*, which require students to generate an answer autonomously. In these instances, the respondent may be required to write a couple of sentences (*open response items*), write a single word, or insert a number as an answer (*closed-response items*). On the other hand, we have *selected-response items*, a widespread testing procedure, especially in the financial literacy framework, requiring individuals to pick an alternative out of an already determined set of options (*simple* or *multiple choice items*, according to the number of possible options the student has to choose from). An example of both question formats from the 2012 wave can be found in the appendix Figure A.2 and A.1.

Besides this first classification based on the item format, the OECD financial literacy framework categorizes each question according to the level of difficulty, the area of tested knowledge (*content*), the situation in which the financial knowledge is applied, ranging from personal to global (four different *contexts*), and the four cognitive processes required to answer (*process*). Appendix A.1 and the corresponding Table A.1 report the exact classification of items according to format, by content, context, and process and dig into the details of these classifications. As can be seen, there is an overall balanced division of questions characteristics, although different formats lend themselves to assess specific skills and questions. Previous research has already suggested that different groups of students may display heterogeneous response patterns according to the item formats, which is one of the arguments for OECD to retain a mixture of selected and closed-response items, as the format of the item should not affect the interpretation of the intended object of measurement (OECD (2017a)).

Several features make PISA data an interesting choice for our analysis. Most importantly, PISA provides to us the only, to the best of our knowledge, large-scale financial literacy assessment combining a set of multiple choice and open response questions. This is crucial for our analysis in that it allows us to compare the response behavior across gender by question format, rather than focusing exclusively on multiple-choice as most financial knowledge assessments do. Moreover, according to the overall PISA 2015 scores, there is no evidence of gender gaps in the financial knowledge domain, contrary to most of the existing literature and studies (see Arellano et al. (2018) and Alessie et al. (2021) among others). This is a puzzling fact by itself, given the robust evidence on financial literacy gender gaps worldwide, and it deserves some attention.

Another crucial feature of PISA is that it is a low-stake exam whose results are not directly affecting students' outcomes in or outside school. Students do not receive a grade nor the correct solutions to the questions, and feedback about the test performance is only returned to the schools'

principles in an aggregate format. The existing studies analyzing the issue of differential gender patterns in testing formats take into consideration high-stake exams, such as university entrance exams or school-exit tests (Montolio and Taberner (2021), Saygin (2020)). In such contexts, the pressure to perform is undoubtedly higher, and these studies aim to analyze how gender differences in propensity to guess, levels of pressure, or anxiety affect students' outcomes while interacting with the multiple-choice format. Our study can consider the issue from a different perspective, discarding the role that pressure or competitiveness may induce in high-stake exams to focus solely on the question format.

Finally, PISA contains a rich set of information about the students' background, demographics, and cognitive skills, a type of information rarely available in household surveys that assess adults' financial literacy. The richness of the data allows investigating all factors potentially affecting the financial knowledge of individuals who have gained relatively little exposure to financial markets.

2.2 Sample and Descriptive Statistics

In our analysis, we make use of the PISA 2015 financial literacy assessment. The reason to privilege the 2015 wave is double. On the one hand, the 2015 assessment took place in a computer-based mode, as compared to the 2012 wave, which was paper-based. This allows, first, to minimize human mistakes and discretion in grading. Secondly, it provides a set of earlier unavailable information about the way students behave during the test (timing of response, number of clicks, and other information automatically recorded while the students answer the PISA assessment, OECD (2017a)). On the other hand, compared to PISA 2018, the 2015 wave provides a richer array of information on students' non-cognitive traits, such as anxiety and motivation, and students' engagement in financial behaviors.

The full sample includes more than 30,000 pupils living in 14 different countries[3]. Throughout the paper, we use the final student weights, which allow us to scale the sample up to the size of the countries' 15^th-years-old student population and account for the oversampling of specific regions. Appendix Table A.3 provides a set of descriptive statistics for the sample of interest, split by gender of the students. We can see that, broadly, students present similar characteristics across gender in terms of socio-economic status, age, or immigration background, in line with the OECD efforts to provide representative samples of the country populations (OECD (2016). However, it also comes as no surprise that some differences in the two groups can be observed as far as concern behaviors (with guys being, for example, more actively involved in working activities, as compared to girls), non-cognitive characteristics (with girls being more motivated in school as compared to guys) and cognitive skills, with boys outperforming girls in math literacy, and vice-versa in reading literacy.

[3]The participating countries are: Belgium, Brazil, Canada, Chile, Italy, Lithuania, Netherlands, Peru, Poland, Russia, Slovakia, Spain, USA, China. See Appendix Table A.2 for sample sizes specific to each country.

A preliminary set of descriptive evidence can better motivate our analysis. As can be seen from column (1) of Table 1, there is no statistical support to the hypothesis that girls perform worse than boys in the financial literacy in the world sample. The gender coefficient estimated through OLS, despite being negative, is not statistically different from zero as far as concern the overall score for financial literacy[4]. This is not surprising and in line with what is already underlined in OECD (2017b), where Italy is reported to be the only country with a significant underperformance of girls as compared to boys in financial literacy. However, the emerging picture is quite different when we dismiss the general PISA score and decompose the questions according to their format (i.e., selected- and constructed-response). Columns (3) and (4) show that girls, on average, answer significantly fewer multiple-choice questions than boys do, while such association is not found for open questions. It is important to underline here that we are not employing the official PISA proficiency scale. The financial literacy score, as all literacy scales in PISA, is usually computed following the OECD guidelines (OECD (2016)), through IRT methodologies that take into account the weighted presence of differential question formats. Hence, when looking at correlations between gender and financial literacy by question format, we could be observing spurious correlations related to the choice of our dependent variable (the proportion of correctly answered questions, in Table 1). Different dependent variables reported in Appendix Table B.1, however, confirm such result: when analyzing exclusively open-response questions, girls in PISA 2015 perform significantly better than boys, whereas, in multiple-choice questions, the finding is reverted. Figure 1 visually reports these preliminary descriptive findings.

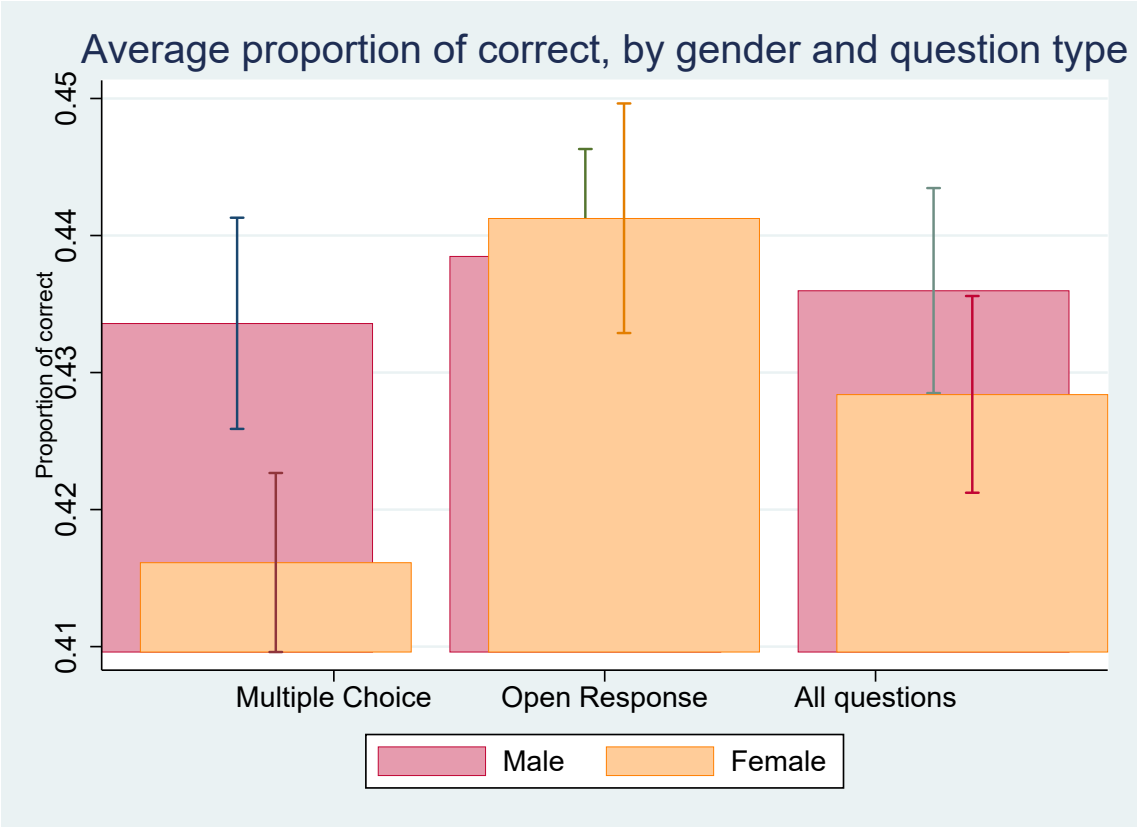
Table 1: Unconditional correlations between gender and financial literacy

	(1) PISA FL score	(2) Prop of correct	(3) Prop of correct OR	(4) Prop of correct MC
Female	-3.190	-0.007	0.003	-0.017***
	(2.339)	(0.004)	(0.004)	(0.004)
Constant	478.001***	0.436***	0.438***	0.434***
	(2.160)	(0.004)	(0.004)	(0.004)
N	33189	33189	33189	33189

Note: The table reports estimates from OLS regressions. The dependent variable are: in (1) the PISA financial literacy score computed following OECD procedure; in (2) the proportion of correct answers out of all the questions; in (3) the proportion of correct among the 21 open response questions; and in (4) the proportion of correct among the 22 multiple choice questions. All regressions include a constant. Errors are clustered at the school level (in parenthesis), data are weighted and BRR replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[4]PISA test scores are computed through Item Response Theory, using a two-parameters Rasch Model

Figure 1: Proportion of correct answers, by gender and question format



Notes: for each question format, by gender, the figure reports the average proportion of correct answers out of the total (22 multiple choice questions, 21 open-response questions). Errors are clustered at the school level, data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights.

