Questioni di Economia e Finanza

The glass drop ceiling: composition effects
or implicit discrimination?

by Claudia Biancotti, Giuseppe Ilardi and Clair Lavinia Moscatelli

Questioni di Economia e Finanza

(Occasional papers)

The glass drop ceiling: composition effects
or implicit discrimination?

by Claudia Biancotti, Giuseppe Ilardi and Clair Lavinia Moscatelli

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it *.*

THE GLASS DROP CEILING: COMPOSITION EFFECTS OR IMPLICIT DISCRIMINATION?

by Claudia Biancotti*, Giuseppe Ilardi* and Clair Lavinia Moscatelli**

Abstract

We analyze data from the Bank of Italy's most recent recruitment rounds, in an effort to explain why men consistently score better than women. We focus on the pre-screening stage of the hiring process, a multiple-choice test, where men acquire a preliminary advantage. After observing a higher incidence of questions left blank for women, and a negative correlation between the share of unanswered questions and the final score, we run an experiment on scoring formulas to check for implicit discrimination linked to risk aversion; no evidence of such discrimination is found. Based on a follow-up questionnaire, we also study the role of composition effects. Nearly 40 per cent of the gap in test scores depends on the quality of the candidates: male graduates appear to self-select into the applicant pool more frequently than females do. A further 34 per cent is explained by the fact that the same characteristics tend to produce different effects based on gender. The remaining 26 per cent remains unexplained.

Contents

_____

* Bank of Italy, Economic and Financial Statistics Department.

** Bank of Italy, Human Resources Department.

1 Introduction

Gender[1] gaps still exist in most labor markets. According to the 2011 Statistical Abstract (U.S. Census Bureau, 2011), in 2009 the labor force participation rate for American males was 72 per cent, against 59.2 per cent for females [2]. Men working full-time earned a median weekly wage of 819 dollars, 24.6 per cent more than women (657 dollars); they also filled the vast majority of management positions. The OECD Employment Outlook for 2010 (OECD, 2010) shows that during the recent crisis women, along with the young and the unskilled, were most liable to lose their jobs. Several international sources confirm these asymmetries: in most countries if not all, women are less likely to work, keep their job through difficult times, draw high salaries, and reach peak positions.

Differences in participation, earnings and career paths are only partially accounted for by composition effects, i.e. heterogeneity in the level of education, experience, and effort. The residual, unexplained gaps may reveal the existence of discrimination, either explicit or implicit. Explicit discrimination can be defined as a conscious preference for men on the part of employers: a positive value is intentionally associated to maleness in itself, resulting in men being hired with more ease than women, or being paid more for identical jobs. Implicit discrimination can be defined as the result of hiring or wage-setting behavior that, while formally independent of gender, inadvertently favors men for reasons unconnected to job-relevant skills. It comes in different varieties, whose common factor is lack of intentionality: unconscious gender bias on the part of recruiters, working hours that are unappealing for mothers and lack of employer-sponsored childcare facilities, personnel selection techniques that reward psychological traits more prevalent among males and not relevant for the positions being offered.

Explicit discrimination is illegal in many countries; while it is not extinct, at least it is deterred by the law.[3] Implicit discrimination, on account of its unintentional and indirect nature, presents a bigger challenge. First of all, it must be identified: a daunting task, since it often works through subtle mechanisms that leave no *prima facie* evidence. Then it must be eliminated: in the absence of a culprit with malicious intent, this generally requires a complex mixture of actions, some of which must touch upon slow-moving cultural processes.

In this paper, we present an exercise in the analysis of gender gaps and the investigation of implicit discrimination. We look at the recruitment of young professionals at the Bank of Italy, which presents an interesting peculiarity: for the past fifteen years, the male-female ratio has consistently hovered around 0.7 for applicants, and 1.5 for new hires. Most of the reversal occurs during the first stage of the hiring procedure, a multiple-choice test. We find that composition

---

[1]A previous version of this paper was presented at the conference "Women and the Italian Economy" organized by the Bank of Italy, held in Rome on March 7th, 2012. The views expressed therein are those of the authors and do not necessarily reflect those of the Bank of Italy.

[2]For Italy, it was 51.9 per cent and 38 per cent respectively

[3]In the US, sexist hiring and pay schemes are forbidden by the Equal Pay Act of 1963 and the Civil Rights Act of 1964; in the EU, Council Directive 75/117 of 1975 mandates equal pay for equally qualified workers.

effects related to job-relevant variables explain approximately 40 per cent of the performance gap, while a further 34 per cent accrues on gender-specific effects of the same variables. However, a small unexplained gender gap in performance remains that might indicate implicit discrimination. We test the hypothesis of non-neutrality of the test with respect to two psychological traits that are known to be correlated to gender, but are not indicated by recruiters as relevant to the job: risk aversion and self-esteem. We find no evidence of discrimination with respect to risk aversion; the results on self-esteem are inconclusive, prompting further research. Section 2 reviews related literature. Section 3 presents the hiring procedure, an overview of the data and some preliminary evidence. Section 4 illustrates the main results. Section 5 concludes.

2   Related literature

Our study draws on four different bodies of research, respectively concerned with implicit discrimination, gender differences in risk aversion, gender differences in self-esteem, and multiple-choice tests.

2.1   Implicit discrimination

In the economic literature, work on implicit discrimination has so far focused on subconscious bias, i.e. involuntary, unverified association of traits such as ethnic origin and gender with employability and work performance (Bertrand et al., 2005). Many recent studies are based on implicit association tests (IATs) (Greenwald et al., 1998; Sriram and Greenwald, 2009). IATs are laboratory experiments requiring subjects to look at words and/or pictures flashing on a screen and associate them under a tight time constraint following certain rules; these can be neutral (e.g. "click on all words written in blue") or correlated to known cultural biases against certain groups, either directly (e.g. "click on all words related to family" displayed alongside the picture of a woman) or inversely (e.g. "click on all words related to work"). Researchers examine response times and associations that violate the rules in order to detect whether and how much subjects are affected by certain prejudices. According to Mo and Weiksner (2009), voters who score highly on measures of implicit association between gender and leadership tend to choose representatives of the preferred gender beyond what candidate qualifications would predict. Levinson and Young (2010) administer two IATs to a pool of law students, one on gender and the legal profession, one on gender and career choices. Both experiments show the existence of subconscious prejudice: students generally report low scores on the Modern Sexism Scale, a measure of conscious preference for traditional gender roles, but under time pressure they still tend to associate "man" with "judge" and "woman" with "paralegal" or "homemaker" even when instructed to do the contrary. Latu et al. (2011) use IATs to highlight a discrepancy between explicit appreciation and implicit mistrust of female managers, especially on the part of men. Beaman et al. (2009) provide evidence of association of female names with domestic tasks, and male names with leadership positions; they also show that exposure to female leaders can

attenuate such preconceptions, consistently with earlier findings by Dasgupta and Asgari (2004). The Gender/Science IAT, one of the most popular experiments, consistently shows that males are assumed to be more proficient at science and mathematics than females; the stereotype might be self-fulfilling in that it discourages women from investing in the study of these subjects (Nosek et al., 2009).

Attention has also been given to the potential biasing effect of certain environmental conditions and evaluation procedures, although the expression "implicit discrimination" has not been used in this context. Hunter (1997) talks of "indirect discrimination" when presenting a court case originating from the selection of a male applicant on the part of an all-male hiring panel, to the detriment of a female applicant who appeared better qualified. She argues that the decision, while not taken on the basis of gender per se, stemmed from a preference for interviewing styles more common among men, and hence more familiar to the panel. Charles et al. (2009) find a positive correlation between male sexism, as measured by attitude questions in the General Social Survey, and the relative outcomes of women in local labor markets, both in terms of employment rates and wage gaps. Marx and Roman (2002) show that girls perform better in mathematics when tests are administered by a positive role model, i.e. a female teacher perceived as competent in the subject. Boyd et al. (2010) document that male judges are about ten per cent less likely than female judges to rule in favor of the litigant in sex discrimination cases, possibly because of an implicit predisposition to dismiss some accusations as instrumental or preposterous. Peresie (2005) provides similar results for collegial decision-making on the Federal Appellate circuit.

A few studies exist on appropriate policy responses to implicit discrimination. Jolls and Sunstein (2006) argue in favor of "debiasing through law": involuntary behaviour cannot be punished, but regulators can work on debunking the underlying, unconscious assumptions, much in the same spirit of mandating tobacco companies to print health warnings on cigarette packs in order to make smokers more aware of the consequences of their actions. Blair et al. (2001) focus on the effectiveness of counterstereotypic mental imagery, making the case for the relatively malleable, reversible nature of some implicit associations.

