ISSN 2042-2695

CENTRE for ECONOMIC PERFORMANCE

CEP Discussion Paper No 1314

November 2014

Gender Differences in Response to Big Stakes

Ghazala Azmat, Caterina Calsamiglia and Nagore Iriberri

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract

In the psychology literature, “choking under pressure” refers to a behavioural response to an increase in the stakes. In a natural experiment, we study the gender difference in performance resulting from changes in stakes. We use detailed information on the performance of high-school students and exploit the variation in the stakes of tests, which range from 5% to 27% of the final grade. We find that female students outperform male students relatively more when the stakes are low. The gender gap disappears in tests taken at the end of high school, which count for 50% of the university entry grade.

Keywords: Stakes, gender gaps, performance
JEL codes: D03; J16; I21; C30

This paper was produced as part of the Centre’s Labour Markets Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Ghazala Azmat acknowledges financial support from ECO2011-30323-C03-02. The authors acknowledge support from the Institut Català de la Dona (U-8/10). Caterina Calsamiglia acknowledges financial support from the Ministerio de Economía y Competición (SEJ2005-01481, SEJ2005-01690, SEJ2006-002789-E and FEDER) and from the “Grupo Consolidado de tipo C” (ECO2008-04756), the Generalita de Catalunya (SGR2005-00626) and the Severo Ochoa program. Nagore Iriberri acknowledges financial support from Ministerio de Economía y Competición (ECO2012-31626), Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT869-13) y Ministerio de Ciencia e Innovación (ECO2011-25295).

Ghazala Azmat, Queen Mary, University of London, and associate of Centre for Economic Performance, London School of Economics. Caterina Calsamiglia, Universitat Automoma de Barcelona and Barcelona GSE. Nagore Iriberri, University of the Basque Country UPV/EHU, IKERBASQUE, Basque Foundation for Research.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1. Introduction

Pressure is a defining feature of many social and economic environments. Final exams in school, the last round of a job interview, giving a speech and answering questions at a press conference are some examples in which stakes are high. In many instances, principals (evaluators), whether in a competitive or non-competitive setting, use a one-shot process to gather information or evaluate the agent. This process is likely to induce pressure since agents understand that they will not have the opportunity to repeat the process—or that doing so will be costly. The psychology literature has identified a number of factors that produce "choking under pressure" (Baumeister, 1984). The most relevant sources of pressure seem to be the presence of an audience, competition with others, personal traits, and one's own ego-relevant threat (see Ariely et al., 2009). Research in psychology has shown that an emphasis on the importance of the process can harm the individual's capacity to exhibit her "true" capability (Beilock, 2011).[1] Moreover, the strain of pressure on performance is heterogeneous across individuals.

In this paper, we provide empirical evidence showing that males and females react differently to increases in pressure, as defined by the size of the stakes at hand. Using detailed information over a period of 12 years on several cohorts of high school students who take numerous tests with varying stakes, we show that there is a change in the gender gap in performance when the tests' stakes increase. In particular, we find that females do relatively better on tests with low stakes but that this difference is reduced, and can even disappear, when the stakes are high. To the best of our knowledge, our paper is the first to provide empirical evidence of a gender gap in the impact of choking under pressure.

We follow several cohorts of students through their six years of high school (ages 12 to 18). For all subjects in each academic year (typically ten or 11 subjects), students undertake three types of tests with varying stakes (low, medium, and high). During the year, students undertake several low-stakes tests, two medium-stakes tests at the end of

[1] In her book "Choke: What the Secrets of the Brain Reveal about Getting it Right When You Have To," Beilock (2011) summarizes the existing literature in social psychology and emphasizes the similarity between students, athletes and business people who choke when the stakes are high. The mechanisms that could produce such a reaction include increased arousal, narrowing of attention, and the preoccupation with the reward itself. It has also been identified that consciously thinking about a task that is usually done automatically can be detrimental to performance. Increased pressure can induce a shift from this "automated" to a "controlled" procedure. See Yerkes and Dodson, 1908; Langer and Imber, 1979; Camerer, Lowenstein and Prelen, 2005 for details on this literature.

each semester, and then a high-stakes test at the end of the academic year. The test stakes vary from five percent to 27 percent of the final grade, which, for each subject, is a weighted average of all the tests taken throughout the academic year. At the end of high school, students undertake national-level (standardized) tests, similar to the SAT in the United States, which are very high-stakes. The grade in the national level test counts for 50 percent of the university entry grade.

This natural experiment provides an ideal testing ground for whether pressure generates gender gaps in performance. There is a large literature that documents gender differences in labor-market outcomes (see Altonji and Blank, 1999; Bertrand, 2009). Understanding if differential reactions to pressure exist can potentially explain part of this gap, but using labor-market information is problematic. Once in the labor market, men and women will have made choices shaped by their professional environment and personal circumstances and, potentially, by their preference for and reaction to pressure. In our setting, we focus on a period in individuals' educational career when, for the last time, they are exposed to a homogeneous and compulsory procedure that will affect their future success in higher education and in the labor market.

The analysis shows that female students outperform male students by 0.18 standard deviations of the mean in low-stakes tests but by only 0.11 standard deviations in high-stakes tests. Moreover, in the national-level exams, the gap is reversed, such that male students outperform female students by 0.02 standard deviations, although this difference is not statistically significant. Interestingly, consistent with the gender differential reaction to the underlying stakes, in the year prior to college, just before the national exams occur, we find that the gender differential response to pressure occurs only for those subjects who are evaluated in the national exams. Our results persist over time, within and between academic years, and throughout the performance distribution.