3 Empirical Analysis

3.1 Female Premium associated to Different Question Formats

The results above, while being suggestive of the existence of a significant correlation, may not be too informative since the specification in Table 1 does not allow me to compare directly selected- and constructed-response questions, nor to control for features other than individual characteristics. Hence, to account for these problems in the estimation of item format effects, I rely on within-student variations in correct answers across different questions. Based on this approach, I examine whether differences in the probability to answer across students' gender are systematically associated with questions differences in formats. The identification strategy relies upon the idea that students' characteristics across schools, such as their innate ability or non-cognitive traits, are the same for all the financial literacy test. The only difference resides in the fact that some questions present a different answer format. To analyze formally the probability of correctly answering a given question and how this varies according to gender and questions format, I estimate within-student models in a quasi-panel specification at the question-by-student level:

$$C_{iqs} = \beta_0 + \beta_1 female_{is} + \beta_2 MC_{iqs} + \beta_3 female_i * MC_q + \lambda X_{is} + \gamma School_s + \delta I_q + (\mu_i + \eta_s) + \epsilon_{iqs} \quad (1)$$

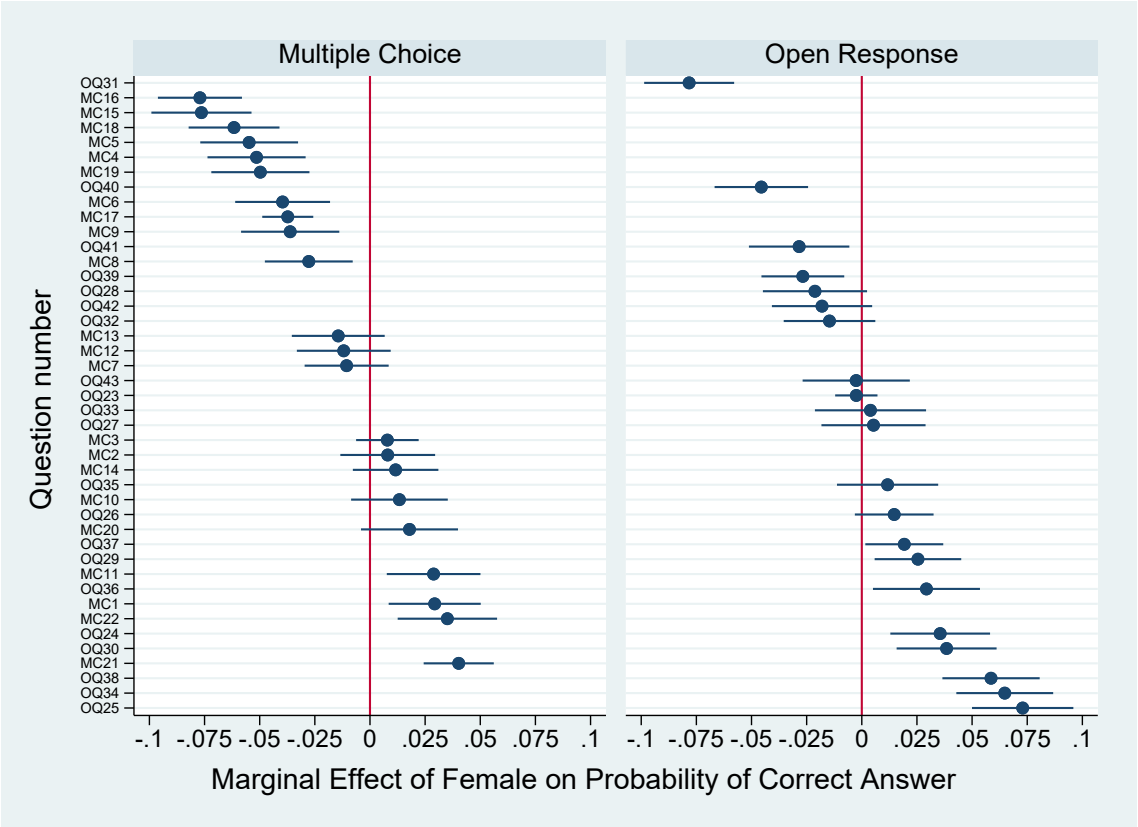
where the binary outcome, C_{iqs}, indicating whether the ith student in the sth school has correctly answered the qth question, is a function of different variables. X_{is} is a vector of student-specific characteristics measured in 2015, such as age, migration status, and socioeconomic background, $School_s$ is a vector of school-invariant characteristics, and $female_{is}$ is the gender of the student. MC_q captures the question format, being equal to one if the specific item is of the *selected-response* format (i.e., a multiple-choice question) as opposed to the constructed-response one. Finally, I_q is an indicator for questions characteristics other than the format and $(\mu_i + \eta_s)$, used in subsequent stages of the analysis, represent the student and school fixed effects, capturing underlying characteristics such as students' cognitive skills or schools resources. ϵ_{iqs} is the remaining error term, which we cluster at the student level in the estimation procedure. The estimate of interest is β_3, associated with the interaction term between gender of the student and question format. Hence, β_3 measures the differential relationship between financial literacy (i.e., the probability of correct answer) and question format for females relative to male students.

Figure 2 give us a visual inspection of the issue of interest, as for each question we look at girls probability of correct response as compared to boys. Although negative (and positive) premia

for female students, as opposed to males, emerge in both sets of questions, it clearly stands out how the majority of negative premium the fact that such probability is more likely to appear for girls in multiple-choice questions as compared to open-response ones.

The results of the estimation of Equation 1 are reported in Table 2. From column (1), we observe, first, that on average, all students appear to struggle more with multiple-choice as compared to open-response questions, as the estimate for β_2 is significantly smaller than zero. Moreover, if there is no statistical evidence supporting the existence of a generic gender differential (β_1 being not significantly different from zero), we observe that girls, as compared to boys, have a lower probability of answering multiple-choice question items as compared to open-response ones. In particular, once we account for baseline characteristics such as age, immigrant background, school type, and socioeconomic status, female students' probability of answering financial literacy multiple-choice questions is lower, on average, by two percentage points, as compared to male students. To put the estimate into context, while a boy experiences, on average, a decrease in the probability of correct answer of roughly 9 percent when answering multiple-choice questions as compared to open-response ones, being a female is associated with an additional decrease of about 5 percent for the same probabilityThis is calculated as $\frac{\hat{\beta_2}}{C_{isq,mean}} = 0,09$ and $\frac{\hat{\beta_3}}{C_{isq,mean}} = 0,05$.. The magnitude of the effect is not negligible, considering that, for example, a one-standard-deviation increase in the students' socioeconomic status is associated with a 12 percent increase in the probability of correctly answering financial literacy questions, while being a first-generation immigrant with a decrease of 3 percent in financial literacy. The results of the estimation of Equation 1 are reported in Table 2. From column (1), we observe, first, that on average all students appear to struggle more with multiple choice as compared to open-response questions, as the estimate for β_2 is significantly smaller than zero. Moreover, if there are no statistical evidence supporting the existence of a generic gender differential (β_1 being not significantly different from zero), we can clearly observe that girls, as compared to boys, have a lower probability to answer multiple choice items as compared to open-response ones. In particular, once we account for baseline characteristics such as age, immigrant background, school type and socio-economic status, female students probability to answer financial literacy multiple choice questions is lower, on average, by two percentage points, as compared to male students.

Figure 2: Female premium associated to each question



Notes: The figure plots the probability of correctly answering a given question, i.e. the female premium associated to each question. Each point represents the estimated $female$ coefficient in a regression of the type $C_{is} = \beta_0 + \beta_1 female + e_{is}$, estimated for each question in the test. Data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights.

Table 2: Probability of answering correctly as a function of item format

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.008	0.008	0.009*	0.009*	0.009*	-0.087
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.219)
MC	-0.046***	-0.046***	-0.045***	-0.045***	-0.045***	-0.045***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female ×MC	-0.023***	-0.023***	-0.023***	-0.023***	-0.023***	-0.022***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age		-1.475*	-1.563*	-1.550*	0.005	0.002
		(0.849)	(0.811)	(0.810)	(0.007)	(0.010)
Socio-Econ status			0.052***	0.052***	0.051***	0.047***
			(0.002)	(0.002)	(0.002)	(0.003)
First Generation				-0.000	-0.002	-0.005
				(0.010)	(0.010)	(0.016)
Second Generation				-0.015	-0.018	0.015
				(0.012)	(0.012)	(0.017)
Pre-Vocational school					-0.134***	-0.135***
					(0.009)	(0.014)
Vocational school					0.013**	0.019**
					(0.005)	(0.008)
Modular school					0.035***	0.030***
					(0.006)	(0.009)
Constant	0.531***	12.129*	12.855**	12.751**	0.467***	0.512***
	(0.004)	(6.710)	(6.409)	(6.407)	(0.110)	(0.163)
Age		X	X	X	X	X
ESCS			X	X	X	X
Immigrant				X	X	X
School type					X	X
Interactions						X
N	1143805	1143805	1143805	1143805	1143805	1143805

Note: The table reports results from OLS estimation. The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and the last column includes interactions of all controls with the $female$ variable. Errors are clustered at the student level (in parenthesis), data are weighted and replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.2 Robustness Checks

These results are highly robust to the addition of several other variables (see Table 3) that may differentially influence the ability of the students to answer a question correctly across genders. First, in columns (1) and (2), we add controls for the students' cognitive skills, namely the math and reading PISA test results. As pointed out by Skagerlund et al. (2018), cognitive skills and numeracy, more specifically, are important determinants of financial literacy. While there is not a perfect intersection between the content of financial literacy and mathematics literacy in PISA (OECD (2017a)), it is undoubtedly true that a certain level of numeracy and familiarity with computations are somehow a prerequisite for financial literacy. Similarly, students with poor reading skills may struggle in general to answer any written test.