2.2 Gender differences in risk aversion

A strand of the vast literature on risk aversion focuses on gender differences. Several studies find that males are more likely to engage in risky behavior than females. While the evidence on substance abuse, criminal behavior, and perception of risks connected to both catastrophic events and standard activities appear to be solid, results on financial risk-taking are somewhat mixed (see Eckel and Grossman, 2008, for a meta-analysis). Jianakoplos and Bernasek (1998) examine household portfolios, showing that males have a stronger preference for risky assets compared to females. Schubert et al. (1999) find a significant gender gap in risk aversion in the context of abstract gamble experiments. Hinz et al. (1996) show that American women invest their pension

assets more conservatively than men. Sapienza et al. (2009) link risky financial behavior with the level of circulating testosterone, which is generally higher in males. Booth and Nolen (2012) argue that gaps in risk aversion may depend on socialization instead: in an experimental context, young women from single-sex schools appear to be more risk-prone than their coed peers, suggesting that cautious behavior may enter the construction of female identity by contrast rather than naturally.

2.3   Gender differences in self-esteem

Self-esteem was originally defined by James (1890) in these terms: "[O]ur self-feeling in this world depends entirely on what we back ourselves to be and do. It is determined by the ratio of our actualities to our supposed potentialities; a fraction of which our pretensions are the denominator and the numerator our success". Subsequent studies in psychology have expanded the concept to encompass "the level of global regard that one has for the self as a person" (Harter, 1985), "individuals' experience that they are appropriate to life and to the requirements of life" (Branden, 1990), and "the positivity of the person's self-evaluation" (Baumeister, 1998). Different schools of thought describe the origin of self-esteem in different terms. The definition proposed by James is at the root of what is now known as the competencies model, according to which people draw self-esteem from their objective achievements, mediated by a subjective evaluation on the importance of each achievement. The reflected appraisal model, pioneered by Cooley (1902), conversely posits that people derive self-esteem from social interactions: what ultimately matters is acceptance on the part of others, possibly filtered through subjective perception. Some works bring the two approaches together: for example, Greenberg et al. (1986) argue that "self-esteem derives from succeeding at what is valued in a given socio-cultural niche".

The existence of a "confidence gap" has been the subject of much attention, both in academic research and the public debate. Mary Pipher (1994), drawing on her work as a psychotherapist for adolescent females, first brought to the attention of the general public that girls tend to feel worse about themselves compared to boys. In a large meta-analysis, spanning more than 145,000 individuals, Kling et al. (1999) show that indeed males have higher global self-esteem than females, especially during the transition to adulthood, although the difference is smaller than generally believed.

Where does this gap come from? As a general background element, most parents encourage gender-typified behavior (Lytton and Romney, 1991), which leads to children "internalizing the gender ideology of their larger culture" (Tenenbaum and Leaper, 2003). Lamke (1982) shows that self-assessed masculinity is a predictor of high self-esteem for both males and females. If cultural norms exist that place special positive value on maleness, men will think more highly of themselves compared to women, especially when their convictions resonate with expected life outcomes (see for example Fortin, 2005, for the negative cross-country correlation between progressive gender role attitudes and wage gaps). In a study of clinical depression, a condition often accompanied by very

low self-esteem, Nolen-Hoeksema (2001) argues that women are simultaneously more exposed to stressful life events and less equipped to deal with them compared with men, both because of biological reasons and socialization. Disregulation of the hypothalamic-pituitary-adrenal axis, which governs stress responses, has higher prevalence among women; also, coping ability is emphasized as very important in the upbringing of males, less so for females. McMullin and Cairney (2003) show that females are less self-confident than males throughout the life span, but age and social class affect the size of the gap, suggesting that self-esteem may be a proxy of relative powerlessness of a given group.

Josephs et al. (1992) define a general framework where social expectations lead men and women to extract self-esteem from different sources: men evaluate themselves mainly based on autonomy and individual achievements, women put greater emphasis on connectedness to others. Variations on this idea are incorporated in several models of self-concept definition (Banaji and Prentice, 1994; Cross and Madson, 1997; Benenson and Heath, 2006). The role of culturally prescribed behavior as a component of the sense of self was also recently picked up by economists, first in the context of occupational choice, then as the foundation of a broader "identity economics" approach (Akerlof and Kranton, 2010; Bénabou and Tirole, 2011). Drawing on these models of differential trait relevance, some authors claim that women's poor self-esteem can be partly traced back to body image dissatisfaction (Kostanski and Gullone, 1998; Tiggemann, 2005), an issue not as relevant to men both because success of males in dyadic relationships does not depend as heavily on looks, and because men's self-worth is less affected by this specific type of interdependence compared to women (Gabriel and Gardner, 1999).

2.4 Multiple-choice tests

Several papers exist on the relationship between scoring formulas, chance and performance in multiple-choice tests (see for example Burton, 2001; Bush, 2001; Simkin and Kuechler, 2005; Jennings and Bush, 2011). Guidelines are also provided by various actors in the education and academic community, such as the National Council on Measurement in Education and the National Board of Medical Examiners, on how to maximize the knowledge-signaling value of an instrument that presents definite advantages in terms of efficiency but is also exposed to the distortionary effects of lucky guesswork. The potential gender bias of this instrument has also been studied: a male advantage was detected in some experiments, but not in others (Simkin and Kuechler, 2005; Chan and Kennedy, 2002; Ng and Chan, 2009).

2.5 Previous work similar to ours

To our knowledge, the four areas of study referenced above have seldom been brought together in a systematic way. Correlations between risk aversion, self-assessed competence and scores in multiple-choice tests have been sketched by Walker and Thompson (2001) in a field experiment related to

the adoption of non-traditional response rules. Espinosa and Gardeazabal (2007) study rationality, risk aversion and the effect of gender on test performance, conditional on scoring formulas. A largish literature is concerned with how gender differences in attitudes toward competition, risk and stress may put women at a disadvantage in the classroom and on the job. Lorenzo et al. (2006) show that the achievement gap in physics between male and female undergraduates can be partially traced back to competitive, non-interactive teaching methods. Niederle and Vesterlund (2007, 2011) examine gender differences in taste for competition and the phenomenon known in the psychological literature as "stereotype threat", showing how men's preference for competitive settings and women's fear of confirming negative preconceptions might play a role in explaining why the former outperform the latter in math tests.

3 Background: hiring procedure and gender gaps, 1998-2009

Almost every year, the Bank of Italy opens a number P_y of entry-level positions for young professionals. Eligibility is restricted to candidates under 40 with a degree in economics, law, political science or similar fields, with a final overall grade of at least 105 over 110. Applications are submitted online through a form requesting basic demographic and academic information.[4] The first stage of the recruitment process entails a multiple-choice test with 120 questions organized in three sections: specialist topics, foreign languages and logic. The questions change for each recruitment round; they are drawn from a pre-existing superset, following criteria aimed at ensuring that the level of difficulty of the test remains approximately constant over time. The test is a pre-screening instrument, aimed at filtering very large applicant pools, and only serving to determine who goes on to the actual competitive selection procedure (usually, those applicants who obtain the best $P_y * 10$ scores). Hires are then decided based exclusively on the sum of grades achieved in a written exam with essay-type questions and a panel interview. Those who surpass a certain threshold are drafted as potential hires; jobs are offered starting from the best performer and following in descending sequence the ordering of grades, until all P_y positions are filled or the end of the list of potential hires is reached.

Over the period 1998-2009, 61.5% of applicants were female, a slightly higher percentage compared to the 56.2% share of graduates in relevant fields in the general population.[5] However, women only accounted for 35.5% of new hires (Table 1).

In the multiple-choice test, which appeared to discriminate between genders more strongly than the remaining stages, women appeared to do worse across the board: the ratio between average female and male scores was respectively 87.5%, 88.2% and 88.1% in the three sections. Regressions controlling for all observable characteristics provided in the application form did not give sufficient insight into the phenomenon. Their results were mostly to be expected, e.g. applicants with a

---

[4]Date and place of birth, current residence, type of degree, awarding university, final overall grade.
[5]Source: OECD database at `http://stats.oecd.org/`.

history of good grades or coming from prestigious colleges were more likely to pass the test. This notwithstanding, being male remained a significant predictor of success (Table 2) [6].