Gender differences in academic attainment and achievement have been widely documented (see, for example, Goldin, Katz and Kuziemko, 2006). Differences in the nature of tests—in particular, their objectiveness, their competitive nature and the skills they measure—have been exploited to identify potential channels that explain these gaps. Some noteworthy examples are Cornwell, Mustard, and Van Parys (2013), Lavy (2008), and Örs, Palomino and Peyrache (2013).[2] Our setting uses a quasi-experimental

[2] Cornwell, Mustard, and Van Parys (2011) show that the apparent advantage that female students had in grades provided by a teacher versus those provided by an external evaluation disappears once non-cognitive traits are controlled for. Similarly, Lavy (2008) compares blind and non-blind scores on college matriculation exams of male and female students and finds evidence of gender stereotyping and discrimination against male students by teachers. Örs, Palomino and Peyrache (2013) show that when

design in the evaluation system to study gender differences in performance, where the variation exploited is the test stakes, while other factors, such as the evaluators, the competitiveness of the environment, the material evaluated or format of the exam, are held constant.

A recent literature shows that women underperform in competitive environments and, when given the opportunity, shy away from competition (Gneezy, Niederle and Rustichini, 2003; Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007; Antonovics, Arcidiacono, and Walsh, 2009; Shurchkov, 2012; Buser, Niederle, Oosterbek; 2013; Iriberri and Rey-Biel, 2014). The emphasis in this literature is on the gender of the opponent(s) and the task at hand. In our setting, since pressure is not defined as having a competitive nature, the rewards are independent of the performance of others. Here, the gaps in performance result from the pressure that arises due to variation in the size of the tests' stakes. In an experimental literature that studies how stakes affect performance through effort choice, stakes are defined as monetary incentives. Camerer and Hogarth (1999) review the experimental results and conclude that the size of the monetary payoff has little effect on effort choice.[3] Ariely et al. (2009) provide experimental evidence on choking under pressure becoming relevant as the size of the stakes is increased. They suggest that there is an optimal amount of motivation/pressure that leads to a maximal performance, and deviation from this will reduce performance.

The rest of the paper is organized as follows. In Section 2, we describe the data and the evaluating procedure that generated it. In Section 3, we present the main results and perform a number of robustness checks. Section 4 concludes by discussing the implications of our findings.

---

female and male students compete against each other to enter college, male students perform better than predicted by their previous high school grades, and woman perform worse.

[3] Lavy and Angrist (2009), in an educational setting, show that providing monetary incentives to students in low-achieving schools improves the matriculation rates of girls, but has no effect on boys.

2. Study Design

2.1. Evaluation System

The data come from a high school in Barcelona, Spain that offers four years of Compulsory Secondary Education (*ESO*), for ages 12 to 15, and *Bachillerato,* which comprises the two years prior to university, for ages 16 to 18. There are, therefore, six levels, which we refer to as Levels 1 to 6. At the end of Level 6, students who plan to pursue a university degree take externally designed and graded national exams (*Selectividad*).

During each academic year, students take several exams in each of their subjects, and the final score for each subject is determined as follows. In each trimester, students take bi-monthly tests and an end-of-trimester test, except in trimester three, when they have an end-of-year test. The final grade is determined by a weighted average of the first, second, and third trimester tests and the end-of-year test. The weights on the bi-monthly tests in each trimester are around 2.5 percent. The end-of-trimester test in trimester one and two is worth 11 percent and the end-of-year test 27 percent.[4] We define the bi-monthly, end-of-term, and end-of-year tests as “low,” “medium,” and “high” stakes, respectively. The evaluation system is summarized in Figure 1.

Students in Level 6 take national-level exams, which, together with their high-school test scores in Levels 5 and 6, determine their college entry grade.[5] The weight assigned to the Level 5 and Level 6 end-of-year grade is 25 percent each, and 50 percent is assigned to the national-level exam grade. We, therefore, define the national exams as being “super-high” stakes tests. The subject material evaluated in the national exams is the same as that covered in Level 6. However, not all subjects are tested in the national exam, which allows for some interesting variation that we will exploit later.

2.2. Data Description

We have panel data on students’ test scores for all subjects, through levels 1 to 6, between the academic years 2000 to 2012, giving a total sample of 1,404 students. For each subject per academic year, we observe eight measures of performance in low-

---

[4] The weights described are the reduced-form weights resulting from the following compounded formula. Every trimester, a grade is constructed as follows. In the first two trimesters, the trimester-grade is constructed giving 60 percent of the weight to the (average) bi-monthly exams and 40 percent to the end-of-trimester exam. In the third trimester, the grade is determined only by the bi-monthly exams. The final grade is determined by giving 27 percent of the weight to trimester one and two, 18 percent to trimester three and 27 percent to the end-of-the-year exam.

[5] College admission in Spain is administered through a centralized system. Applicants submit a single application with a list of up to eight major-university options. Students are then ordered according to their grades and are assigned to their preferred option following that order.

stakes tests, two in medium-stakes tests, and one in high-stakes tests.[6] All test scores are standardized to a distribution with zero mean and a unit standard deviation. The standardization is done by academic year, level, subject and test type. The sample size varies by subjects, including most students in compulsory subjects but far fewer in elective subjects. In addition, we have information on the scores for the national exams (super-high stakes). We also have information on subject teachers for some, but not all, years. Table 1 presents the descriptive statistics, separately for each level.