Further, in column (3), I introduce controls for students' familiarity with financial products (such as debit cards and bank account) and their primary sources of money (whether they received money as a result of proper work activities, gifts from the family, or allowances). Columns (4) to (7) progressively add country, region, and school fixed effects (η_s), to take into account overall differences in school systems around the world, which may systematically affect students in a differential way across gender. Finally, in (8), we introduce students' fixed effects (μ_i), capturing individuals' fixed characteristics such as innate ability, family background, and non-cognitive skills. Controlling for individual fixed effects, in this setting, allows us to control as well for school unobserved characteristics. Given that we exploit within-student variation, it is not possible to estimate directly the coefficient associated with *female* since gender is a student-invariant characteristic. We can, however, estimate its interaction with the question format.

In column (9), in addition to students' fixed effects, we also control for question-related characteristics. As explained in Section 2, we do not know the exact content of each question, but we have information related to their context, content, and process. The estimate of our coefficient of interest, the interaction between questions format and students' gender, is basically unaffected in magnitude and significance. Hence, a robust association between gender and item format emerges from our analysis, highlighting how girls perform worse in multiple-choice questions than boys. The observed association does not seem to be explained by the common school- and student-level characteristics suggested by previous literature.

Finally, our results are robust to estimating a Probit model instead of a linear probability model to better account for the binary nature of the dependent variable[5].

[5]Results not reported but available upon request.

Table 3: Probability of answering correctly as a function of item format: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.028***	0.006*	0.004	0.005	0.005	0.004	0.002		
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)		
MC	-0.037***	-0.036***	-0.039***	-0.036***	-0.036***	-0.037***	-0.035***	-0.034***	-0.032***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Female × MC	-0.023***	-0.024***	-0.027***	-0.024***	-0.024***	-0.023***	-0.024***	-0.025***	-0.025***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
N	1143805	1143805	825745	1143805	1143805	852372	1143805	1143805	1143805
Baseline Controls	X	X	X	X	X	X	X		
Math Literacy	X	X	X	X	X	X	X		
Reading Literacy		X	X	X	X	X	X		
Financial Behavior			X						
Country FE				X					
Regions FE					X	X			
School characteristics						X			
School FE							X		
Student FE								X	X
Question characteristics									X

Note: The table reports results from OLS estimation. The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant; baseline controls are the ones included in column (5), Table 2; the 10 controls for financial behavior included in column (3) refers to the source of money for the student (7 categories), whether the student has a bank account, a debit card and savings. School characteristics in column (6) includes controls for school size, teacher-to-student ratio, class size, private school dummy, school enrollment by gender, and indexes of students and teachers beahvior. Errors are clustered at the student level (in parenthesis), data are weighted and replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.3 Heterogeneity Analysis

3.3.1 Generalizability of the Results

Inflation, Investing and Income Tax: Different Questions Topics

The PISA test is quite different from the tests commonly implemented in Household Finance research. A relevant difference relates to the measure of financial literacy. Commonly, financial literacy is measured with a limited set of questions on inflation, interest rate, and risk diversification, first introduced in the Health and Retirement Study in 2004 (Lusardi and Mitchell (2008)). As explained in Section 2, the measure of financial literacy in PISA assessment has a pretty different content, as it is directed to 15-year-old students with limited financial experiences. It is straightforward to understand how some of the concepts commonly tested in household finance research are very much foreign to 15-years-old students, who have, for example, no experience with the stock market. Given the differences in financial literacy measures, whether our result is generalizable to other contexts may be debatable. However, a subset of the PISA questions is somehow more relatable to the standard measure of financial literacy, being centered around concepts such as income tax, interest rate, bank statement and investing[6].

Table B.3 reports separate estimates for such subset of questions (in column 1) and the remaining others (column 2). While the multiple-choice-by-gender pattern clearly emerges in both subgroups, the estimates reveal an interesting result. The negative female premium in selected-response questions is more than twice in magnitude when focusing on standard household finance questions topics, and such difference is statistically different from zero. This suggests that, first, there are some heterogeneous effects according to the topic of the question. Secondly, our results may also hold in a context where the measure for financial literacy coincides with the standard one in the literature.

Financial Behavior

A second reason why PISA is quite different from the standard household finance context is that 15 -years-old students have probably had few chances of gaining experience with financial products, contrary to the adult respondents that generally participate in household finance surveys. However, evidence from the PISA data shows that students already have day-to-day experiences related to managing essential financial products and money in many countries. The student background questionnaire, in fact, includes a wide array of questions aimed at analyzing students' relationship with money, and, more specifically, whether they discuss money matters with parents[7], whether they hold basic financial products (bank accounts and debit cards[8]), and whether they receive money from different sources (in the form of allowances and gifts, or as a payment for different types of work[9]). This array of information can be informative as boys and girls may have, already from a very young age, quite different opportunities for being exposed to financial matters and having direct experiences with money. Such heterogeneity could also reflect on their financial literacy skills.

From a visual inspection of the answers to the questions of interest (see Figures B.1, B.2 and B.3), we observe that as far as concern experience of financial products and discussion about money, there is no significant difference between the male and the female population. This information had also been included in the estimation of Table 2 and did not relevantly affect our result. On the contrary, we observe some differences between boys and girls on where their

[6] As already specified, the exact content of the question is not disclosed by the OECD, which however provides a name for each question item (e.g. "Phone plan", "Music store" etc.). This allows us to isolate ten questions appearing under the following names: "Interest rate", "Investing", "Bank statement", "Income tax".

[7] The exact question is: "How often do you discuss money matters with: Parents guardians or other adult relations?/Friends?".

[8] The exact question is: "Do you have either of the following? Bank account/Pre-paid debit card".

[9] The exact question wording is: "Do you get money from any of these sources?", repeated for different categories "An allowance or pocket money for regularly doing chores at home", "An allowance or pocket money, without having to do any chores", "Working in a family business", "Working outside school hours (e.g. a holiday job, part-time work)", "Occasional informal jobs (e.g. baby-sitting or gardening)", "Gifts of money from friends or relatives", "Selling things (e.g. at local markets or on eBay)".

money comes from and whether they hold a bank account. In general, it appears to be the case that girls are more likely than boys to receive money from gifts or allowances, whereas boys are more likely to be involved in selling and regular work activities (whether inside the house or from jobs) receiving money in exchange for that. Based on this evidence, we estimate again Equation 1 for subgroups of our sample reflecting students' source of money and use of financial products[10]. The estimates in Table B.2 reveal that also students who engaged in financial behavior and got exposed to financial matters exhibit a negative probability of answering multiple-choice questions. Hence, the negative female premium associated with multiple-choice questions is similar, regardless of a student's experience with money. This may also hint at the fact that, even in an adult population, with considerably higher financial experience, these results could emerge, nonetheless.

3.3.2 Teaching to the Test

In this subsection, we perform the estimation outlined in Equation 1 by subgroups, according to a specific school policy. To be more precise, we look at whether a school implements the practice of assessing students' competencies through standardized tests. Students (and more so female students) who are more used to such types of tests, and hence more used to answer multiple-choice questions in their daily school-life, may display differential probabilities of correctly answering selected-response items. The School Questionnaire administered by PISA provides information on how often each participating school tests its student via standardized tests[11].

Table B.4 reveals that, indeed, the estimated interaction between gender and items format is no longer significant once we exclusively consider schools that consistently implement such policy. Not only, also the overall students struggle with multiple-choice questions (the estimated β_2) is marginally lower in these schools, as compared to those who never test their students with standardized tests. This result is quite important as it suggests that the negative performance of female students in selected-response questions can be mitigated if the schools actively engage in policies aimed at training their students for such a format of tests.

[10] Because the response rate to these questions was limited, the sample analyzed is considerably smaller, as can be seen in Table B.2.

[11] The exact wording of the question was: "How often are students assessed with mandatory standardized tests?". Possible answers were "Never", "1-2 times a year", "3-5 times a year", "Monthly", "More than once a month".

4 Investigating Possible Mechanisms

4.1 The Role of Individual Characteristics

In Table 3 we showed how our results are robust to the inclusion of a wide array of factors, basically any student-invariant characteristics, such as unobserved ability or family background. However, the inclusion of students' fixed effects in the model does not take into account that some individual characteristics, such as cognitive traits, may play a differential role in the probability of answering a question, according to the format of the question. There is ample evidence about how men and women differ in terms of specific psychological traits, such as motivation, anxiety in stress-situations, self-confidence or competitiveness, and how these differences, in turns, correlate with various cognitive outcomes and financial behaviors (Barber and Odean (2001), Dahlbom et al. (2011), Arellano et al. (2018)). If non-cognitive traits' effect on financial literacy differs by question formats, this might explain the observed gender difference in performance.

PISA Students' Questionnaire includes a set of information that can be useful to report such mechanism, as students are requested to answer about their level of motivation and anxiety in different situations[12]. Girls appear to be more anxious than boys on average (OECD (2017a)), and this could negatively affect their likelihood to correctly answer multiple-choice questions more than other formats. Anxiety may hinder their capability to rule out alternative hypotheses or affect their ability to guess an answer (all actions that are involved in the process of addressing selected-response items) (Ben-Shakhar and Sinai (1991)). Similarly, higher motivation to perform well could be an incentive to guess answers when the correct answer is unknown, especially in a context such as the PISA test, which does not include penalty points for incorrect answers. Table 4 reports estimates for our preferred specification where gender and items format are further interacted with students' non-cognitive traits to uncover the existence of a differential gender-by-format effect according to different levels of motivation or anxiety. While both motivation and anxiety play a role in the probability to answer a question correctly, the effect is not heterogeneous across gender nor questions formats, and our estimate of β_3 is mostly unaffected.