4 Results

Starting from the 2010 selection rounds for law graduates, we investigated the determinants of the gender gap with two complementary instruments: an experiment involving a modification of the scoring formula, and a voluntary follow-up survey on education history, work experience, family background, motivation, strategies deployed to prepare for the test, and selected psychological traits (for the full text of the questionnaire, see Appendix A). The results presented in sub-section 4.1 refer to a total of 2,441 applicants, i.e. everyone who participated in the 2010 and 2011 multiple-choice tests for law graduates; those in sub-section 4.2 refer to the 1,156 test-takers who participated in the survey (Table 3).[7]

4.1 Scoring formula modification

Up to 2009, the multiple-choice test was scored according to the following formula: +1 point for each correct answer, -0.3 points for each incorrect answer, 0 for each missing answer. Considering that each question has four response options, and test-takers are allowed to either pick a single option or leave the question blank, the expected value of a random guess was $(1 - 0.9) * 0.25 > 0$. Rational respondents should have always provided an answer, independent of their level of knowledge. However, missing items were observed, reflecting lack of time and/or risk aversion. The share of missing items appeared to be negatively correlated with the final score, and higher for females. Based on the literature referenced in Section 2, it was hypothesized that the test might implicitly discriminate women by rewarding risk-taking behavior more prevalent in males and independent of competence. Starting from the 2010 recruitment round, the scoring formula was modified to +1, -0.7 and 0 points for correct, incorrect and missing answers respectively: the gender gap in the share of missing items closed, but the differences in overall performance remained (Tables 4 and 5). These results might imply that, while differential risk aversion might affect response behavior, the gap in performance observed in the past did not accrue on lucky guesses on the part of male applicants.

4.2 Follow-up questionnaire

The follow-up questionnaire provides previously unavailable insight on gender differences in the applicant pool (Table 6), and how they affect performance (Tables 7 and 8). Males are more likely

---

[6] Quintile regressions produced similar results, suggesting that the effect of gender is homogeneous across the skill distribution

[7] No systematic differences were found between respondents and non-respondents with respect to demographic variables available for both sets. On average, respondents achieved slightly higher scores in the test, which is explained by the fact that the population of non-respondents is composed almost entirely of rejected applicants.

to come from Northern Italy, an area associated with better quality of schools and universities, higher social capital, and lower unemployment rates. The percentage of men whose parents have university degrees is higher compared to women, a relevant element in a country where social mobility is historically low and transmission of culture within the household is a key component of education. Considering that place of birth and parental education can safely be considered exogenous with respect to gender, these results appear to indicate that more skilled men self-select into applying for the Bank of Italy compared to skilled women.

Males are also more likely to hold a postgraduate degree and a full-time job, which might indicate better market-relevant abilities, greater determination, or both. It could, however, also reflect the presence of a culture-wide gender bias, whereby male students are more encouraged to pursue advanced degrees compared to female students, and are more likely to be hired once they graduate.

When asked to choose from a list of motivations for applying, both genders focus on reputation and professionalism; men are more responsive to salary expectations, women to guarantees of job security. Consistently with this fact, a higher share of women reports applying for another government job; however, these applications were successful less often compared to those of men.

Where evaluation of the test is concerned, males are less likely than females to report that they found the questions very difficult, and that the time allotted to answer was insufficient.

Indications on risk aversion are mixed. On the one hand, when presented with two standard lottery questions, males show a definite taste for riskier options; on the other hand, when asked to describe their behavior during the test, they state more often than women that they only answered items on which they were certain or almost certain.

Considering that both "difficulty of the test" and "certainty about the correct response option" in this context are subjective, self-esteem may come into play in explaining these differences. For an identical level of knowledge, self-confident test-takers might feel that the test was easier compared to self-doubting ones, and might feel certain that they are right more often than those who have lower trust in their intuition. Unfortunately, self-esteem scores as measured by the questionnaire can neither confirm nor deny the existence of this correlation: the variance is minimal on the whole sample, probably signaling measurement error so significant that the measure is deprived of any informational power (See Appendix B for a critical discussion of our self-esteem measure and a sketch of a testable hypothesis).

In order to assess the impact of composition effects on performance, we estimate two models on the pool of applicants who participated in the survey. Table 7 presents the results of a linear model where the dependent variable is the final score obtained in the test, normalized with respect to the distribution of scores observed for the specific round of recruitment in order to control for small misalignments in difficulty levels of the test and in average quality of each vintage of applicants. Table 8 presents the results of a logistic model where the dependent variable is a

dummy indicating whether an applicant passed the test or not. The two sets of estimates offer a common picture, although significance levels for individual variables are generally higher for the linear model on account of higher precision in measuring the outcome. Applicants who reside in Northern or Central Italy, and those who graduated at universities in these areas, do better than those who reside and/or studied in the South. Obtaining a degree at a young age and with a distinction is a predictor of success, as is a modicum of experience accumulated prior to sitting the test. Too long a gap between first degree and application is, however, negatively related to scores. Those who state they did not have time to prepare for the test in any way did worse than others; cautious test-takers, choosing to only answer questions they were certain or almost certain about, obtained higher scores. Risk-lovers also have an advantage. These results are consistent with men performing better than women, based on the different composition of the two sub-populations discussed above. On account of the aforementioned measurement issues, self-esteem scores are not significant, even when broken down in sub-scores evaluating respectively the propensity to agree with positive and with negative statements about one's self.

Gender in and of itself still remains, somewhat problematically, a significant determinant of achievements: when all other characteristics are controlled for, female applicants score approximately 0.26 normalized points less than males. The coefficient is 27 per cent lower, in absolute terms, compared to the coefficient estimated for a specification that does not include the variables collected in the supplementary questionnaire.

The Blinder-Oaxaca decomposition for the linear model including all available covariates shows that approximately 40 per cent of the total gender gap in performance is attributable to composition effects (Table 9); 34 per cent can be explained with differences in coefficients for single-gender models, i.e. different effects of the same covariates on final scores for male and female test-takers; the remaining 26 per cent is the unexplained component, signifying either composition effects for skills relevant to the job but not observed through the questionnaire, or implicit discrimination acting through an unidentified channel.

5 Conclusions

In this paper, we analyzed data from the Bank of Italy's 2010 and 2011 recruitment rounds for law graduates, in an effort to explain why men consistently perform better than women. We focused on the pre-screening stage of the hiring process, a multiple-choice test determining who is admitted to a competitive selection procedure consisting of an essay-type written exam and a panel interview; men typically do better at this stage, acquiring an advantage that is then maintained throughout. The test consists of 120 questions with four response options each; test-takers can either choose a single option or leave the item unanswered. After observing a higher incidence of missing items for women, and a negative correlation between the share of missing items and the final score, we checked for implicit discrimination related to gender differences in risk aversion

through an experiment: the scoring formula was changed so as to increase the cost of selecting response options randomly. While the heterogeneity in response behavior disappeared, the gap in performance did not, suggesting that the difference in scores did not accrue on lucky guesswork on the part of risk-loving men.

We also distributed a follow-up questionnaire to test-takers, in an effort to gather insight on composition effects related to job-relevant traits, and to ascertain whether implicit discrimination related to gender differences in self-esteem was at play. The survey was answered by approximately 50 per cent of the applicant pool. We found that nearly 40 per cent of the gap in scores depends on quality factors: male applicants are more likely to possess traits that predict success, such as holding a degree from a good university, being in full-time employment, coming from a highly educated family, preparing for the test thoroughly, and choosing to answer only questions they were certain or almost certain about. These results suggest that skilled males self-select into the applicant pool more frequently than skilled females; further investigation is needed on the reasons why. A further 34 per cent is explained by the differential impact of the same covariates for men and women. The remaining 26 per cent is hitherto unexplained, signifying either composition effects for yet unobserved skills relevant to the job, or implicit discrimination acting through an unidentified channel. Minimal variance in self-esteem scores, probably dependent on social desirability bias affecting respondent behavior, prevented us from testing any hypothesis related to the effects of this psychological trait; more research is needed, after acquiring data through instruments that are relatively immune to conscious manipulation on the part of respondents (e.g. implicit association tests).