From Table 1, we can see that there are an equal number of female and male students. Each year, students take around eight subjects.[7] In Levels 1 to 4, all subjects are common and compulsory. High school in Spain is compulsory until age 16 (end of Level 4). From the table, we see that around 20 percent of students leave at the end of Level 4. While some students may choose not to pursue Levels 5 and 6 and, thus, the national-level exams, others might choose to do so in another school. Some, but fewer, students leave at the end of other levels. Students performing very well in the low-stakes and medium-stakes tests, such that they are getting the highest possible grade throughout the course, may be given an exemption from the high-stakes test. Table 1 shows that around five percent of students are exempt from the test. More importantly, across genders—for both those who leave and those who are exempt—there is no difference at any level.[8]

3. Analysis and Results

3.1. Baseline Regressions

We start by estimating gender differences in school performance on different types of tests:

$$(1)\ Y_{ilsy} = \alpha + \beta Female_i + \varepsilon_{ilsy},$$

where the outcome variable, Y_{itlsy}, is the standardized score for student *i*, in level *l*, subject *s*, and academic year *y*. *Female* takes value one if the student is female and zero otherwise. Table 2 shows the estimation results.

[6] In the data, the test scores for the low-stakes tests are typically provided for each month and are an average of the bi-monthly tests.

[7] Table A.1 in the Appendix summarizes all the subjects that students take per level.

[8] We will discuss this further in Section 3.

Panel A, column 1 presents the results for the gender difference in the final end-of-year test score. Columns 2 to 5 present the results for tests with different stakes: low-stakes, medium-stakes, high-stakes, and super-high-stakes, respectively. In Panel B, we restrict the analysis to final-year students (Level 6) who take the super-high-stakes test (*Selectividad*).

Overall, we see that in school, female students outperform men by 0.16 standard deviations of the mean. Looking at different types of tests, we find that the gender gap in performance falls as the stakes of the test increase. Outperformance is highest in low-stakes tests, 0.18 standard deviations, and lowest in high-stakes tests, 0.11 standard deviations. Moreover, in super-high-stakes tests (Panel B), the sign of the coefficient reverses (-0.03), such that male students outperform female students. This difference, however, is not significant. Panel B shows that the same patterns persist when we restrict the analysis to level 6 only. Female students perform relatively better in low-stakes tests—0.18 standard deviations—than in high-stakes tests—0.08 standard deviations.

In the following analysis, we study students' relative performance on different types of tests. We estimate the following regression:

$$(2)\ Y_{ilsyt} = \alpha + \beta Female_i + \gamma Low\_Stake\_Test_{ilsyt} + \delta Female_i * Low\_Stake\_Test_{ilsyt} + \varepsilon_{ilsyt},$$

where the outcome variable is the standardized test score; *Female* refers, as before, to a dummy for female students; and *Low_Stake_Test* is a dummy variable that takes value one when the test score refers to that of low-stakes test and zero otherwise.[9]

Table 3 presents the estimated parameters for equation (2). Columns 1 and 2 compare the low-stakes versus high-stakes tests; columns 2 and 3 compare the low-stakes versus medium-stakes tests; and, finally, columns 5 and 6 compare the medium-stakes versus high-stakes tests. The estimates are done with and without the inclusion of

[9] Similarly, for the comparison of medium stakes and high stakes, the dummy we define is *Medium Stake Test*, which takes the value of 1 when the test is that of medium stakes.

student fixed effects. In all regressions, the coefficient of interest is the interaction between female and the lower-stakes test type.

Consistent with Table 2, we find that female students perform significantly better on low-stakes than on high-stakes tests, by 0.06 standard deviations, and compared to medium-stakes tests, by 0.05 standard deviations. Furthermore, the results do not change when using an individual fixed-effect estimation. Hence, from now on, we will show the results using an OLS regression. From columns 5 and 6, we see that there is no significant gender difference when comparing medium- and high-stakes tests. Thus, in the analysis that follows, we will focus only on comparisons between low- and high-stakes tests.

To understand whether gender differences change along the students' academic career, we explore the gaps for each academic level separately. Table 4 presents the estimation results separately for Levels 1 to 6. Overall, we see similar patterns across all levels, with the exception of Level 5, where female students outperform male students in low-stakes relative to high-stakes tests, but the difference is not significant. The magnitudes do, however, vary. In Levels 1 to 3, the difference is 0.05 standard deviations, while in Levels 4 and 6the effect is double (0.10 standard deviations). In Level 4, students decide whether or not to stay at the same school for the last two years prior to entering university. Since the grades obtained in the last two years count toward the university entry grade, this decision is important, as it will affect students' access to university. In Level 6, the students sit their Selectividad exams and the weight on their high-school grades is sizeable.

Looking at the distribution of performance (Table 5), we find that female students outperform male students in low-stakes relative to high-stakes tests at all points of the distribution. Although the magnitudes vary, there does not appear to be any systematic pattern moving from the lowest to highest performers.

Students undertake several subjects, typically ten or 11, per academic year. In Table 6, we disaggregate the analysis by subject type, classifying subjects as either Arts or Science.[10] Women and men have traditionally shown performance differences that depend on the type of subjects, where, following the stereotype, women are, on average, more likely to outperform men in the Arts rather than in Science.