Unfortunately, we do not have specific measures for confidence or risk-aversion, which may further support our results. However, we perform analyses similar to the one presented above using alternative proxies provided by the PISA Students' Questionnaire. First, we introduce controls (in the form of a triple-interaction) for two measures of self-efficacy and confidence in science. Optimally, we would like to have information about the same traits for the financial literacy domain, but this is unfortunately not available in PISA. Secondly, previous literature has shown that parents working in STEM-related occupations are associated with higher confidence of sons and daughters in the mathematics domain. Hence, we introduce as well controls for mothers' field of employment. These additional checks also leave our main estimate, β_3, unaffected[13]. Baldiga (2014), in her analysis of gender differences in the tendency to skip questions, also finds a limited role for confidence or risk-preferences. In line with hers, our results suggest that ultimately individual characteristics might not be the driving mechanism beyond the gender-by-item-format difference in financial literacy.

[12] Students gave statements about themselves on a four-point Likert scale ("strongly agree", "agree", "disagree", and "strongly disagree), recoded as dummy categories. The exact questions for anxiety are: "I often worry that it will be difficult for me taking a test"; "I worry that I will get poor grades"; "Even if I am well prepared for a test I feel very anxious"; "I get very tense when I study for a test"; "I get nervous when I don't know how to solve a task at school". For achievement motivation: "I want top ¡grades¿ in most or all of my courses"; "I want to be able to select from among the best opportunities available when I graduate"; "I want to be the best, whatever I do"; "I see myself as an ambitious person"; "I want to be one of the best students in my class".

[13] Results are not reported but available upon request.

Table 4: Effect of Motivation and Anxiety on questions' answers

Motivation						
	(1) Top grades	(2) Opportunities	(3) Want to be the best	(4) Ambitious	(5) Want to be best student	(6) Motivation Index
Female	0.033***	0.050**	-0.002	0.002	0.003	0.008*
	(0.012)	(0.025)	(0.010)	(0.009)	(0.008)	(0.005)
MC	-0.044***	-0.023*	-0.043***	-0.039***	-0.047***	-0.044***
	(0.007)	(0.014)	(0.005)	(0.005)	(0.004)	(0.002)
Female × MC	-0.026***	-0.049***	-0.025***	-0.013**	-0.020***	-0.022***
	(0.009)	(0.018)	(0.007)	(0.006)	(0.006)	(0.003)
Motivation	-0.011	0.075***	-0.019**	0.047***	0.018**	0.010**
	(0.011)	(0.021)	(0.007)	(0.008)	(0.007)	(0.004)
Female × Motivation	-0.028**	-0.044*	0.014	0.014	0.008	0.003
	(0.013)	(0.025)	(0.011)	(0.011)	(0.010)	0.003
MC× Motivation	0.000	-0.022	-0.002	-0.007	0.004	0.000
	(0.007)	(0.014)	(0.006)	(0.006)	0.004	(0.003)
Female × MC × Motivation	0.003	0.027	0.003	-0.016**	-0.005	-0.004
	(0.010)	(0.018)	(0.010)	(0.007)	(0.007)	(0.004)
N	1132983	1128489	1126286	1120239	1127114	1130364
Anxiety						
	(1) Worry about test	(2) Worry about grades	(3) test Anxious	(4) Tens when study	(5) Nervous at school	(6) Anxiety Index
Female	0.007	0.007	0.013	0.007	0.012	0.017***
	(0.008)	(0.011)	(0.008)	(0.006)	(0.008)	(0.005)
MC	-0.039***	-0.038***	-0.046***	-0.048***	-0.044***	-0.044***
	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.002)
Female ×MC	-0.023***	-0.033***	-0.022***	-0.019***	-0.017***	-0.022***
	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.004)
Anxiety	-0.059***	-0.032***	-0.059***	-0.052***	-0.022***	-0.028***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.004)
Female × Anxiety	0.014	0.006	0.005	0.016*	0.000	0.006
	(0.010)	(0.012)	(0.010)	(0.010)	(0.010)	(0.005)
MC× Anxiety	-0.008*	-0.009*	0.003	0.009**	-0.000	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)
Female × MC ×Anxiety	0.002	0.014*	-0.002	-0.011	-0.009	-0.001
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.003)
N	1133753	1130368	1128991	1127742	1129840	1132154

Note: The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for immigrant status, age, family socio-economic status, type of school, mathematic and reading literacy. Errors are clustered at the student level (in parenthesis), data are weighted. Each column (1) to (5) use a different questions about anxiety/motivation, in a dummy format. Columns (6) use indexes of the 5 questions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2 The Role of Question Characteristics

The second mechanism, which we explore in this section, relies on the PISA assessment's uniqueness. PISA allows to analyze the questions' in-depth characteristics and to go beyond the selected- and constructed-item classification, a possibility often not available to many previous studies analyzing the interconnection between gender and student assessment methods. For example, while Saygin (2020) finds stark differences in gender gaps between high school GPA and standardized tests (with females over-performing in the former and under-performing in the latter), the two assessment methods may be in principle very different in terms of content. A direct comparison is difficult to perform. Most previous studies exclusively analyze multiple-choice tests (contrary to our, which is directly comparing them with open-response ones), imposing some sort of exogenous variation on the grading methodology (see, for example, the work from Baldiga (2014)). In our setting, we can easily classify items according to their content, process, and context, meaning that we can directly compare the difference in the likelihood to answer between two otherwise identical questions, apart from their format. As observed in Table 3, column (9), question characteristics, *per sé*, do not explain the format-by-gender pattern, meaning that the item format is not capturing characteristics other than the format itself. However, it could be the case that the financial literacy gap varies with items format because different formats are used to preferentially measure different skills dimensions. These skills, in turn, may differ between boys and girls, i.e., we may be missing an additional interaction between gender, items format, and items content, context, or process. To give a practical example, multiple-choice questions are more often employed to test individual choice and decisions than education-related topics. They are also most often framed in a context where money is discussed (see Table A.1). If girls are less familiar with such topics or context, this may be driving the observed gender gap in financial literacy.

Hence, we estimate how the likelihood of correct answers between boys and girls varies by item format according to the different content, context, or cognitive process underlying the specific question. The estimated model is presented in Table 5.

The first striking result emerging from the set of estimations is the heterogeneity of the gender effect across all the available categories. We observe that females generally outperform males when the context analyzed relates to home and family or societal issues compared to individual-context items[14]. When moving to question content, results suggest a role played by the content framework when analyzing the interaction between gender and item formats. Girls are doing comparatively better than boys in all questions related to planning and managing finances compared to the other content areas.

Finally, the biggest differences emerge when considering the cognitive process required to answer the question. Female students perform better than males in questions that require them to evaluate financial issues, i.e., "recognizing or constructing financial justifications and explanations, ...which also involves cognitive activities, such as explaining, assessing and generalizing" (OECD (2017b)). On the contrary, boys overperform girls when the question calls for analysis or identification of financial information. Importantly, when controlling for the interactive effect of cognitive processes, it no longer appears to be the case that items format has such a strong effect on the girls' probability to answer. In fact, the estimated coefficient for β_3 is no longer significant.

Hence, when accounting for all possible interactions between questions format, cognitive process, and gender, the negative effect of format disappears (see also Appendix Figure B.4). Overall, we provide suggestive evidence that students' performance is possibly dependent on the test format and the cognitive process required to answer the question correctly.

[14]The results on the societal context should be taken with caution as the category includes only one selected-response question.

Table 5: Questions characteristics and the gender gap

Question characteristics	(1) Context	(2) Content	(3) Cognitive Process
Female × MC	-0.020***	-0.022***	-0.005
	(0.004)	(0.003)	(0.004)
Female × Edu and work	0.002		
	(0.005)		
Female × Individual	-0.020***		
	(0.004)		
Female × Societal	0.007		
	(0.006)		
Female × Money		-0.033***	
		(0.004)	
Female × Risk		-0.037***	
		(0.005)	
Female × Financial		-0.017***	
		(0.004)	
Female × Evaluate			0.034***
			(0.005)
Female × Apply			-0.009*
			(0.004)
Female × Identify			-0.017***
			(0.005)
N	1143805	1143805	1143805
Baseline Controls	X	X	X

Note: The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for immigrant status, age, family socioeconomic status, type of school, mathematics and reading literacy. Errors are clustered at the student level (in parenthesis), data are weighted. Each column (1) to (3) introduce a set of controls related to question content(1), context(2) and cognitive process(3). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Conclusion

This paper empirically analyzes how gender differences in financial literacy are affected by the question format used to test financial literacy. Our analysis finds the following: (1) While on average 15-years-old girls do not show lower overall financial literacy than boys, a gap is found when differentiating between multiple-choice and open-response questions. More specifically, boys over-perform girls when answering financial literacy items formulated in a selected-response format, but no gap exists for constructed-response ones. (2) The within-student estimates are shown to be robust to the inclusion of a wide array of individual and contextual factors. On average, females appear to have a 2.5 percentage points lower probability to correctly answer multiple-choice financial literacy questions than males. This corresponds, roughly, to a 5 percent decrease in the probability of correct answer. Importantly, students' direct experience with money and financial products does not seem to directly affect their financial literacy. (3) The gender-by-format effect appears to be absorbed by the gender-by-process interaction. Once controls for the cognitive processes required to answer the questions and their interactions with gender are included, the gender differential effect of items format disappears. The result suggests that multiple-choice questions are more commonly used to assess skills related to analyzing and identifying financial information, where males tend to perform better (Breland et al. (1994), Taylor and Lee (2012)). At the same time, girls are particularly successful when answering questions that require them to evaluate or explain financial issues, commonly assessed through open-response items. From a policy perspective, our results are relevant for several reasons. First, given the methodology broadly used to assess financial knowledge in the population, it is relevant to be aware of what the specific format in use is measuring. While there are indubitable advantages in using multiple-choice based tests, both in the financial literacy domain and elsewhere, because of their efficient and easy-to-implement nature, we highlight a potential issue in the exclusive use of such testing format. When an assessment tool is employed, it is, at the very least, crucial to understand whether there exist group differences related to the measurement mechanism in use. Moreover, previous works have suggested that the gender gap in multiple-choice testing is to attribute to gender differences in non-cognitive traits, such as risk preferences or confidence (Burns et al. (2012), Riener and Wagner (2017)). Our results, in line with Baldiga (2014), suggest that individual-level characteristics do not play a major role while possibly additional factors (such as other questions characteristics) are at work. In our analysis, we uncover how multiple-choice questions may be reflective of skills other than financial knowledge and point to the existence of gender differences in cognitive processes used by students to address a specific question.