A The questionnaire

<div align="center">General background & education</div>

1. Where do you live?

   a. North

   b. Centre

   c. South/Islands

2. Did you attend university away from your home residence? (yes/no)

3. If yes, for what reason?

   a. In order to attend a prestigious university

   b. The faculty I had chosen was not available near my home

   c. Other

4. Have you got children under the age of 14? (yes/no)

5. If yes, how many?

6. Have you achieved or are you working towards a MSc/MA relevant to one or more subjects in the recruitment notice? (yes/no)

7. Have you achieved or are you working towards a PhD relevant to one or more subjects in the recruitment notice? (yes/no)

8. Have you attended professional specialization courses / have you been in professional training (i.e. legal practice)? (yes/no)

9. Have you achieved a professional certification (i.e. lawyer or accountant)? (yes/no)

10. Have you recently studied for other public-sector positions in subjects relevant to your degree? (yes/no)

11. Have you already been hired for other public-sector positions in subjects relevant to your degree? (yes/no)

12. What is your current working status?

    a. Full time

    b. Part time

    c. Occasional

d. Unemployed

e. Student

f. Other

13. Are you working in the public sector? (yes/no)

14. Does one (or both) of your parents work in the public sector? (yes/no)

15. Does one (or both) of your parents have a university degree? (yes/no)

Preparation effort & evaluation

16. How did you prepare for the test?

   a. I studied all the subjects and practiced answering multiple choice tests

   b. I studied all the subjects but didn't practice answering multiple choice tests, because I had previous experience with them

   c. I studied all the subjects but didn't practice answering multiple choice tests; I had no previous experience with them

   d. I didn't have time to prepare for the test

17. Regardless of the results you achieved, how difficult did you find the test?

   a. Not at all difficult

   b. Not very difficult

   c. Quite difficult

   d. Very difficult

18. Which section of the test did you find most difficult?

   a. Logic

   b. Law

   c. Foreign language

   d. None in particular

19. How did you perceive the amount of time given for the test?

   a. Not enough

   b. Just enough

   c. More than enough

20. When answering did you follow the order in which the questions were presented? (yes/no)

21. Which of the following statements best describes the strategy you followed in answering?

    a. I only answered the questions of which I was absolutely sure of the answer.

    b. I answered the questions of which I was quite sure of the answer.

    c. I tried to answer most of the questions even if I wasn't sure of the answer

    d. I tried in any case to answer all of the questions.

Motivation

22. Please order the following items on the basis of how strongly they influenced your decision to apply to this recruitment notice (from the most important to the least important):

    ☐ Job security

    ☐ Salary

    ☐ Reputation of the Institution

    ☐ Work-life balance

    ☐ Level of professionalism guaranteed

    ☐ Job in line with my studies

    ☐ Interest in the functions carried out by the Central Bank

23. Which of the following functions are you interested in (multiple answers are allowed):

    a. Monetary policy

    b. Banking supervision

    c. Economic research

    d. Other

24. For the same wage and the same job where would you prefer to work?

    a. Public sector

    b. Private sector

    c. No preference

Personality traits

25. Some of the literature on multiple-choice tests state that the results achieved can depend also on innate features of the candidates. The following questions test this aspect.

    Imagine you are playing a game in which there are 2 wheels of fortune and you have to choose which to spin. Each wheel has 100 numbers and the probability that a number is drawn is always 1/100. You can choose only one wheel and spin it only once. You have the following two possibilities: Wheel A: numbers from 1 to 10 correspond to a prize of 500.000, numbers from 11 to 100 don't win anything. Wheel B: numbers from 1 to 5 correspond to a prize of 1.000.000, numbers from 6 to 100 don't win anything. Which would you choose? (wheel A, wheel B, it makes no difference)

26. If, instead, you had the following 2 options: Wheel C: numbers from 1 to 100 correspond to a prize of 100.000. Wheel D: numbers from 1 to 9 correspond to a prize of 500.000, numbers from 10 to 97 win 100.000, numbers from 98 to 100 don't win anything. Which would you choose? (wheel C, wheel D, it makes no difference)

27. Below are 10 statements that describe the opinion you have of yourself; it is a test introduced by M. Rosenberg. For each statement please indicate your level of agreement: I agree, I strongly agree, I disagree, I strongly disagree.

    - I feel that I am a person of worth, at least on an equal plane with others.
    - I feel that I have a number of good qualities.
    - All in all, I am inclined to feel that I am a failure.
    - I am able to do things as well as most other people.
    - I feel I do not have much to be proud of.
    - I take a positive attitude towards myself.
    - On the whole, I am satisfied with myself.
    - I wish I could have more respect for myself.
    - I certainly feel useless at times.
    - At times I think I am no good at all.

B Measuring self-esteem and testing for implicit discrimination

B.1 Ratio and limitations of our self-esteem measure

Self-esteem can be measured either through direct means, by administering a questionnaire, or through IATs or other indirect routes. Direct elicitation is somewhat problematic because of social desirability bias (Schlenker, 1980): respondents with low self-esteem may be reluctant to answer truthfully. Also, self-deception (Paulhus, 1984) or lack of knowledge of oneself (Wilson et al., 2000)

might lead to inaccurate answers even when the respondent wants to be honest. While IATs are generally considered more reliable (Rudolph et al., 2008), we did not implement one because the benefits in terms of precision did not appear to outweigh the data protection and programming costs accruing on IATs administered to a large sample outside of a lab setting. We choose instead to run a survey, based on evidence that certain questionnaires work well: they deliver results that are both internally consistent and highly correlated with those derived from alternative measurement strategies (Rosenberg, 1965; Schütz and Sellin, 2006). We chose the Rosenberg questionnaire on account of its widespread use in the literature and of its limited length; the survey was voluntary and we did not want to increase respondent burden. This questionnaire can be found in Appendix A, as the final set of items of the survey. It consists of five positive affirmations and five negative affirmations about one's self; the respondent is asked to express their level of agreement with each, using a four-item Likert scale. The individual responses are transformed and aggregated with a standard formula, summing up to a single indicator ranging from 0 (very low self-esteem) to 30 (very high self-esteem).

Unfortunately, test-takers in our sample provided very homogeneous responses on most individual items, resulting in minimal variance in the final indicator (Table 10).

This is not consistent with results observed in the literature; it suggests that a mixture of sample selection mechanisms and social desirability biases might be at work. A considerable share of applicants who fail to be hired in one recruitment round re-apply the following year. Unsuccessful test-takers that are planning to re-apply are probably more likely to participate in the survey. They might also erroneously assume that their answers are going to influence their chances of success in the following round, hence choosing response options according to their expectations on recruiter preferences. If most respondents share a stereotypical idea of what a "good" answer is, the heterogeneity in measurements will be very limited independent of true heterogeneity in the underlying trait.

B.2   A testable model of self-esteem and performance

We intended to test for implicit discrimination based on gender gaps in self-esteem with the following model. Consider a multiple-choice test with M questions and N candidates. Each candidate $i \in [1, 2, ..., N]$ is characterized by ability A_i and self-esteem Δ_i . Each question $j \in [1, 2, ..., M]$ has K response options $k \in [a, b, ..., K]$; only one of them, denoted by k_j^*, is correct. The scoring formula for the test is represented by $s' = [s_c, s_m, s_e]$, where the elements are, respectively, the number of points awarded for a correct answer, a missing answer, and an erroneous answer. Candidate i can either choose one of the options, which will then be identified as k_j^*, or leave the question unanswered. We posit that the cognitive process works as follows, for any candidate i and any question j. The candidate, after reading through all the response options, implicitly assesses how likely each of them is to be correct; in formal terms, she estimates the distribution of $p_{ijk} = P(k_j = k_j^*)$ over all

values of k. Since response options are mutually exclusive, this can be represented as a realization of a k-variate Dirichlet distribution. The estimate, formulated under strict time pressure, results from a combination of subject matter knowledge and self-confidence: knowledge determines the initial intuition, and self-confidence determines how strongly the candidate trusts this intuition. The probability distribution resulting from the initial intuition is

$$\boldsymbol{p}_{ij0} = [p_{ij0a}, p_{ij0b}, ..., p_{ij0K}]^{'} \sim Dir(\Psi_{ijk}),$$

with $\Psi_{ijk} = f(Ai)$, and the probability of k^* maximizing p_{ijk} increasing with A_i.