[10] Table A.1 in the Appendix classifies subjects between science and arts subjects.

Table 6 shows the estimation results for gender difference for science and arts subjects separately, for the final grade (columns 1 and 2), for low-stakes tests (columns 3 and 4), for high-stakes tests (columns 5 and 6). In columns 7 and 8 we bring the specifications together by interacting gender with test stakes for science and art subjects, respectively. First, looking at the final scores, female students outperform male students in Arts subjects, while there is no significant gender difference in Science subjects. We clearly see that in low-stakes tests, females outperform males in Arts subjects but show no performance difference in Science subjects. In high-stakes tests, we find that while females outperform males in Arts subjects (0.25 standard deviations), they significantly underperform relative to male students in Science subjects (-0.10 standard deviations). The gender differences depend on the stakes of the tests, as well as on the type of subject. The interactions in columns 7 and 8 show that gender differentials regarding the stakes are significant and positive in both Arts and Science subjects.

3.2. Robustness of the main finding

We perform three robustness checks with respect to the baseline results. The results are shown in Table 7. For ease of exposition, column 1 replicates the overall gender difference in tests with different stakes (column 1 in Table 3). In column 2, we include the teacher fixed effects, in order to see whether teachers' information explains any of the gender differences at different stakes.[11] In column 3, we restrict the analysis to a balanced sample of students who complete all six levels. Finally, in column 4, we exclude student-subjects who have been exempt from high-stakes test. The last two robustness checks rule out the possibility that the gender differential effect is driven by students who leave school and who are exempt from taking high-stakes tests.

From column 2, we find that including teacher fixed effects does not change the size or the significance of the gender difference in low-stakes versus high-stakes test performance. This suggests that the effect is not driven by teachers, as the size and significance of the coefficient does not change by adding teachers' fixed effects. Similarly, in column 3, we find that restricting the analysis to students who stay in school has a quantitatively small effect. We again see that the interaction effect is significant and positive. Finally, students performing very well in low-stakes and

[11] We have teacher information for academic years 2000 to 2009.

medium-stakes tests, such that they are getting the highest possible grade throughout the course, may be given an exemption from the high-stakes test. Although this is typically a very small number of students per subject, sometimes none, we check the robustness of our results from excluding them. In column 4, we see the same result, both quantitatively and qualitatively, thus showing the robustness of the gender differential effect that depends on the size of the stakes for each test.

In Appendix Tables A.3 and A.4, we complement some of this analysis by looking at students' likelihood of leaving the school and of being exempt from a high-stakes exam, respectively. From Table A.3, we see that there is no gender difference in leaving school. Moreover, although high-performing students, as characterized by the number of exemptions, are more likely to stay at the school, there is no gender differential in this. Table A.4 shows that, conditional on the low-stakes test score, female and male students have a similar exemption probability. In Level 1, the interaction is negative and significant, suggesting that female students have a smaller chance of exemption. However, across levels, there is no systematic pattern in the coefficient, which with similar magnitudes is positive, although insignificant, in some levels (Levels 3, 4 and 5).

3.3. Extension: Understanding the gender differential in reaction to stakes

In this section, we investigate some hypotheses that might help to clarify the reasons for gender differences in performance on tests of different stakes. We focus on two main hypotheses: (1) the timing of the tests; and (2) the pressure of the test itself.

The timing for the low-stakes and high-stakes tests differs. Low-stakes tests are administered throughout the academic year, while high-stakes tests are administered only at the end of the academic year (see Figure 1 for the timing of exams). One argument might be that, rather than stakes, these estimates capture a gender difference in exam timing. To understand if this is the case, we look at the performance in each of the eight low-stakes exams. Table 8 reports the results. If the gender difference is caused by the later-in-the-academic-year characteristic of the high-stakes test, we should expect that as the academic year progresses, the gender difference in low-stakes relative to the high-stakes test will become smaller. Note that the low-stakes tests in Low-Stakes Test 8 are held up until one week before the high-stakes exam. From Table 8, we see that there is no significant difference in performance across different low-stakes tests (column 1). Moreover, the estimation results clearly show that there is no gender differential effect that depends on the timing of the different low-stakes tests. This helps to rule out that the different timing of the low- and high-stakes tests drives the identified gender differential effect. We do not observe a gender difference in time allocation in the low-stakes test. That is, it does not seem that male students wait until the end of the year to study since they do not exhibit a larger increase than girls in their test score in the later low-stakes tests.

Finally, the structure of the Level 6 test system provides us with an efficient way to check that it is the pressure of high-stakes tests that explains the gender differences. We can exploit the variation at the subject level, which naturally occurs in Level 6. In this level, as in all the other levels, students take around ten or 11 subjects. In high school, all subjects are examined in the same way as in all other levels. However, one important difference is that the national-level examination (i.e., the Selectividad) tests students on only around five subjects. The format and material covered in Selectividad subjects are the same as in high school. Recall that for the university entry grade, the high school Level 6 test score (for all ten or 11 subjects) counts for 25 percent of the final grade. The super-high-stakes test scores of the Selectividad (for the smaller set of subjects) counts for 50 percent. We can, therefore, classify subjects as those that "count" (i.e., Selectividad subjects) and those that "do not count" (i.e., non-Selectividad subjects). Whether the subject counts or not will have implications for their stakes. Thus, we use the variation in whether the exam counts to see if female and male students perform differently in their high school tests.