Encouragingly, our results also show that the estimated negative premium associated with items format can be alleviated with ad-hoc school policies aiming at getting students used to the format of standardized assessment tests (and to the format of multiple-choice questions). Trying to generalize from the school context, we do not find evidence that the financial literacy gender gap is rooted in a gender gap in cognitive skills; instead, individuals trained in this survey methodology may improve their performance.

Appendices

A Appendix A: the Data

A.1 *Financial Literacy in PISA*

The Financial Literacy Expert Group (FEG) from the OECD developed a working definition of financial literacy that encompasses the PISA definitions of already existing literacies, the skills and behaviors characteristics to the financial domain and the purposes for developing the particular literacy. As read in OECD (2017a):

> *Financial literacy is knowledge and understanding of financial concepts and risks, and the skills, motivation and confidence to apply such knowledge and understanding in order to make effective decisions across a range of financial contexts, to improve the financial well-being of individuals and society, and to enable participation in economic life.*

Clearly, financial literacy in PISA is intended to fall in the range of personal and household finance activities that are experienced and understood by a 15-year-old, such as making a purchase of a music player, and it is in this sense distinguished from a broader concept of economic literacy, including for example concepts of demand and supply, market structures and so on.

Each one of the 43 PISA financial literacy item is categorised according to its content, process and contexts (OECD (2017b)). The **content** of a question concern the areas of knowledge required to answer. The OECD identified four topics for the financial literacy assessment:

- (1) *money and transactions*, which test the general awareness about the different forms and purposes of money and monetary transactions;
- (2) *planning and managing finances* over the short and long term, where finances includes income, expenditure, wealth or savings and wealth creation;
- (3) *risk and reward* in context of uncertainty, where risks relates to both unexpected financial losses and the risk inherent to financial products;
- (4) *financial landscape*, related to understanding of overall characteristics and features of the financial world, such as consumer protection, financial contracts, the consequences of changes in economic conditions and public policies and so on.

The four **process** categories relate to the cognitive processes that a student needs to engage in, when completing a task. A typical task might ask students to *identify financial information*, in terms of recognising financial terminology or specific features in contracts or bank statements. Students can also be required to *analyse information in a financial context*, extrapolating information that are not explicitly provided, or *evaluate financial issues*, engaging in critical thinking to explain or assess in specified contexts. Finally, the fourth process, *apply financial knowledge and understanding*, is characteristics of tasks that involve performing calculations and using knowledge of financial concepts to solve a problem.

Also the context in which the tasks is presented is relevant to cover a broad variety of situations in which students may need to use their financial knowledge. The questions are framed in general-life situations, familiar to 15-years-olds, such as *education and work* or *home and family*. Also the *individual* dimension is quite important though, as students take many decisions that entirely relate to their own personal benefits and gratification. Finally, the OECD also recognizes that students are living in a global and interdependent context, where individual financial decisions often falls into a *societal* context.

Similarly to the other PISA literacy constructs, the items are also distributed along a proficiency scale, according

to their relative difficulty estimated by considering the proportion of test takers answering each question correctly. The financial literacy assessment includes five levels of proficiency and scores scaled on the participating countries, with a mean of 500 and a standard deviation of 100.

To give a concrete example, Figure A.2 and A.1 report the text of two questions that were part of the 2015 financial literacy field trial. Both questions ask students to analyse information in a situation that should be familiar to many students, the choice of a mobile phone plans. However, in Figure A.2 the student is asked to deal with a multiple choice item that belongs to "Home" context and "Financial Landscape" content, while evaluating a financial issue, whereas in Figure A.1 the student is faced with a open-response question, belonging to the "Planning and Managing Finances" content category, the "Analyse information in a financial context" process and the "Individual" context. While this last question has a proficiency level of 3, the multiple choice one is considered to be slightly easier and is classified as of level 2.

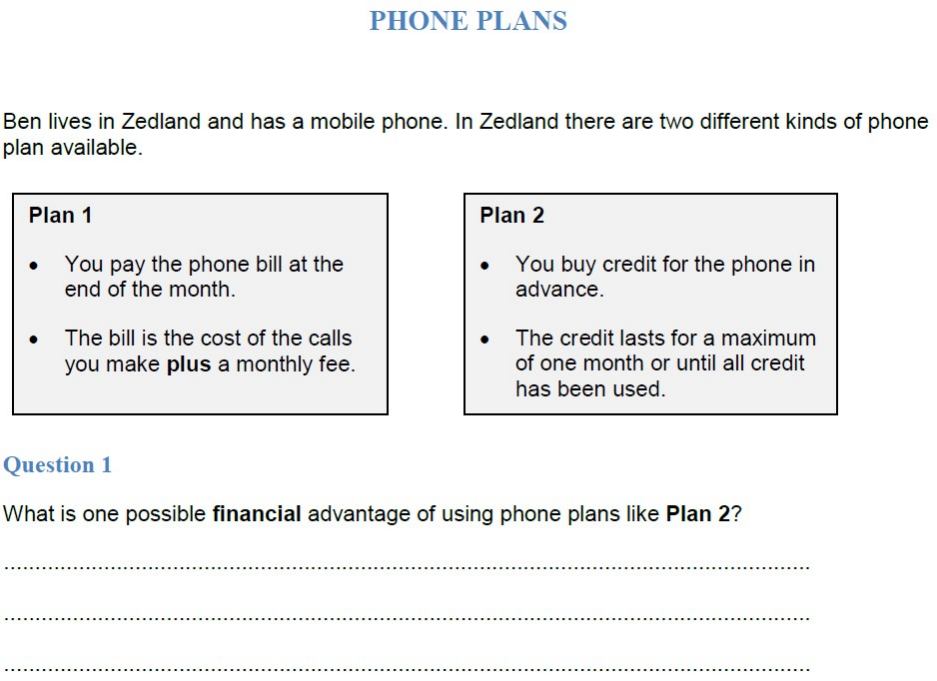
Besides cognitive elements of financial literacy, PISA investigates as well non-cognitive aspects that may influence financial behaviors, focusing especially on three aspects. First, the OECD FEG identified as relevant the process of financial socialization, so the sources of financial information that may be available to students. Parents, peers and schools are considered to be the three main sources through which students can acquire knowledge about personal finance. Second, plausibly, real-life experiences influence students' familiarity with financial literacy concepts, hence, asking about experiences with banks, payment cards, remuneration from work activities and so on is also important to frame the relationship of young people with financial matters. Finally, while it is trues that 15-years old do not have much of a direct experience in financial activities, it is still possible to measure their actual behavior in simple activities such as saving and spending (OECD (2017b)).

Table A.1: Number of questions by format, context, content, characteristic

	Multiple choice	Open Response	Total
Context:			
Home and family	9	6	15
Education and work	2	4	6
Individual	10	7	17
Societal	1	4	5
Content:			
Planning and Managing Finances	7	7	14
Money and Transactions	7	4	11
Risk and reward	4	5	9
Financial Landscape	4	5	9
Process:			
Analyse information in a financial context	8	3	11
Evaluate financial Issues	3	12	15
Apply financial knowledge and understanding	6	4	10
Identify financial information	5	2	7

Note: PISA 2015. Number of questions by each subgroup

Figure A.1: Example of PISA questions: Constructed-Response Item



PHONE PLANS

Ben lives in Zedland and has a mobile phone. In Zedland there are two different kinds of phone plan available.

Plan 1

- You pay the phone bill at the end of the month.
- The bill is the cost of the calls you make **plus** a monthly fee.

Plan 2

- You buy credit for the phone in advance.
- The credit lasts for a maximum of one month or until all credit has been used.

Question 1

What is one possible **financial** advantage of using phone plans like **Plan 2**?

.....

.....

.....

Notes: Example of an open-response item from the "Individual" content area. Source: PISA 2015 (OECD (2013)).

Figure A.2: Example of PISA questions: Selected-Response Item

MOBILE PHONE CONTRACT

Alan wants a mobile phone but he is not old enough to sign the contract.

His mother buys the phone for Alan and signs a one-year contract.

Alan agrees to pay the monthly bill for the phone.

After 6 weeks, Alan's mother discovers that the bill has **not** been paid.

Question

Is each statement about the mobile phone bill true or false?

Circle "True" or "False" for each statement.

Statement	Is the statement about the mobile phone bill true or false?
Alan's mother is legally responsible for paying the bill.	True / False
The mobile phone shop must pay the bill if Alan and his mother do not.	True / False
The bill does not have to be paid if Alan returns the mobile phone to the shop.	True / False

Notes: Example of a multiple choice item from the "Home and Family" content area. Source: PISA 2015 (OECD (2013)).