The effect of self-doubt comes into play through a random confounding factor corresponding to a completely random guess:

$$\boldsymbol{\varepsilon}_i \sim Dir(0.25, 0.25, 0.25, 0.25),$$

resulting in a final assessment vector defined as

$$\boldsymbol{p}_{ij} = [p_{ija}, p_{ijb}, ..., p_{ijK}]^{'} = (1 - \alpha_i)\boldsymbol{p}_{ij0} + \boldsymbol{\varepsilon}_i,$$

where $0 < \alpha_i < 1$ is a weighting factor inversely related to the self-esteem measure δ_i. The preferred answer k_{ij}^* is such that $p_{ijk}^* = \max_k(p_{ijk})$; the more self-confident a candidate is, the lower the impact exerted by the random confounding factor in choosing the preferred answer, and the higher the impact exerted by knowledge. Let the expected score from answering each question be

$$E[s_{ij}] = p_{ijk^*}s_c + (1 - p_{ijk^*})s_e$$

A risk-neutral rational rational candidate will choose to answer if and only if $E[s_{ijk^*}] > s_m$, that is equal to

$$p_{ijk^*} > \frac{s_m - s_e}{s_c - s_e}.$$

Since candidate have heterogeneous levels of risk aversion, the subjective response threshold for candidate i can be defined as

$$\lambda_i = \frac{s_{mi} - s_{ei}}{s_{ci} - s_{ei}}.$$

Hence, a candidate answers the question if $p_{ijk^*} > \lambda_i$. Let $\boldsymbol{r}_i = [ric, rim, rie]^{'}$ describe the final performance of candidate i, where the elements are, respectively, the number of correct answers, missing answers, and erroneous answers. The final score achieved by i can be computed as $S_i = \boldsymbol{r}_i^{'}\boldsymbol{s}$. The model's main predictions should be:

$$\frac{\partial S_i}{\partial (A_i \cdot \Delta_i)} > 0$$
$$\frac{\partial S_i}{\partial A_i} > 0$$
$$\frac{\partial S_i}{\partial \Delta_i} = 0$$

signifying that test-takers who are both skilled and correctly aware of their skills will do better than competitors.

References

Akerlof, G. A., and R. E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being.* Princeton: Princeton University Press.

Banaji, M., and D. Prentice. 1994. The self in social contexts. *Annual Review of Psychology* 45(1): 297–332.

Baumeister, R. F. 1998. *The self*, 680–740. Handbook of social psychology, New York: McGraw-Hill.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova. 2009. Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics* 124(4): 1497–1540.

Bénabou, R., and J. Tirole. 2011. Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics* 126(2): 805–855.

Benenson, J., and A. Heath. 2006. Boys withdraw more in one-on-one interactions, whereas girls withdraw more in groups. *Developmental Psychology* 42(2): 272–282.

Bertrand, M., D. Chugh, and S. Mullainathan. 2005. Implicit Discrimination. *American Economic Review* 95(2): 94–98.

Blair, V. I., E. Ma, and A. Lenton. 2001. Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology* 81(5): 828–841.

Booth, A. L., and P. Nolen. 2012. Gender differences in risk behaviour: does nurture matter? *The Economic Journal* 122(558): 56–78.

Boyd, C. L., L. Epstein, and A. D. Martin. 2010. Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science* 54(2): 389–411.

Branden, N. 1990. What Is Self-Esteem? Technical report, Paper presented at the International Conference on Self-Esteem, Oslo.

Burton, R. 2001. Quantifying the Effects of Chance in Multiple Choice and True/False Tests: question selection and guessing of answers. *Assessment & Evaluation in Higher Education* 26(1): 41–50.

Bush, M. 2001. A Multiple Choice Test that Rewards Partial Knowledge. *Journal of Further and Higher Education* 25(2): 157–163.

Chan, N., and P. E. Kennedy. 2002. Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple-Choice and "Equivalent" Constructed-Response Exam Questions. *Southern Economic Journal* 68(4): 957–971.

Charles, K. K., J. Guryan, and J. Pan. 2009. Sexism and Womens Labor Market Outcomes. Mimeo.

Cooley, C. H. 1902. *Human Nature and Social Order*. New York: Charles Scribnern's Son.

Cross, S. E., and L. Madson. 1997. Models of the self: Self-construals and gender. *Psychological Bulletin* 122(1): 5–37.

Dasgupta, N., and S. Asgari. 2004. Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology* 40(5): 642 – 658.

Eckel, C. C., and P. J. Grossman. 2008. *Men, Women and Risk Aversion: Experimental Evidence*, vol. 1 of *Handbook of Experimental Economics Results*, chap. 113, 1061–1073. Elsevier.

Espinosa, M. P., and J. Gardeazabal. 2007. Optimal Correction for Guessing in Multiple-Choice Tests. DFAEII Working Papers 08, University of the Basque Country - Department of Foundations of Economic Analysis II.

Fortin, N. M. 2005. Gender Role Attitudes and the Labour-market Outcomes of Women across OECD Countries. *Oxford Review of Economic Policy* 21(3): 416–438.

Gabriel, S., and W. Gardner. 1999. Are there "his" and "hers" types of interdependence? The implications of gender differences in collective versus relational interdependence for affect, behavior, and cognition. *Journal of Personality and Social Psychology* 77(2): 642–655.

Greenberg, J., T. Pyszczynski, and S. Solomon. 1986. *The causes and consequences of a need for self-esteem: A terror management theory*, 189–212. Public self and private self, New York: Springer-Verlag.

Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz. 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Soclal Psychology* 74(8): 1464–1480.

Harter, S. 1985. *Competence as a dimension of self-evaluation: Toward a comprehensive model of self-worth*. The Development of Self, Orlando: Academic Press, Inc.

Hinz, R. P., D. D. McCarthy, and J. A. Turner. 1996. Are Women Conservative Investors? Gender Differences in Participant Directed Pension Investments. Pension Research Council Working Papers 17, Wharton School Pension Research Council, University of Pennsylvania.

Hunter, R. 1997. *Indirect Discrimination in the Workplace*. Sydney: Federation Press.

James, W. 1890. *The Principles of Psychology*. New York: Cosimo.

Jennings, S., and M. Bush. 2011. A Comparison of Conventional and Liberal (Free-Choice) Multiple-Choice Tests. *Practical Assessment, Research & Evaluation* 8(6): 1–5.

Jianakoplos, N. A., and A. Bernasek. 1998. Are women more risk averse? *Economic Inquiry* 36(4): 620–630.

Jolls, C., and C. R. Sunstein. 2006. Debiasing through Law. *The Journal of Legal Studies* 35(1): 199–242.

Josephs, R., H. Markus, and R. Tafarodi. 1992. Gender and self-esteem. *Journal of Personality and Social Psychology* 63(3): 391–402.

Kling, K., J. Hyde, C. Showers, and B. Buswell. 1999. Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin* 125(4): 470–500.

Kostanski, M., and E. Gullone. 1998. Adolescent Body Image Dissatisfaction: Relationships with Self-esteem, Anxiety, and Depression Controlling for Body Mass. *Journal of Child Psychology and Psychiatry* 39(2): 255–262.

Lamke, L. K. 1982. The Impact of Sex-Role Orientation on Self-Esteem in Early Adolescence. *Child Development* 53(6): 1530–1535.

Latu, I. M., T. L. Stewart, A. C. Myers, C. G. Lisco, S. B. Estes, and D. K. Donohue. 2011. What We "Say" and What We "Think" About Female Managers: Explicit Versus Implicit Associations of Women With Success. *Psychology of Women Quarterly* 35(2): 252–266.

Levinson, J. D., and D. Young. 2010. Implicit Gender Bias in the Legal Profession: An Empirical Study. *Duke Journal of Gender Law & Policy* 18(1): 1–44.

Lorenzo, M., C. Crouch, and E. Mazur. 2006. Reducing the gender gap in the physics classroom. *American Journal of Physics* 74(2): 118.

Lytton, H., and D. Romney. 1991. Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin* 109(2): 267–296.

Marx, D. M., and J. S. Roman. 2002. Female Role Models: Protecting Women's Math Test Performance. *Personality and Social Psychology Bulletin* 28(9): 1183–1193.

McMullin, J. A., and J. Cairney. 2003. Self-esteem and the intersection of age, class, and gender. *Journal of Aging Studies* 18(1): 75–90.

Mo, C. H., and M. G. Weiksner. 2009. The Sexist Vote: Results from a Lab Election Experiment. Stanford Working Paper.

Ng, A. W. Y., and A. H. S. Chan. 2009. The Testing Methods and Gender Differences in Multiple-Choice Assessment. *American Institute of Physics Conference Proceedings* 1174(1): 236–243.

Niederle, M., and L. Vesterlund. 2007. Do Women Shy Away from Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics* 122(3): 1067–1101.

———. 2011. Gender and Competition. *Annual Review of Economics* 3(1): 601–630.

Nolen-Hoeksema, S. 2001. Gender Differences in Depression. *Current Directions in Psychological Science* 10(5): 173–176.

Nosek, B. A., F. L. Smyth, N. Sriram, N. M. Lindner, T. Devos, A. Ayala, Y. Bar-Anan, R. Bergh, H. Cai, K. Gonsalkorale, S. Kesebir, N. Maliszewski, F. Neto, E. Olli, J. Park, K. Schnabel, K. Shiomura, B. Tulbure, R. W. Wiers, M. Somogyi, N. Akrami, B. Ekehammar, M. Vianello, M. R. Banaji, and A. G. Greenwald. 2009. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106(26): 10593–10597.