Table 9 shows the results for the final grade, low-stakes and high-stakes tests, respectively. Three results are noteworthy: first, the first three columns show that female students outperform male students in all three types of tests, whether or not it is a Selectividad subject. However, as in the baseline specifications, the gender gap is smaller in high-stakes exams. Second, students perform relatively better in subjects in which they will sit a Selectividad exam than in those that they will not, regardless of the exam's stakes. Finally, from columns 4 to 6, looking at the interaction between gender and whether or not the subject is a Selectividad subject, we clearly see that the differential gender effect is negative and significant for the final score, as well as for the low-stakes tests. This is indicative of the importance of exam stakes. All subjects, whether or not they are included on the Selectividad, count the same for the high school final grade. It might be that students prepare more for the Selectividad subjects because the national exams carry greater weight, as is suggestive of the difference in performance of all students. However, the gender differential, especially in high-stakes exams, shows that pressure, in the form of what the exam will eventually count for, generates a different reaction among female and male students. Female students underperform relative to male students in subjects that are considered higher-stakes compared to those considered lower-stakes. This result further confirms that men and women respond differently to different stakes.

4. Discussion and Implications

A recent literature has acknowledged that a behavioral reaction referred to as "choking under pressure" can impact individuals' future success. An increase in stakes generally brings an increase in incentives and, therefore, an improvement in performance. However, if the increase in stakes is too large, individuals may "choke" and actually do worse. We provide empirical evidence that the propensity to choke under an increase in the stakes of an exam is different for females and males. In particular, outperformance by females is reduced or even disappears as the stakes are increased substantially.

It is important to understand the consequences of gender differences in choking under pressure. Here, we discuss the potential labor-market implications of our findings that males and females have different levels of pressure-tolerance.

If selection into jobs is unaffected by individuals' pressure-tolerance, and if pressure tolerance is payoff-relevant to the worker through his or her productivity, then we would expect that, all else equal, female workers would earn a lower wage than male workers. In particular, for jobs in which pressure is high, we may expect female workers to underperform relative to their male counterparts.

Prior to entering the labor market, however, individuals typically must go through a selection process. Individuals self-select into occupations and firms, and employers in those firms select those that they want to hire. Pressure in the selection process may lead to disregarding individuals with lower pressure-tolerance, at the cost of other, potentially productive skills. A firm will interview workers to measure their skills, and candidates, when interviewed, provide a signal of their skills that may be affected by their tolerance of pressure. The level of pressure in the signaling process may impose a tradeoff in the productivity of the selected candidates. That is, increased pressure in the selection process will lead to a workforce with a higher tolerance for pressure, but this might come at the expense of other relevant skills.

In anticipation of the importance of pressure in the signaling process and in the firms' valuation of pressure in the workplace, individuals concerned with their labor-market success may self-select into certain occupations and firms. For instance, an individual with low pressure-tolerance will avoid firms that reward pressure-tolerance, either in the interview stage or on the job, unless other skills can compensate for this low tolerance. This might be one explanation of occupational segregation, whereby there is an over- (under-) representation of females in some professions while not in others. Moreover, recent evidence highlights that females, unless top-performing, abandon certain college majors, unlike their male counterparts.[12] The concern is that these majors tend to lead to high-paying jobs and, again, might explain gender wage gaps. In line with our reasoning, if these jobs involve high pressure, lower-performing female students are rationally switching majors.

In this paper, we find that the increased pressure arising from an increase in the stakes induces different reactions in females and males. So far, we have denoted pressure as a single-dimension variable. However, as the literature in social psychology has acknowledged, there are different sources of pressure, and the reaction to each form may be different for different individuals (see Beilock, 2011). For instance, the economics literature has also emphasized the asymmetric reactions of males and females to competitive environments, even when the stakes are not particularly high. Understanding the different sources of pressure and how they affect different groups in society is important to understanding the potential inefficiencies in the labor market.

[12] A recent article in the *Washington Post*, "Women should embrace the B's in college to make more later," by C. Rampell, cites P. Arciadiacono's words: "STEM majors, as with economics, begin with few women enrolling and end with even fewer graduating. This *leaky pipeline* has been somewhat puzzling, because women enter college just as prepared as men in math and science." Similarly, work by Goldin is cited showing that the fraction of females graduating in economics is highly decreasing with their grade in their introductory economics class in the first year, which is not true for male students.

References

Altonji, J. and R. Blank (1999), "Race and Gender in the Labor Market" in O. Ashenfelter and D. Card, (eds), Handbook of Labor Economics, Volume 3c, Elsevier Science B.V.: 3144-3259.

Angrist, J. and V. Lavy (2009), "The Effects of High Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial", *The American Economic Review,* 99 (4): 1384-1414.

Antonovics, K., P. Arciadiacono and R. Walsh (2009), "The Effects of Gender Interactions in the Lab and in the Field", *The Review of Economics and Statistics,* 91 (1): 152-162.

Ariely, D., U. Gneezy, G. Loewenstein and N. Mazar (2009), "Large Stakes and Big Mistakes", *Review of Economic Studies*, 76 (2): 451-469.

Baumeister, R. F. (1984), "Chocking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance", *Journal of Personality and Social Psychology,* 46 (3): 610-20.

