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

Despite women's and men's convergence in many economic outcomes, women consistently display lower levels of financial literacy than men, a finding that is widespread across many countries and contexts (Hasler and Lusardi, 2017). To the extent that financial literacy is a driver of financial inclusions (Grohmann et al., 2018) and more savvy financial behaviors[1] (Rooij et al., 2011; Lusardi and Mitchell, 2014, among others), it is a policy-relevant goal to understand the extent of women's disadvantage in financial knowledge and the reasons behind it. Gender differences in financial literacy levels have been hard to explain, and scholars have not yet found a definitive answer (Lusardi and Mitchell, 2014), even though previous literature has explored factors such as marital status, educational levels, labor force participation, non-cognitive skills, and expectations (Arellano et al., 2018; Driva et al., 2016). What has not yet received much attention in the financial literacy gender gap literature is *the way* financial literacy is measured and how this may relate to the observed differential patterns across genders. This essay, based on a chapter of my PhD thesis, aims at filling this gap.

Analyzing the difference in gaps related to the use of multiple-choice testing formats versus open-response ones, I question whether the tools used to measure financial literacy can themselves amplify the gender gap in performance. In order to do so, I employ the financial literacy module of the PISA 2015 assessment, hence focusing on gender differences among country representative samples of 15th years old students.

Since its introduction in the 2004 Health and Retirement Study (Lusardi and Mitchell, 2011), the standard measure of financial literacy in empirical studies consists of a set of 3 to 7 multiple choice questions testing the understanding of basic concepts such as interest, inflation, and risk diversification. The reason behind the use of multiple-choice to measure literacy levels is relatively straightforward, as it allows to effectively and objectively assess large fractions of the population (Wuttke et al., 2020). In the specific financial literacy context, the test proposed by Lusardi and Mitchell is particularly convenient, as it can easily be added to existing household surveys with a relatively low organizational effort. Unlike the vast majority of financial literacy surveys available, PISA consists of a weighted mix of open-response and multiple-choice test items. It also does not report 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). From the Economics Education literature, there is evidence that there exist dependencies between respondents' test results and test format, with multiple-choice testing tending to favor males over females (and the opposite for constructed-response questions, Lumsden and Scott, 1987, Ferber et al., 1983), and still, as pointed out by Wuttke et al., 2020 research on the issues of test bias in financial literacy assessments is scarce.

In the chapter of my PhD thesis, I exploit the unique characteristics of PISA data to identify the effect of question formats on gender gaps in financial literacy. 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). 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

---

[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)

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). Having answers to both multiple-choice and open-response questions for each student allows to estimate within-student effects of question format on the probability to answer correctly financial literacy test items for boys and girls. 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. This specification also allows to control for student-invariant characteristics, such as unobserved cognitive skills, family background, and underlying non-cognitive skills.

As a first result, I find that while on average 15-years-old girls do not show lower overall financial literacy as compared to boys (in line with OECD, 2015}), a gap is found when differentiating between multiple-choice and open-response questions. More specifically, girls' likelihood to correctly answer financial literacy test items is about 5% lower when the question is formulated in a selected-response format. In contrast, no gap is found for constructed-response ones. Secondly, the result is highly robust to the addition of several other variables that may differentially influence the ability of the students to answer a question correctly across genders. First, I add controls for the students' cognitive skills, namely the math and reading PISA test results, as 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, 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). I also progressively add country, region, and school fixed effects, to take into account overall differences in school systems around the world, which may systematically affect students in a differential way across gender. Finally, I introduce students' fixed effects, capturing individuals' fixed characteristics such as innate ability, family background, and non-cognitive skills. Moreover, I suggest that the gender-by-format effect appears to be explained by the cognitive process and mental strategies that test-takers need to engage in to answer assessment items. 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.

The population analyzed in PISA is quite different from the one typically tested in household surveys. In light of that, the financial literacy construct in PISA may be hardly comparable to the one tested in household surveys, making the results difficult to generalize outside the school context. However, I also provide evidence that the results, even on the student sample, are likely to be consistent, and hence of interest, when the material assessed relates to topics such as inflation and investing, typical of the household finance setting. Hence, the results are quite relevant from a policy perspective, even if it is worth noting that our identification strategy, based on within-student variation, has some limitations compared to the strength of the causal evidence that one may draw from experimental settings.

My project contributes to the existing literature in different ways. First, it provides an alternative explanation to a well-consolidated finding in the household finance literature and rises awareness on the general way financial literacy is tested in household surveys. The existence of a persistent gender divide in financial knowledge, across surveys and countries, has been examined from different perspectives, considering traditional socio-economic factors, such as educational level or marital status (Grohmann et al., 2018, Bucher-Koenen et al., 2021, Mahdavi and Horton, 2014 among others) and within household specialization (Fonseca et al., 2012). A different set of works have rather focused on the importance of personality and non-cognitive traits, (Arellano et al., 2018), such as differences in individual disposition and interest in financial issues (Luhrmann et al., 2015) or in expectations (Driva et al., 2016). However, available research has also suggested that a definitive answer has not yet been found, and a considerable portion of the differences in performances remain unexplained (Fonseca et al., 2012, Lusardi and Mitchell, 2014, Alessie et al., 2021).

My work also speaks to the growing literature on the role of standardized assessments and on the question of what they precisely measure, as personality traits and non-cognitive skills have been found to be relevant on the final test performance across different groups of individuals (Borghans et al., 2016, Borgonovi, 2021, Brunello et al., 2021, Anaya et al., 2021). A rich literature on Economics of Education has found a gender difference in performance in multiple-choice tests, in part explained by women higher tendency to skip questions in standardized tests and by overall gender differences in non-cognitive traits such as willingness to guess, confidence and risk-aversion (Baldiga, 2014, Alessi et al., 2021, Iriberri and Rey-Biel, 2021, Brunello et al., 2021, Pekkarinen, 2015), while studying how different penalties schemes on wrong answers can reduce the gender gap in questions skipped (Riener and Wagner, 2017, Coffman and Klinowski, 2020, Saygin and Atawer, 2021). There is also limited evidence for subjects other than mathematics. A paper that is closely related to this work is Siegfried et al., 2019, which tested economic knowledge on a sample of roughly 200 German students using constructed- and selected-response formats for similar question contents. Their analysis suggests that a mix of test formats can compensate for gender differences, although they find no evidence of a gender differential effect of item formats. These studies provide valuable insights into the issue of gender differences in answering different question formats. However, they are often affected by issues related to small sample size, the inability to compare directly multiple-choice and open-response items, or the fact that they focus on specific subjects' samples and high-stake contexts (such as entrance and university exams). To the best of my knowledge, this is the first large-scale empirical analysis that attempts to extend the validity of such previous studies while investigating the interaction between gender and question formats in the financial literacy domain. The results are quite relevant, especially in light of the very 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. Moreover, because of the data we employ, we are able to highlight a mechanism behind our results that previous studies had not been able to investigate[2], i.e., the multidimensionality of test items and how questions' characteristics other than the format (in our case its content, context or process) can also interact with the gender dimension and explain the observed gender-by-format effect. Understanding the dimensions of financial literacy the tests in use measure and how specific questions formats can favor certain groups in the population is crucial to fighting the existing inequalities.

From a policy perspective, my results are relevant for several reasons. First, given the methodology almost exclusively 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

---

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

easy-to-implement nature, I 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)). My 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 the analysis, I 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, the 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.