OECD. 2010. *OECD Employment Outlook.* Paris.

Paulhus, D. L. 1984. Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46(3): 598–609.

Peresie, J. L. 2005. Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts. *The Yale Law Journal* 114(7): 1759–1790.

Pipher, M. 1994. *Reviving Ophelia: Saving the Selves of Adolescent Girls.* New York: Riverhead Books.

Rosenberg, M. 1965. *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rudolph, A., M. Schröder-Abé, A. Schütz, A. P. Gregg, and C. Sedikides. 2008. Through a glass, less darkly? Reassessing convergent and discriminant validity in measures of implicit self-esteem. *European Journal of Psychological Assessment* 124(3): 273–281.

Sapienza, P., L. Zingales, and D. Maestripieri. 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences* 106(26): 1526815273.

Schlenker, B. R. 1980. *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations.* Monterey: Brooks-Cole.

Schubert, R., M. Brown, M. Gysler, and H. W. Brachinger. 1999. Financial Decision-Making: Are Women Really More Risk-Averse? *American Economic Review* 89(2): 381–385.

Schütz, A., and I. Sellin. 2006. *Die multidimensionale Selbstwertskala (MSWS)*. Göttingen: Hogrefe.

Simkin, M. G., and W. L. Kuechler. 2005. Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education* 3(1): 73–98.

Sriram, N., and A. G. Greenwald. 2009. The Brief Implicit Association Test. *Experimental Psychology* 56(4): 283–294.

Tenenbaum, H., and C. Leaper. 2003. Parent-child conversations about science: The socialization of gender inequities. *Developmental Psychology* 39(1): 34–47.

Tiggemann, M. 2005. Body dissatisfaction and adolescent self-esteem: Prospective findings. *Body Image* 2(2): 129 – 135.

U.S. Census Bureau. 2011. *Statistical Abstract of the United States, 2010*. 130th ed. Washington, DC: U.S. Census Bureau.

Wilson, T. D., S. Lindsey, and T. Y. Schoole. 2000. A Model of Dual Attitudes. *Psychological Review* 107(1): 101–126.

Statistical tables

Table 1: Summary statistics, recruitment rounds 1998-2011

Selection	Test				Written Assessment				Oral Assessment				Hired			
	Male	Female	Total	Perc. F	Male	Female	Total	Perc. F	Male	Female	Total	Perc. F	Male	Female	Total	Perc. F
								Original Scoring Rule (1998-2009)								
Law 1	256	524	780	67.2%	57	43	100	43.0%	10	15	25	60.0%	7	13	20	65.0%
Law 2	391	797	1188	67.1%	113	107	220	48.6%	23	19	42	45.2%	16	16	32	50.0%
Law 3	320	638	958	66.6%	86	64	150	42.7%	23	6	29	19.4%	22	5	27	18.5%
Law 4	714	1389	2103	66.0%	108	93	201	46.3%	27	24	51	47.1%	25	24	49	49.0%
Law 5	543	1121	1664	67.4%	107	143	250	57.2%	24	16	40	40.0%	18	10	28	35.7%
Law 6	331	727	1058	68.7%	57	65	122	53.3%	11	9	20	45.0%	11	9	20	45.0%
Total Law	2555	5196	7751	67.0%	528	511	1039	49.2%	120	89	209	42.6%	99	77	176	43.8%
Eco 1	445	464	909	51.0%	139	61	200	30.5%	28	10	38	26.3%	24	10	34	29.4%
Eco 2	418	430	848	50.7%	204	96	300	32.0%	35	13	48	27.1%	26	9	35	25.7%
Eco 3	322	363	685	53.0%	207	143	350	40.9%	29	15	44	34.1%	25	14	39	35.9%
Eco 4	592	807	1399	57.7%	161	89	250	35.6%	44	16	60	26.7%	40	14	54	25.9%
Eco 5	410	495	905	54.7%	143	57	200	28.5%	31	11	42	26.2%	29	9	38	23.7%
Eco 6	273	341	614	55.5%	143	97	240	40.4%	22	21	43	41.0%	21	14	35	40.0%
Eco 7	139	148	287	51.6%	35	13	48	27.1%	10	10	31	32.3%	9	3	12	25.0%
Total Eco	2599	3048	5646	54.0%	1032	556	1588	35.0%	200	91	302	30.1%	174	73	247	29.6%
Total	5154	8244	13397	61.5%	1560	1067	2627	40.6%	320	180	511	35.2%	273	150	423	35.5%
								New Scoring Rule (2010-2011)								
Law 7	595	1208	1803	67.0%	115	135	250	54.0%	29	34	63	54.0%	25	31	56	55.4%
Law 8	255	363	618	58.7%	79	41	120	34.2%	17	7	24	29.2%	10	4	14	28.6%
Total	850	1571	2421	64.9%	194	176	370	47.6%	46	41	87	47.1%	35	35	70	50.0%

Table 2: Linear Regression Results. Dependent variable: standardized score in the multiple-choice test. Recruitment rounds 1998-2009, original scoring rule

	I	II
Intercept	6.88***	6.58***
Gender: female	-0.49***	-0.49***
Age at graduation	-0.80***	-0.77***
Age at graduation, squared	0.01	0.01***
Age at time of test	0.36***	0.34***
Age at time of test, squared	-0.01	0.00
Residence: Central Italy	-0.10***	-0.13***
Residence: Southern Italy	-0.13***	-0.22***
University: Central Italy	-0.24***	-0.42
University: Southern Italy	-0.44***	1.01
Final grade: 106	0.07**	0.06*
Final grade: 107	0.10***	0.09***
Final grade: 108	0.16***	0.15***
Final grade: 109	0.19**	0.14**
Final grade: 110	0.22***	0.22***
Final grade: 110 honors	0.43***	0.43***
N. obs	16427	16426
R-squared:	0.19	0.23
Adjusted R-squared:	0.19	0.22
University Fixed -Effects	No	Yes
Selection Dummies	Yes	Yes

Table 3: Test-takers and survey respondents, 2010-2011

	Law 7					
	All Participants			Survey Participants		
	Failed	Passed	All	Failed	Passed	All
Male	26.6	6.4	33.0	23.2	9.2	32.4
Female	59.5	7.5	67.0	57.2	10.3	67.6
All	86.1	13.9	100.0	80.5	19.5	100.0
	Law 8					
	All Participants			Survey Participants		
	Failed	Passed	All	Failed	Passed	All
Male	28.5	12.8	41.3	23.5	14.8	38.3
Female	52.1	6.6	58.7	51.9	9.8	61.7
All	80.6	19.4	100.0	75.4	24.6	100.0

Table 4: Summary statistics, response behavior in the multiple-choice test, 1998-2011 (percentages)