Beilock, S. (2011), *Choke: What the Secrets of the Brain Reveal About Getting it Right When You have To,* by Simon and Schuster, Free Press.

Bertrand, M. (2010), "New Perspectives on Gender", in O. Ashenfelter and D. Card, (eds), Handbook of Labor Economics, Volume 4b, Elsevier Science B.V.: 1545-1592.

Buser, T., M. Niederle and H. Oosterbeek (2014), "Gender, Competitiveness and Career Choices", *The Quarterly Journal of Economics*, 129 (3).

Camerer, C. F. and R. Hogarth (1999), "The Effect of Financial Incentives in Experiments: a Review and Capital-Labor-Production Framework", *Journal of Risk and Uncertainty,* 19 (1), 7-42.

Camerer, C.F., G. Lowenstein and D. Prelec (2005), "Neuroeconomics: How Neuroscience Can Inform Economics", *Journal of Economic Literature,* 43 (1): 9-64.

Cornwell, C., D. Mustard and J. Van Parys, (2013), "Non-cognitive Skills and Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School", *Journal of Human Resources*, 48 (1): 236-264.

Gneezy, U., M. Niederle and A. Rustichini (2003), "Performance in Competitive Environments: Gender Differences", *The Quarterly Journal of Economics,* 118 (3), 1049-1074.

Gneezy, U. and A. Rustichini (2004), "Gender and Competition at a Young Age", *American Economic Review P&P,* 94(2), 377-381.

Goldin, C., L. Katz, and I. Kuziemko, (2006) “The Homecoming of American College Women: The Reversal of the Gender Gap in College”, *Journal of Economic Perspectives,* 20:133-156.

Iriberri N. and P. Rey-Biel (2011) “On Women’s Underperformance in Competitive Environments: When and Why", UPF working paper 1288.

Langer, E. J. and L.G. Imber (1979), “When Practice Makes Imperfect: The Debilitating Effects of Overlearning”, *Journal of Personality and Social Psychology,* 37 (11), 2014-2024.

Lavy, V. (2008) “Do Gender Stereotypes Reduce Girls’ or Boys’ Human Capital Outcomes? Evidence from a Natural Experiment”, *Journal of Public Economics,* 92 (10-11): 2083-2105.

Niederle, M. and L. Vesterlund (2007), “Do Woman Shy Away from Competition? Do Men Compete too Much?” *The Quarterly Journal of Economics,* 122 (3): 1067-1101.

Örs, E., F. Palomino and E. Peyrache (2013), “Performance Gender-Gap: Does Competition Matter?” *Journal of Labor Economics,* 31 (3): 443-449.

Shurchkov, O. (2012), “Under Pressure: Gender Differences in Output Quality and Quantity Under Competition and Time Constraints,” *Journal of the European Economic Association* 10(5): 1189–1213.

Yerkes, R.M. and J. D. Dodson (1908), “The Relationship of Strength of Stimulus to Rapidity of Habit-Formation,” *Journal of Comparative Neurology of Psychology,* 18 (5): 459-482.

Tables and Figures

Table 1: Descriptive Statistics

	OVERALL			Level 1			Level 2			Level 3			Level 4			Level 5			Level 6		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
No. Low-Stakes Tests	38146	6.78	1.19	8994	7.01	0.97	7974	7.03	0.97	6904	6.83	0.99	6435	6.82	1.02	4626	6.6	1.62	3213	5.59	1.47
No. Medium-Stakes Tests	38146	1.98	0.16	8994	2.00	0.07	7974	2.00	0.04	6904	1.97	0.16	6435	2.00	0.05	4626	1.97	0.23	3213	1.93	0.37
No. High-Stakes Tests	38146	1.00	0.06	8994	1.00	0	7974	1.00	0	6904	1.00	0.00	6435	1.00	0.00	4626	0.99	0.11	3213	0.97	0.17
No. Students	38146	309.53	97.53	8994	85.12	6.49	7974	78.17	5.37	6904	69.74	5.12	6435	59.46	7.67	4626	41.7	13.7	3213	35.49	17.06
Prop. Female Students	38146	0.50	0.06	8994	0.49	0.07	7974	0.50	0.05	6904	0.51	0.06	6435	0.52	0.05	4626	0.51	0.07	3213	0.51	0.06
No. Subjects	38146	8.11	1.09	8994	7.96	0.25	7974	7.91	0.31	6904	7.83	1.02	6435	8.76	1.54	4626	8.86	0.99	3120	7.22	1.56
Prop. Leavers	38146	0.10	0.03		--	--	7974	0.03	0.04	6904	0.07	0.05	6435	0.1	0.05	4626	0.21	0.04	3213	0.1	0.05
Prop. Students Exempt	38146	0.05	0.08	8555	0.04	0.07	7636	0.04	0.06	6553	0.05	0.07	5753	0.09	0.10	4334	0.05	0.10	3157	0.00	0.02

Notes: The descriptive statistics are calculated over years 2000 to 2012. Each observation refers to a student-subject. *No. Low-Stakes Tests* are the number of low-stakes tests taken by students for a given subject in a given academic year. Students typically take low-stakes tests bi-monthly (per subject), such that they take around six low-stakes tests in terms 1 and 2 and four in term 3. We have information only on the average monthly low-stakes test scores. *No. Medium-Stakes Tests* the number of medium-stakes tests taken by students for a given subject in a given academic year. *No. High-Stakes Tests* are the number of high-stakes tests taken by students for a given subject in a given academic year. *No. Students* is the number of students in a given academic year. *Prop. Female Students* is the proportion of female students in a given academic year. *No. Subjects* is the number of subjects that students take in a given academic year. *Prop. Leavers* is the proportion of students who leave at the end of the previous academic year. *Prop. Students Exempt* is the proportion on students given an exemption from a high-stakes test in one (or more) of the subjects in a given academic year.