A.2 *Our Sample*

Table A.2: Countries Samples Size and Financial Literacy Score

Countries	N	Average FL score	Average Male Score	Average Female Score
Belgium	1,433	549.2233	551.8722	546.5125
Brazil	6,078	392.857	388.8341	396.7373
Canada	3,409	533.249	529.2615	537.6662
Chile	1,809	433.126	432.2826	433.9227
Spain	2,684	466.4105	463.7432	469.0406
Italy	3,034	485.9293	492.558	479.1712
Lithuania	1,720	448.0876	434.0714	461.694
Netherlands	1,365	518.7785	515.1535	522.1678
Peru	1,804	403.274	403.2572	403.2902
Poland	1,739	485.5897	479.0915	492.1177
China	2,555	568.5358	572.4119	564.0633
Russia	1,558	511.1433	514.5452	507.9494
Slovakia	1,629	445.848	436.4859	455.5406
USA	2,372	491.7515	494.6429	488.9616
World sample	33,189	474.8852	474.8716	474.8987

Note: The table reports country level PISA score for the financial literacy domain. Average scores are weighted to be representative of the student population in each county and BRR replication weights are employed. Source: PISA 2015.

Table A.3: Descriptive Statistics

	Males			Females				
Variable	N	Mean	Std.	N	Mean	Std.	Min.	Max
Age	15,636	15.81123	0.2926241	15,642	15.81944	0.2911354	15.25	16.42
Socio-Economic Status	15,636	-0.3400792	1.157958	15,642	-0.3799719	1.163285	-5.3505	3.6566
Native	15,636	0.8990913	0.3012174	15,642	0.9013552	0.298194	0	1
Second Generation	15,636	0.0642793	0.2452577	15,642	0.0634245	0.2437326	0	1
Firs Generation	15,636	0.0366293	0.1878561	15,642	0.0352203	0.184342	0	1
School Type:								
General	15,636	0.9181644	0.2741227	15,642	0.9337592	0.2487103	0	1
Pre-Vocational	15,636	0.0045775	0.0675041	15,642	0.0041175	0.0640374	0	1
Vocational	15,636	0.0553612	0.228691	15,642	0.0425915	0.2019404	0	1
Modular	15,636	0.021897	0.1463518	15,642	0.0195319	0.1383893	0	1
Anxiety	15,827	0.0414127	0.9617849	16,053	0.4257317	0.9160259	-2.505	2.5493
Motivation	15,808	0.1956701	0.94731	16,001	0.2457969	0.9300368	-3.0877	1.8543
Math Literacy	15,636	468.6508	107.4295	15,642	454.8321	102.1075	70.796	836.719
Reading Literacy	15,636	464.9903	108.7632	15,642	484.1723	102.7889	36.158	799.943
Father:								
Higher Education	14,342	0.2943475	0.4557648	14,330	0.2647961	0.44124	0	1
Unemployed	13,921	0.0195307	0.1383859	14,059	0.0153568	0.1229718	0	1
Finance job	13,921	0.0048388	0.0693955	14,059	0.008254	0.0904792	0	1
Mother:								
Higher Education	16,248	0.2894859	0.4535377	16,339	0.26989	0.4439161	0	1
Unemployed	14,113	0.1030557	0.3040424	14,570	0.0997679	0.2997006	0	1
Finance job	14,113	0.0203541	0.1412134	14,570	0.0192166	0.1372903	0	1
Financial education:								
At school	13,059	0.6928874	0.4613142	12,649	0.6703574	0.4701019	0	1
Outside school	12,812	0.4793497	0.4995929	12,402	0.4347806	0.4957482	0	1
Discussing with parents	12,885	2.523934	0.9868518	12,537	2.563643	0.9820402	1	4
Discussing with friends	12,763	2.026915	1.008041	12,405	1.904644	0.9593447	1	4
Source of Money:								
Gifts	12,112	0.7943401	0.4041997	11,774	0.8276001	0.3777437	0	1
Proper job	12,088	0.4138535	0.4925432	11,640	0.3383967	0.4731845	0	1
Selling things	11,870	0.4006648	0.4900538	11,423	0.2539443	0.4352852	0	1
Informal jobs	11,877	0.3780564	0.4849222	11,489	0.3785527	0.4850475	0	1
Work in family business	11,886	0.239837	0.4270018	11,473	0.1571386	0.3639472	0	1
Pocket money	12,014	0.3640236	0.4811753	11,650	0.3865102	0.4869707	0	1
Doing home chores	12,235	0.4492362	0.4974367	11,815	0.4249451	0.4943556	0	1
Having a bank account	12,707	0.4540633	0.4979049	12,251	0.4329293	.4955013	0	1
Saving money	12,650	2.73503	1.372606	12,225	2.864755	1.381815	1	6
Having a debit card	12,335	0.19285	0.3945523	11,971	0.1851763	0.3884567	0	1

Note: Standard errors are clustered at the school level, data are weighted and BRR replication weights are used.
Source: PISA 2015.

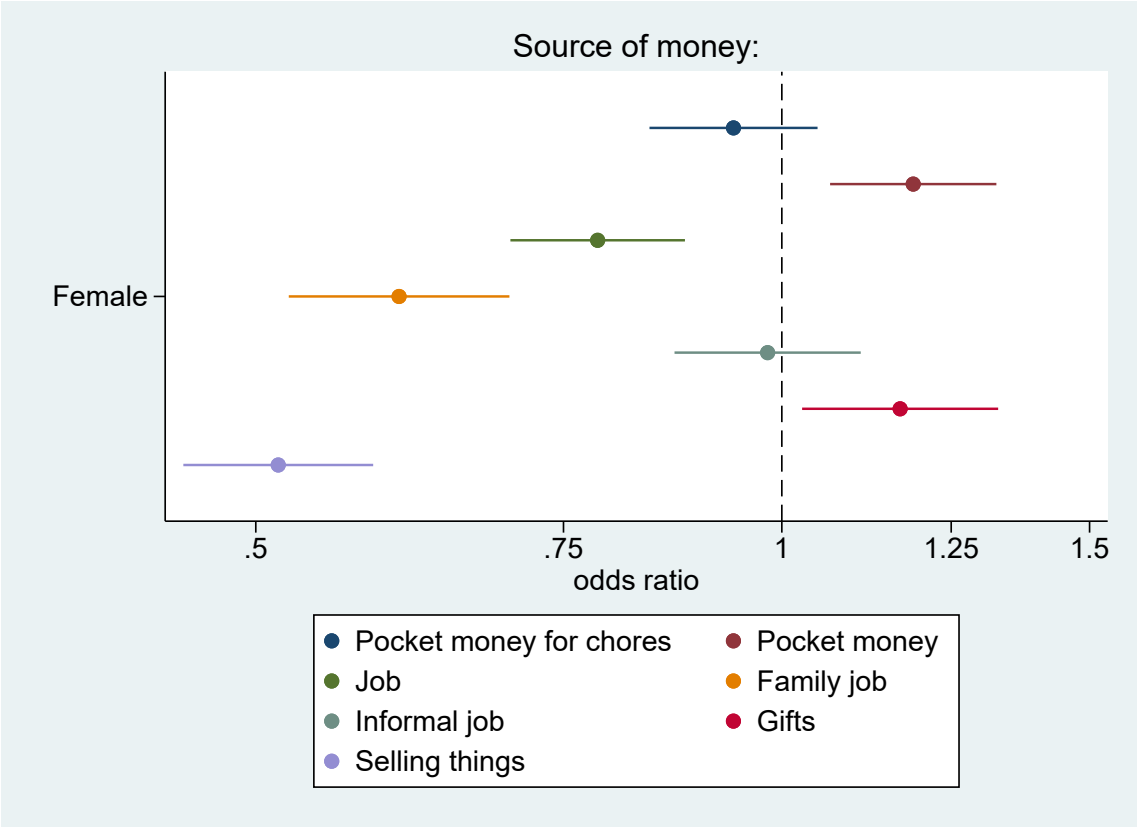
B Appendix B: Further Tables and Figures

Table B.1: Unconditional correlation between gender and financial literacy: different dependent variables

	(1) Prob of N correct above mean	(2) N correct excluding missing	(3) Prob of at least 16 correct	(4) Proportion of correct
	Only multiple choice questions			
Female	-0.026***	-0.324***	-0.051***	-0.071***
	(0.010)	(0.094)	(0.008)	(0.016)
Constant	0.491***	10.838***	0.177***	-0.267***
	(0.008)	(0.093)	(0.007)	(0.016)
N	33189	22856	33189	33189
	Only open response questions			
Female	0.004	0.136	-0.010	0.011
	(0.009)	(0.116)	(0.008)	(0.018)
Constant	0.500***	11.881***	0.186***	-0.247***
	(0.008)	(0.108)	(0.005)	(0.016)
N	33189	14146	33189	33189