	Correct Answers						Incorrect Answers						Missing Answers					
Selection	Q_1	Q_{25}	Q_{50}	Mean	Q_{75}	Q_{100}	Q_1	Q_{25}	Q_{50}	Mean	Q_{75}	Q_{100}	Q_1	Q_{25}	Q_{50}	Mean	Q_{75}	Q_{100}
							Original Scoring Rule											
Law 1	0.0	45.8	53.3	52.8	60.4	80.0	0.0	15.0	21.7	22.2	29.2	67.5	0.0	14.6	23.3	24.9	34.2	100.0
Law 2	13.6	39.1	47.3	46.9	54.6	86.4	0.0	18.2	26.4	27.9	35.5	69.1	0.0	13.0	25.0	25.2	35.5	74.6
Law 3	20.9	43.6	51.8	52.5	60.9	86.4	2.7	19.1	26.4	27.4	34.6	66.4	0.0	9.1	19.1	20.1	30.0	72.7
Law 4	18.2	43.6	52.7	52.7	60.9	88.2	2.7	20.0	27.3	28.3	35.5	74.6	0.0	6.4	17.3	19.1	29.1	72.7
Law 5	2.7	49.1	57.3	57.1	66.4	94.6	0.0	15.5	21.8	22.8	29.1	73.6	0.0	10.0	18.2	20.1	28.2	97.3
Law 6	15.8	50.0	59.2	58.7	68.3	95.0	0.8	17.5	23.8	24.4	30.8	62.5	0.0	5.8	14.2	16.9	25.0	79.2
Eco 1	18.3	47.5	55.8	55.1	63.3	85.0	0.0	14.2	20.0	21.4	27.5	62.5	0.0	12.5	22.5	23.5	31.7	70.8
Eco 2	1.8	39.1	47.3	47.5	56.4	79.1	3.6	17.3	23.6	25.6	32.7	69.1	0.0	14.6	26.4	26.9	38.2	89.1
Eco 3	0.0	40.0	48.2	48.0	55.5	80.0	0.0	19.1	25.5	27.5	35.5	64.6	0.0	12.7	22.7	24.5	34.6	100.0
Eco 4	12.7	40.9	49.1	48.8	57.3	89.1	2.7	17.3	24.6	26.1	33.6	77.3	0.0	12.7	24.1	25.1	37.3	80.0
Eco 5	14.6	41.8	50.9	50.5	58.2	83.6	1.8	15.5	22.7	25.2	32.7	71.8	0.0	11.8	23.6	24.3	34.6	73.6
Eco 6	21.7	48.3	55.8	55.5	62.5	86.7	2.5	15.8	21.7	22.6	29.2	58.3	0.0	10.8	20.8	21.9	32.5	70.0
Eco 7	23.3	43.3	51.7	51.6	58.3	88.3	3.3	14.2	20.8	23.5	30.0	64.2	0.0	11.9	25.8	24.9	35.0	65.0
Male	12.7	47.5	55.8	55.8	64.6	95.0	0.0	15.8	22.5	24.0	30.9	74.6	0.0	8.3	18.3	20.2	30.0	79.2
Female	0.0	41.8	50.0	50.2	58.3	90.8	0.0	17.5	25.0	26.2	33.6	77.3	0.0	11.7	22.5	23.6	34.2	100.0
Total 98-09	0.0	43.6	52.7	52.4	60.9	95.0	0.0	16.7	24.2	25.4	32.5	77.3	0.0	10.0	20.9	22.3	32.7	100.0
							New Scoring Rule											
Law 7	10.0	34.2	42.5	42.7	50.8	78.3	1.7	18.3	26.7	27.3	34.2	73.3	0.0	18.3	29.2	30.0	40.8	82.5
Law 8	0.0	30.0	35.0	36.1	42.5	68.3	0.0	19.2	27.5	28.8	36.7	69.2	0.0	23.3	35.8	35.1	47.5	100.0
Male	10.0	35.0	42.5	43.4	50.8	77.5	1.7	17.5	25.8	26.6	34.2	68.3	0.0	18.3	29.6	30.0	40.8	82.5
Female	0.0	31.7	39.2	39.7	47.5	78.3	0.0	19.2	26.7	28.3	35.8	73.3	0.0	20.0	31.7	32.0	43.3	100.0
Total 10-11	0.0	32.5	40.0	41.0	49.2	78.3	0.0	18.3	26.7	27.7	35.0	73.3	0.0	19.2	30.8	31.3	42.5	100.0

Table 5: Distribution of the standardized score, 1998-2011

Selection	Q_1	Q_{25}	Q_{50}	Mean	Q_{75}	Q_{100}
Original Scoring Rule						
Law 1	-4.02	-0.67	0.04	0.00	0.72	2.61
Law 2	-2.47	-0.69	-0.04	0.00	0.71	3.60
Law 3	-2.78	-0.74	-0.04	0.00	0.72	2.87
Law 4	-3.00	-0.71	-0.03	0.00	0.65	3.26
Law 5	-3.19	-0.69	-0.03	0.00	0.70	3.10
Law 6	-2.75	-0.74	0.01	0.00	0.76	3.04
Eco 1	-2.98	-0.63	0.02	0.00	0.76	2.80
Eco 2	-3.01	-0.74	-0.01	0.00	0.70	2.75
Eco 3	-3.12	-0.70	0.00	0.00	0.60	2.99
Eco 4	-3.23	-0.68	-0.03	0.00	0.67	3.81
Eco 5	-2.75	-0.73	-0.04	0.00	0.63	3.00
Eco 6	-2.93	-0.66	-0.01	0.00	0.65	3.15
Eco 7	-1.95	-0.67	-0.11	0.00	0.56	3.20
Gender: male	-3.00	-0.36	0.32	0.32	0.99	3.81
Gender: female	-4.02	-0.88	-0.23	-0.20	0.46	2.99
Total 98-09	-4.02	-0.70	-0.02	0.00	0.69	3.81
New Scoring Rule						
Law 7	-3.09	-0.70	-0.03	0.00	0.66	2.89
Law 8	-2.51	-0.69	-0.07	0.00	0.60	3.80
Male	-2.72	-0.45	0.20	0.24	0.89	3.80
Female	-3.09	-0.80	-0.17	-0.13	0.48	3.01
Total 10-11	-3.09	-0.70	-0.04	0.00	0.65	3.80

Table 6: Surveyed traits, 2010-2011, by gender (percentages)

	Male	Female
Where do you live?		
Northern Italy	14.4	12.4
Central Italy	39.7	43.1
Southern Italy	45.9	44.5
Working Status		
Employed, full-time	42.9	32.5
Employed, part-time	8.8	9.0
Occasionally employed	7.9	9.0
Unemployed	15.9	18.7
Student	12.1	14.0
Other	12.4	16.8
Test Preparation Strategy		
Textbooks and mock tests	38.8	44.1
Textbooks, no mock tests, previous experience	12.1	5.2
Textbooks, no mock tests, no previous experience	12.7	7.8
No time to prepare	36.5	42.9
Subjective: Global Test Difficulty		
Not difficult at all	0.0	0.0
Not very difficult	2.4	1.0
Quite difficult	52.1	51.6
Very difficult	45.6	47.4
Subjective: Most Difficult Section of Test		
Logic	46.5	55.6
Law	29.7	25.2
Foreign Language	8.5	8.4
None in Particular	15.3	10.8
Most Important Motivation for Applying		
Job security	12.1	15.2
Salary	8.5	4.1
Employer reputation	35.9	24.6
Expected work/life balance	0.6	0.6
Professional contents	18.2	18.4
Relevance to own studies	7.7	13.6
Interest in central banking	17.1	23.5
Second Most Important Motivation for Applying		
Job Security	12.9	14.5
Salary	14.7	14.8
Employer reputation	23.8	17.9
Expected work/life balance	1.5	3.1
Professional contents	19.4	18.3
Relevance to own studies	14.7	17.7
Interest in central banking	12.9	13.9
Yes	67.1	62.2
No	32.9	37.8
At Least One Parent with University Degree		
Yes	58.8	48.4
No	41.2	51.6
Answering Strategy		
Only answered when certain	9.4	10.8
Answered when certain or almost certain	52.7	43.4
Answered also based on partial information	33.2	39.7
Answered all questions	4.7	6.2
Subjective: Time Allotted		
Not enough	43.2	48.1
Just enough	49.7	45.1
More than enough	7.1	6.8
At Least One Parent in Public Sector		
Yes	67.1	62.2
No	32.9	37.8

	Male	Female
Risk Aversion: Wheel of Fortune 1		
Averse	70.0	74.5
Lover	20.0	15.0
Neutral	10.0	10.5
Risk Aversion: Wheel of Fortune 2		
Averse	26.8	29.5
Lover	70.0	68.0
Neutral	3.2	2.5
Interest in Banking Supervision		
No	23.2	27.3
Yes	76.8	72.7
Interest in Economic Research		
No	77.1	81.3
Yes	22.9	18.7
Interest in Other Activities	78.5	73.5
No		
Yes	21.5	26.6
Graduated Outside Home Region		
Yes	39.1	40.7
No	60.9	59.3
Postgraduate Education: Msc/MA		
Yes	11.8	9.1
No	88.2	90.9
Postgraduate Education: Ph.D.	13.2	8.1
Yes		
No	86.8	91.9
Professional Training	53.8	59.7
1		
2	46.2	40.3
Admission to Practice Law		
Yes	35.3	32.5
No	64.7	67.6
Applied to Other Government Jobs		
Yes	56.2	55.2
No	43.8	44.8
Hired for Other Government Jobs	20.9	14.9
Yes		
No	79.1	85.1
Employment preference		
Public sector	25.3	16.8
Private sector	74.7	83.2
Final grade		
N/A	0.0	0.2
105	12.7	10.5
106	7.4	5.5
107	5.0	4.9
108	6.5	6.6
109	2.1	0.9
110	11.8	16.7
110 honors	54.7	54.9
Children under the age of 14		
Yes	6.5	4.4
No	93.5	95.6
Age at Graduation (years)	26.0	26.0
Experience after Graduation (years)	4.2	3.6

Table 7: Linear regression results. Dependent variable: standardized score. Recruitment rounds: 2010 and 2011 for law graduates