Table 2: Performance under Different Stakes

A: Performance under Different Stakes: Overall					
	Final Test Score	Low-Stakes Test Score	Medium-Stakes Test Score	High-Stakes Test Score	Super-High-Stakes Test Score
Female	0.159***	0.175***	0.126***	0.118***	-0.0301
	[0.0419]	[0.0438]	[0.0407]	[0.0374]	[0.0382]
Constant	-0.0799***	-0.0884***	-0.0636**	-0.0595**	0.016
	[0.0294]	[0.0309]	[0.0283]	[0.0263]	[0.0272]
Observations	38,146	38,637	38,247	38,857	2,598
R-squared	0.006	0.008	0.004	0.004	0.001

B: Performance under Different Stakes: Level 6 only					
	Final Test Score	Low Stakes Test Score	Medium Stakes Test Score	High Stakes Test Score	Super High Stakes Test Score
Female	0.169***	0.185***	0.0939***	0.0803**	-0.0301
	[0.0344]	[0.0336]	[0.0345]	[0.0332]	[0.0382]
Constant	-0.0820***	-0.0937***	-0.0467*	-0.0407*	0.016
	[0.0244]	[0.0239]	[0.0246]	[0.0236]	[0.0272]
Observations	3,213	3,372	3,207	3,473	2,598
R-squared	0.008	0.009	0.002	0.002	0.001

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Final Test Score* is the students' accumulated test score at the end of the academic year. *Low-Stakes Test Score* is the students' test score in the low-stakes tests; *Medium-Stakes Test Score* is the students' test score in the medium-stakes tests; *High-Stakes Test Score* is the students' test score in the high-stakes tests; and *Super-High-Stakes Test Score* is the students' test score in the national exams, Selectividad, taken in Level 6. All test scores are standardized to a distribution with zero mean and a unit standard deviation. The standardization is done by academic year, level, subject and test type.

Table 3: Performance under Different Stakes with Interactions

	Low-Stakes vs. High-Stakes		Low-Stakes vs. Med.-Stakes		Med.-Stakes vs. High-Stakes	
	Test Score	Test Score	Test Score	Test Score	Test Score	Test Score
Female	0.118***		0.126***		0.118***	
	[0.0374]		[0.0407]		[0.0374]	
Low-Stakes Test	-0.0290***	-0.0277***	-0.0248***	-0.0253***		
	[0.00897]	[0.00899]	[0.00730]	[0.00718]		
Female*Low-Stakes Test	0.0574***	0.0579***	0.0490***	0.0513***		
	[0.0124]	[0.0124]	[0.00988]	[0.00979]		
Med.-Stakes Test					-0.00418	
					[0.00722]	
Female*Med.-Stakes Test					0.00845	0.0048
					[0.0102]	[0.00719]
Constant	-0.0595**	-0.000747	-0.0636**	-0.00026	-0.0595**	-0.00116
	[0.0263]	[0.00309]	[0.0283]	[0.00246]	[0.0263]	[0.00180]
Observations	77,494	77,494	76,884	76,884	77,104	77,104
R-squared	0.006	0	0.006	0	0.004	0
No. of students		1,404		1,404		1,404
Student FE	No	Yes	No	Yes	No	Yes

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Low-Stakes vs. High-Stakes* compares low-stakes test scores with high-stakes test scores. *Low-Stakes vs. Med.-Stakes* compares low-stakes test scores with medium-stakes test scores. *Med.-Stakes vs. High-Stakes* compares medium-stakes test scores with high-stakes test scores. *Low-Stakes Test* takes the value one if the test is a low-stakes test and zero otherwise. *Med.-Stakes Test* takes the value one if the test is a medium-stakes test and zero otherwise. All test scores are standardized.

Table 4: Performance under Different Stakes with Interactions by Level

	Level 1 Test Score	Level 2 Test Score	Level 3 Test Score	Level 4 Test Score	Level 5 Test Score	Level 6 Test Score
Female	0.210***	0.127***	0.119**	0.026	0.0796	0.0803
	[0.0436]	[0.0459]	[0.0486]	[0.0532]	[0.0566]	[0.0595]
Low-Stakes Test	-0.0233*	-0.0247	-0.0244*	-0.0492***	-0.00791	-0.0529**
	[0.0130]	[0.0159]	[0.0148]	[0.0153]	[0.0202]	[0.0218]
Female*Low-Stakes Test	0.0472**	0.0489**	0.0482**	0.0957***	0.0155	0.105***
	[0.0189]	[0.0219]	[0.0215]	[0.0210]	[0.0270]	[0.0332]
Constant	-0.104***	-0.0639**	-0.0601*	-0.0134	-0.0406	-0.0407
	[0.0303]	[0.0324]	[0.0327]	[0.0375]	[0.0421]	[0.0432]
Observations	18,112	16,026	14,156	13,108	9,247	6,845
R-squared	0.014	0.006	0.005	0.002	0.002	0.005

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The regressions compares low-stakes test scores with high-stakes test scores. All test scores are standardized. *Low-Stakes Test* takes the value one if the test is a low-stakes test and zero otherwise.