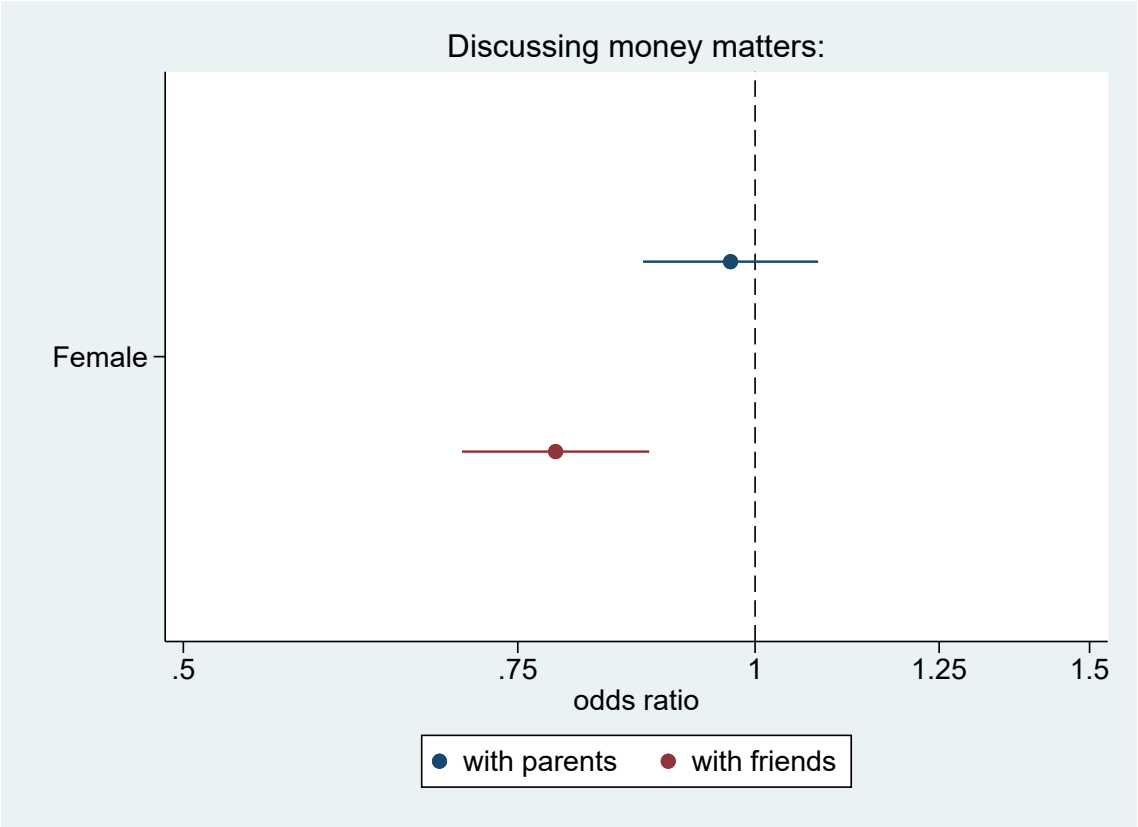
Note: The dependent variable are: in (1) answering correctly more than the average number of correct (dummy); in (2) the number of correct answers excluding missing answers; in (3) the probability of being in the upper quintile for number of correct answers (i.e. at least 16 correct out of 22); in (4) the proportion of correct. (1) to (3) reports marginal effect from OLS, (4) the results of gls estimation with logit as a link function and binomial as a family. All regressions include a constant. Errors are clustered at the school level (in parenthesis), data are weighted are BRR replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure B.1: Likelihood of receiving money from different sources, by gender



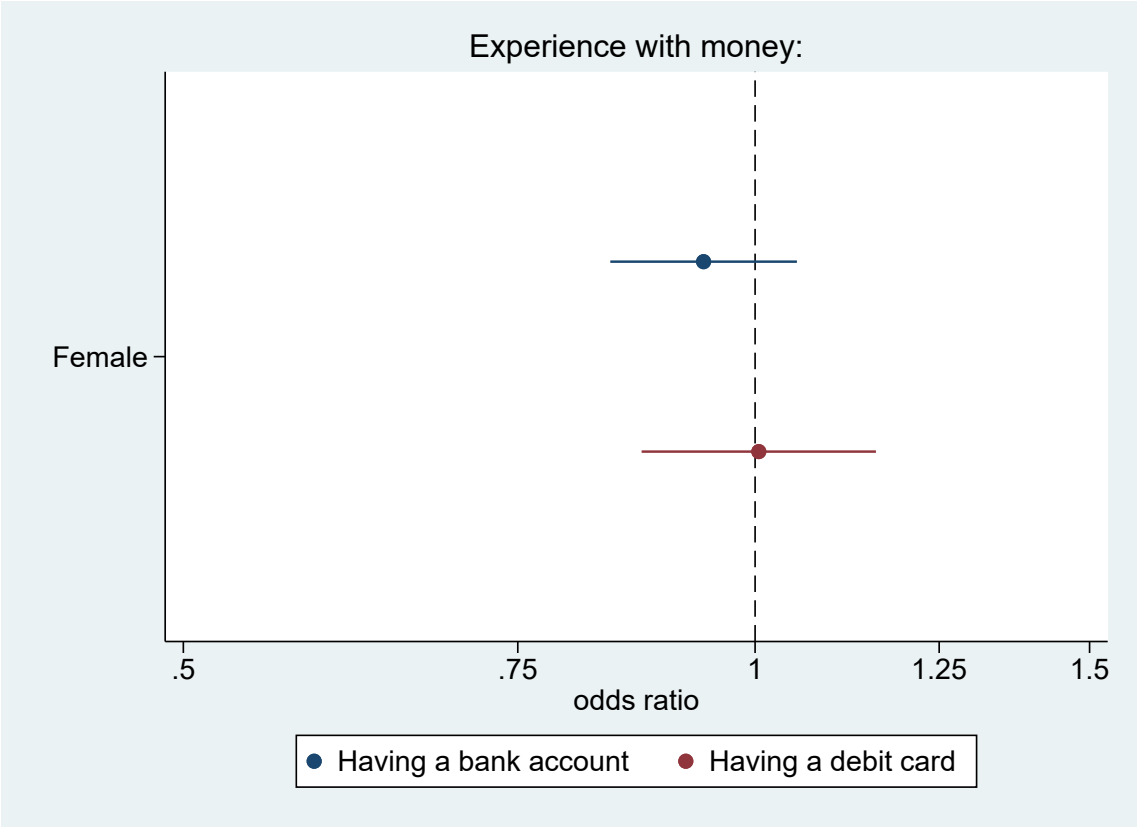
Notes: The figure reports odds ratios. Values greater than 1 imply that girls present on average a higher likelihood for the specific source of money. Each estimate is derived from logistic regressions which control for age, immigrant status, school type, math and reading literacy and socio-economic status. Data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights.

Figure B.2: Likelihood of talking about money, by gender



Notes: The figure reports odds ratios. Values greater than 1 imply that girls present on average a higher likelihood of discussing about money. Each estimate is derived from logistic regressions which control for age, immigrant status, school type, math and reading literacy and socio-economic status. Data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights.

Figure B.3: Likelihood of holding financial products, by gender



Notes: The figure reports odds ratios. Values greater than 1 imply that girls present on average a higher likelihood of holding a specific financial product. Each estimate is derived from logistic regressions which control for age, immigrant status, school type, math and reading literacy and socio-economic status. Data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights.

Table B.2: Heterogeneous effects: students' experience with financial matters

	(1)	(2)	(3)
Experience with finance	Bank account or debit card	Formal source of money	Informal source of money
Female	0.008	0.009***	0.005
	(0.005)	(0.005)	(0.004)
MC	-0.038***	-0.037***	-0.042***
	(0.005)	(0.003)	(0.002)
Female × MC	-0.032***	-0.025***	-0.026***
	(0.00)	(0.005)	(0.004)
N	498,088	492,316	770,517

Note: The table reports results from OLS estimation performed on different subgroups. Each column consider a different subgroup of students, according to their financial experiences. Column (1) includes students who had either a bank account or a debit card, column (2) those who receive money from a formal job, column (3) those who receive money from informal sources (gifts, pocket money). The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for age, immigrant background, socio-economic status, school type, reading and math literacy. Errors are clustered at the student level (in parenthesis), data are weighted and replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Heterogeneous effects: different questions topics

Question Topic:	(1) Interest, investing, income tax	(2) Other questions	(3) Diff
Female	-0.003	0.007*	*
	(0.006)	(0.004)	
MC	0.176***	-0.074***	***
	(0.005)	(0.003)	
Female × MC	-0.038***	-0.015***	**
	(0.007)	(0.004)	
N	254727	889078	

Note: The table reports results from OLS estimation performed on different subgroups. Column (1) are estimates for the set of questions on interest rate, investing, income tax and bank statement, column (2) for the remaining questions. Column three report the statistical significance of the difference between (1) and (2). The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for age, immigrant background, socio-economic status, school type, reading and math literacy. Errors are clustered at the student level (in parenthesis), data are weighted and replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Heterogeneous effects: standardized students' assessments in school

Assessment of student performance:	(1) Never or few times a year	(2) Monthly or more	(3) Diff
Female	0.009**	-0.007	
	(0.004)	(0.012)	
MC	-0.037***	-0.033***	
	(0.003)	(0.008)	
Female× MC	-0.026***	-0.018	***
	(0.004)	(0.011)	
N	687326	43745	

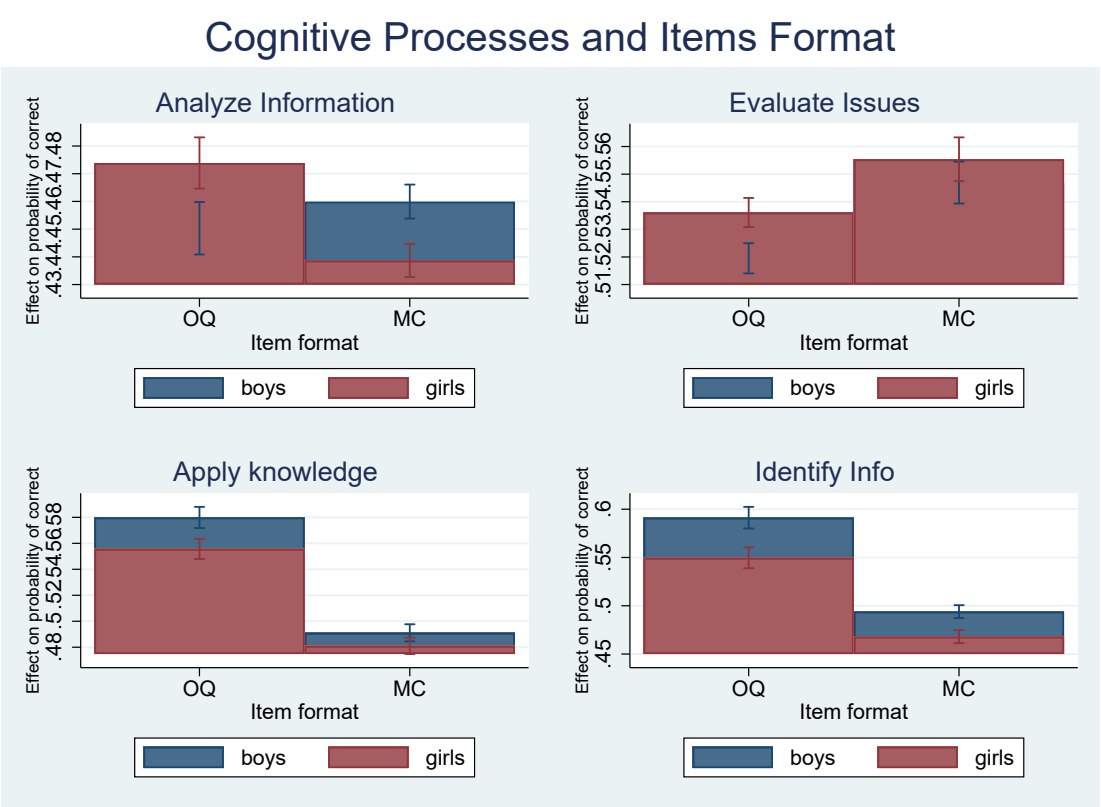
Note: The table reports results from OLS estimation performed on different subgroups, according to how many times the students in a given school are tested with standardized assessments. The exact working of the question is: "How often are students assessed with mandatory standardized tests?". The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for age, immigrant background, socio-economic status, school type, reading and math literacy. Errors are clustered at the student level (in parenthesis), data are weighted and replication weights are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Items Format, Content, Context and Process