	I	II	III	IV	V	VI
Intercept	1.88	2.57	2.17	0.47	1.18	1.30
Round: 2011	0.12*	0.12*	0.13**	0.15**	0.16**	0.14*
Gender: female	-0.36***	-0.36***	-0.36***	-0.31***	-0.26***	-0.26***
Age at graduation	-0.73***	-0.75***	-0.63***	-0.47***	-0.45***	-0.45***
Age at graduation, squared	0.01***	0.01***	0.01***	0.01**	0.01**	0.01**
Age at time of test	0.62***	0.59***	0.49***	0.42***	0.40***	0.39***
Age at time of test, squared	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
Residence: Central Italy	-0.26**	-0.27**	-0.27**	-0.22*	-0.26**	-0.27**
Residence: Southern Italy	-0.39***	-0.40***	-0.36***	-0.34***	-0.36***	-0.39***
University: Central Italy	-0.13	-0.12	-0.14	-0.18*	-0.16	-0.14
University: Southern Italy	-0.25**	-0.24**	-0.31***	-0.34***	-0.32***	-0.29***
Final grade: 106	0.03	0.03	0.02	0.08	0.08	0.04
Final grade: 107	0.08	0.09	0.11	0.10	0.09	0.14
Final grade: 108	0.11	0.11	0.12	0.14	0.13	0.16
Final grade: 109	0.57**	0.55**	0.58**	0.62***	0.68***	0.67***
Final grade: 110	0.12	0.12	0.13	0.13	0.16	0.17*
Final grade: 110 honors	0.42***	0.42***	0.42***	0.42***	0.41***	0.41***
Children under the age of 14		-0.32**	-0.29**	-0.26**	-0.23*	-0.21*
Graduated outside home region			-0.10	-0.08	-0.10*	-0.09
Msc/MA			0.17*	0.14	0.14*	0.13
Ph.D.			0.15*	0.14	0.11	0.11
Professional training			0.14**	0.08	0.08	0.08
Admission to practice law			0.14**	0.13*	0.12*	0.11
Applied for other government jobs				0.18***	0.13**	0.13**
Was hired for other government jobs				0.28***	0.22***	0.23***
Employed, part-time				-0.12	-0.17*	-0.17*
Employed, occasionally				-0.12	-0.12	-0.13
Unemployed				0.13	0.08	0.08
Student				0.16*	0.11	0.10
Other professional condition				-0.03	-0.08	-0.10
Preference for public sector				0.10	0.10	0.08
Civil servants in family				0.01	0.01	0.02
University degree in family				0.21***	0.20***	0.20***
Preparation: textbooks, no mocks, experience					0.23**	0.21**
Preparation: textbooks, no mocks, no experience					0.13	0.12
Preparation: none					-0.12**	-0.10*
Test: quite difficult					-0.46**	-0.40*
Test: very difficult					-0.64***	-0.57**
Allotted time: just enough					0.10*	0.11*
Allotted time: more than enough					-0.02	0.01
Strategy: answered when certain					0.17*	0.12
Strategy: answered when almost certain					0.00	-0.05
Strategy: answered anyway					-0.32**	-0.33**
Most difficult section: law					-0.09	-0.09
Most difficult section: foreign language					-0.22**	-0.21**
Most difficult section: none					0.06	0.04
Sector preference: private					-0.23**	-0.25**
Sector preference: none					-0.03	-0.04
Self esteem: score on positive items						0.00
Self esteem: score on negative items						-0.01
Risk aversion: low						0.18***
Risk aversion: high						-0.25
Multiple R-squared	0.211	0.216	0.230	0.276	0.322	0.327
Adjusted R-squared	0.200	0.204	0.215	0.255	0.294	0.295

Signif. codes: 0.01="***"; 0.05="**"; 0.10="*".

Table 8: Logistic regression results. Dependent variable: successful outcome in the multiple-choice test. Recruitment rounds: 2010 and 2011 for law graduates

	I	II	III	IV	V	VI
Intercept	4.06	6.07	7.12	4.00	4.56	4.57
Round: 2011	0.30	0.30*	0.35*	0.41**	0.47**	0.31
Gender: female	-0.94***	-0.95***	-0.97***	-0.88***	-0.76***	-0.70***
Age at graduation	-1.71***	-1.81***	-1.49***	-1.15**	-1.05*	-1.05
Age at graduation, squared	0.03***	0.03***	0.02**	0.02	0.02	0.01
Age at time of test	1.38***	1.32***	0.96**	0.80**	0.73*	0.72*
Age at time of test, squared	-0.02***	-0.02***	-0.01**	-0.01*	-0.01*	-0.01*
Residence: Central Italy	-0.69**	-0.73**	-0.76**	-0.69**	-0.79**	-0.78**
Residence: Southern Italy	-0.81***	-0.85***	-0.66**	-0.65**	-0.78**	-0.85**
University: Central Italy	-0.21	-0.19	-0.24	-0.35	-0.26	-0.30
University: Southern Italy	-0.46*	-0.44	-0.75**	-0.86***	-0.79**	-0.76**
Final grade: 106	-0.46	-0.44	-0.51	-0.49	-0.46	-0.48
Final grade: 107	0.62	0.65	0.66	0.58	0.66	0.80*
Final grade: 108	0.10	0.11	0.10	0.14	0.28	0.34
Final grade: 109	1.80***	1.79***	1.78***	1.92***	2.24***	2.28***
Final grade: 110	0.38	0.41	0.40	0.39	0.55	0.56
Final grade: 110 honors	0.97***	0.99***	0.97***	1.00***	1.13***	1.17***
Children under the age of 14		-0.79*	-0.69	-0.58	-0.49	-0.46
Graduated outside home region			-0.43**	-0.45**	-0.52**	-0.40*
Msc/MA			0.19	0.12	0.20	0.20
Ph.D.			0.05	0.03	0.01	0.04
Professional training			0.16	0.02	0.09	0.10
Admission to practice law			0.56***	0.58***	0.57**	0.59**
Applied for other government jobs				0.13	0.01	-0.01
Was hired for other government jobs				0.47*	0.35	0.37
Employed, part-time				-0.12	-0.23	-0.31
Employed, occasionally				-0.35	-0.36	-0.45
Unemployed				0.47*	0.38	0.34
Student				0.48*	0.29	0.29
Other professional condition				-0.54*	-0.67**	-0.68**
Preference for public sector				0.26	0.37	0.33
Civil servants in family				0.16	0.18	0.21
University degree in family				0.50***	0.49**	0.46**
Preparation: textbooks, no mocks, experience					0.77***	0.72**
Preparation: textbooks, no mocks, no experience					0.26	0.27
Preparation: none					-0.67***	-0.67***
Test: quite difficult					-1.03*	-0.97
Test: very difficult					-1.44**	-1.37**
Allotted time: just enough					0.40**	0.38**
Allotted time: more than enough					0.60*	0.60*
Strategy: answered when certain					0.39	0.27
Strategy: answered when almost certain					0.11	-0.01
Strategy: answered anyway					-0.52	-0.53
Most difficult section: law					-0.35	-0.35
Most difficult section: foreign language					-0.53	-0.45
Most difficult section: none					0.20	0.17
Sector preference: private					-0.35	-0.41
Sector preference: none					0.18	0.15
Self esteem: score on positive items						0.03
Self esteem: score on negative items						-0.04
Risk aversion: low						0.62***
Risk aversion: high						-0.27
AIC	1044.0	1042.3	1038.3	1020.7	990.1	978.6

Signif. codes: 0.01="***"; 0.05="**"; 0.10="*".

Table 9: Blinder-Oaxaca decomposition for the complete-specification linear model

Average group differences of the standardized scores	0.42	
Of which:		
Gender gap due to differences in distributions of the regressors	0.17	39.54%
Gender gap captured by differences in the regressors' coefficients	0.14	33.98%
Gender gap captured by differences only in the intercept	0.11	26.48%

Table 10: Distribution of Rosenberg's self-esteem score, by gender. Recruitment rounds: 2010-2011 for law graduates

	Direct SE	Inverse SE	Total SE
		Males	
Q_1	4.00	3.00	7.00
Q_{25}	9.25	13.00	23.00
Q_{50}	11.00	16.00	26.00
Mean	10.37	15.08	25.45
Q_{75}	12.00	18.00	28.00
Q_{100}	12.00	18.00	30.00
		Females	
Q_1	3.00	1.00	7.00
Q_{25}	9.00	13.00	22.00
Q_{50}	10.00	15.00	26.00
Mean	10.06	14.61	24.67
Q_{75}	11.00	18.00	28.00
Q_{100}	12.00	18.00	30.00