	(1)	(2)	(3)	(4)
		Panel A: Content		
	Planning Managing	Money and Transaction	Risk and Reward	Financial Landscape
Female	0.022***	-0.005	-0.011**	0.012**
	(0.004)	(0.006)	(0.005)	(0.006)
MC	0.048***	-0.038***	-0.108***	-0.080***
	(0.004)	(0.004)	(0.005)	(0.005)
Female × MC	-0.018***	-0.022***	-0.016**	-0.042***
	(0.006)	(0.006)	(0.007)	(0.007)
N	378616	288257	244358	232574
		Panel B: Context		
	Home and Family	Education and work	Individual	Societal
Female	-0.012**	0.018***	0.002	0.026***
	(0.005)	(0.005)	(0.005)	(0.006)
MC	-0.003	0.060***	-0.085***	-0.334***
	(0.004)	(0.005)	(0.004)	(0.008)
Female× MC	0.014**	-0.030***	-0.034***	-0.121***
	(0.006)	(0.008)	(0.006)	(0.011)
N	436572	159111	420356	127766
		Panel C: Process		
	Analyze info	Evaluate issues	Apply knowledge	Identify Info
Female	0.024***	0.017***	-0.024***	-0.042***
	(0.007)	(0.004)	(0.006)	(0.008)
MC	0.010*	0.027***	-0.089***	-0.097***
	(0.005)	(0.004)	(0.005)	(0.006)
Female× MC	-0.045***	-0.008	0.014**	0.016*
	(0.007)	(0.006)	(0.007)	(0.009)
N	296504	385724	272506	189071

Note: The dependent variable is a dummy for whether the student has correctly answered question q. All regressions include a constant and controls for immigrant status, age, family socio-economic status and type of school. Errors are clustered at the student level (in parenthesis), data are weighted. Each panel split the sample according to the four categories the questions are classified with. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure B.4: Gender gap in financial literacy, by format and process of the questions



Notes: the figure reports linear predictions of Model 1 for males and female when answering open-response versus multiple-choice questions, by one of the four cognitive process categories. Each specification is controlling for age, immigrant status, school type, math and reading literacy and socio-economic status. Data are weighted to be representative of the PISA world population and standard errors are based upon BRR replication weights and clustered at the student level.

References

Akyol, S., J. Key, and K. Krishna (2016). Hit or miss? test taking behavior in multiple choice exams. *NBER Working Paper* (w22401).

Alessie, R. J., T. Bucher-Koenen, A. Lusardi, and M. Van Rooij (2021). Fearless woman: Financial literacy and stock market participation. *CEPR Discussion Paper* (No. DP15913).

Anaya, L., N. Iriberri, P. Rey Biel, and G. Zamarro (2021). Understanding performance in test taking: The role of question difficulty order. *CEPR Discussion Paper* (No. DP16099).

Arellano, A., N. Cámara, and D. Tuesta (2018). Explaining the gender gap in financial literacy: the role of non-cognitive skills. *Economic Notes: Review of Banking, Finance and Monetary Economics 47*(2-3), 495–518.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science 60*(2), 434–448.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics 116*(1), 261–292.

Ben-Shakhar, G. and Y. Sinai (1991). Gender differences in multiple-choice tests: the role of differential guessing tendencies. *Journal of Educational Measurement 28*(1), 23–35.

Borghans, L., B. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences 113*(47), 13354–13359.

Borgonovi, F. (2021). Is the literacy achievement of teenage boys poorer than that of teenage girls, or do estimates of gender gaps depend on the test? a comparison of pisa and piaac. *Journal of Educational Psychology*, Advance online publication.

Breland, H. M., D. O. Danos, H. D. Kahn, M. Y. Kubota, and M. W. Bonner (1994). Performance versus objective testing and gender: An exploratory study of an advanced placement history examination. *Journal of Educational Measurement 31*(4), 275–293.

Brunello, G., A. Crema, and L. Rocco (2021). Some unpleasant consequences of testing at length. *Oxford Bulletin of Economics and Statistics*, 1002–1023.

Burns, J., S. Halliday, and M. Keswell (2012). Gender and risk taking in the classroom. *Southern Africa Labour and Development Research Unit*.

Coffman, K. B. and D. Klinowski (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences 117*(16), 8794–8803.

Dahlbom, L., A. Jakobsson, N. Jakobsson, and A. Kotsadam (2011). Gender and overconfidence: are girls really overconfident? *Applied Economics Letters 18*(4), 325–327.

Driva, A., M. Lührmann, and J. Winter (2016). Gender differences and stereotypes in financial literacy: Off to an early start. *Economics Letters 146*, 143–146.

Ferber, M. A., B. G. Birnbaum, and C. A. Green (1983). Gender differences in economic knowledge: A reevaluation of the evidence. *The Journal of Economic Education 14*(2), 24–37.

Fonseca, R., K. J. Mullen, G. Zamarro, and J. Zissimopoulos (2012). What explains the gender gap in financial literacy? the role of household decision making. *Journal of Consumer Affairs 46*(1), 90–106.

Grohmann, A., T. Klühs, and L. Menkhoff (2018). Does financial literacy improve financial inclusion? cross country evidence. *World Development 111*, 84–96.

Hasler, A. and A. Lusardi (2017). The gender gap in financial literacy: A global perspective. *Global Financial Literacy Excellence Center, The George Washington University School of Business*.

Iriberri, N. and P. Rey-Biel (2021). Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. *European Economic Review 131*, 103603.

Lührmann, M., M. Serra-Garcia, and J. Winter (2015). Teaching teenagers in finance: does it work? *Journal of Banking & Finance 54*, 160–174.

Lumsden, K. G. and A. Scott (1987). The economics student reexamined: Male-female differences in comprehension. *The Journal of Economic Education 18*(4), 365–375.

Lusardi, A. and O. S. Mitchell (2008). Planning and financial literacy: How do women fare? *American economic review 98*(2), 413–17.

Lusardi, A. and O. S. Mitchell (2011). Financial literacy and planning: Implications for retirement wellbeing. Technical report, National Bureau of Economic Research.

Lusardi, A. and O. S. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of economic literature 52*(1), 5–44.

Mahdavi, M. and N. J. Horton (2014). Financial knowledge among educated women: Room for improvement. *Journal of Consumer Affairs 48*(2), 403–417.

Mondak, J. J. and M. R. Anderson (2004). The knowledge gap: A reexamination of gender-based differences in political knowledge. *The Journal of Politics 66*(2), 492–512.

Montolio, D. and P. A. Taberner (2021). Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. *Journal of Economic Behavior & Organization 191*, 1065–1090.

OECD (2013). *PISA 2012 Assessment and Analytical Framework-Mathematics, Reading, Science, Problem Solving and Financial Literacy*. France: OECD Publishing.

OECD (2016). *PISA 2015 Technical Report*. Paris: OECD Publishing.

OECD (2017a). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. Paris: OECD Publishing.

OECD (2017b). *PISA 2015 Results (Volume IV): Students' Financial Literacy*. France: OECD Publishing.

OECD (2020). *PISA 2018 Results (Volume IV): Are Students Smart about Money?* Paris: OECD Publishing.

Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization 115*, 94–110.

Riener, G. and V. Wagner (2017). Shying away from demanding tasks? experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review 59*, 43–62.

Rooij, M. V., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449 – 472.

Saygin, P. O. (2020). Gender bias in standardized tests: evidence from a centralized college admissions system. *Empirical Economics 59*(2), 1037–1065.

Saygin, P. O. and A. Atwater (2021). Gender differences in leaving questions blank on high-stakes standardized tests. *Economics of Education Review 84*, 102162.

Siegfried, C. and E. Wuttke (2019). Are multiple-choice items unfair? and if so, for whom? *Citizenship, Social and Economics Education 18*(3), 198–217.

Skagerlund, K., T. Lind, C. Strömbäck, G. Tinghög, and D. Västfjäll (2018). Financial literacy and the role of numeracy–how individuals' attitude and affinity with numbers influence financial literacy. *Journal of behavioral and experimental economics 74*, 18–25.

Taylor, C. S. and Y. Lee (2012). Gender dif in reading and mathematics tests with mixed item formats. *Applied Measurement in Education 25*(3), 246–280.

Wuttke, E., C. Siegfried, and C. Aprea (2020). Measuring financial literacy with a situational judgement test: do some groups really perform worse or is it the measuring instrument? *Empirical Research in Vocational Education and Training 12*(1), 1–21.