Table 5: Performance under Different Stakes with Interactions by Quantiles

	Q10 Test Score	Q20 Test Score	Q30 Test Score	Q40 Test Score	Q50 Test Score	Q60 Test Score	Q70 Test Score	Q80 Test Score	Q90 Test Score
Female	0.108***	0.106***	0.111***	0.112***	0.140***	0.207***	0.153***	0.104***	0.0814***
	[0.0117]	[0.00938]	[0.0145]	[0.0139]	[0.0127]	[0.0215]	[0.0138]	[0.0181]	[0.0136]
Low-Stakes Test	0.0444***	-0.0170**	-0.0158	-0.0598***	-0.0581***	-0.108***	-0.172***	-0.102***	0.0274
	[0.0119]	[0.00847]	[0.0127]	[0.0143]	[0.0111]	[0.0257]	[0.0205]	[0.0192]	[0.0203]
Female*Low-Stakes Test	0.0425***	0.0636***	0.0769***	0.0832***	0.0679***	0.00819	0.0594**	0.0604**	0.0377*
	[0.0122]	[0.0137]	[0.0182]	[0.0184]	[0.0182]	[0.0267]	[0.0241]	[0.0263]	[0.0217]
Constant	-1.276***	-0.932***	-0.702***	-0.441***	-0.200***	0.130***	0.533***	0.914***	1.341***
	[0.00909]	[0.00657]	[0.0106]	[0.0102]	[0.00615]	[0.0202]	[0.0130]	[0.00966]	[0.00994]
Observations	77,494	77,494	77,494	77,494	77,494	77,494	77,494	77,494	77,494

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Q10* to *Q90* represent the 10^{th} to 90^{th} deciles. The regressions compares low-stakes test scores with high-stakes test scores. All test scores are standardized. *Low-Stakes Test* takes the value one if the test is a low-stakes test and zero otherwise.

Table 6: Performance under Different Stakes with and without Interactions for Science versus Arts Subjects

	Final Test Score		Low-Stakes Test Score		High-Stakes Test Score		Low vs. High Test Score	
	Science	Arts	Science	Arts	Science	Arts	Science	Arts
Female	-0.0689	0.298***	-0.0499	0.312***	-0.101***	0.250***	-0.101***	0.250***
	[0.0452]	[0.0438]	[0.0468]	[0.0458]	[0.0388]	[0.0407]	[0.0388]	[0.0407]
Low-Stakes Test							-0.0252*	-0.0316***
							[0.0134]	[0.00902]
Female*Low-Stakes Test							0.0507***	0.0622***
							[0.0176]	[0.0127]
Constant	0.0358	-0.152***	0.0249	-0.159***	0.0501*	-0.127***	0.0501*	-0.127***
	[0.0322]	[0.0309]	[0.0342]	[0.0319]	[0.0275]	[0.0288]	[0.0275]	[0.0288]
Observations	14,405	23,741	14,589	24,048	14,623	24,234	29,212	48,282
R-squared	0.001	0.023	0.001	0.025	0.003	0.016	0.002	0.02

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Final Test Score* is the students' accumulated test score at the end of the academic year. *Low-Stakes Test Score* is the students' score on the low-stakes tests, *High-Stakes Test Score* is the students' score on the high-stakes tests. *Low-Stakes vs. High-Stakes* compares low-stakes test scores with high-stakes test scores. Science subjects are subjects that are classified as science subjects. Art are subjects that are classified as humanity subjects. A full list of subjects is given in Table A.1.

Table 7: Robustness checks: Teacher Fixed Effects, Staying Students and Students Exempt from the High-Stakes Test

	Test Scores	Test Scores	Test Scores	Test Scores
Female	0.118***	0.158*	0.0887	-0.125***
	[0.0374]	[0.0887]	[0.0553]	[0.0349]
Low-Stake Tests	-0.0290***	-0.0254**	0.0216	0.0325***
	[0.00897]	[0.0106]	[0.0145]	[0.00860]
Female*Low-Stakes Test	0.0574***	0.0502***	0.0458**	-0.0655***
	[0.0124]	[0.0140]	[0.0191]	[0.0125]
Constant	-0.0595**	-0.086	0.194***	0.0619**
	[0.0263]	[0.0670]	[0.0391]	[0.0253]
Observations	77,494	54,520	37,020	72,774
R-squared	0.006	0.021	0.004	0.007
Teacher FE	No	Yes	No	No
Students in sample for all 6 levels	No	No	Yes	No
Student-subjects not exempt	No	No	No	Yes

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The regressions compare low-stakes test scores with high-stakes test scores. All test scores are standardized. *Low-Stakes Test* takes the value one if the test is a low-stakes test and zero otherwise. Column 2 includes teacher fixed effects. Teacher identity is available only for years 2000 to 2009. Column 3 is restricted to cohorts of students who are in the sample for all six levels. Column 4 excludes test scores for when students are exempt from sitting the high-stakes exams.

Table 8: Exploiting the different Timing of Low-Stakes Tests

	Test Score	Test Score
Female	0.142***	0.151***
	[0.0357]	[0.0352]
Low_Stakes_Test2	-9.24E-05	-0.00308
	[0.00561]	[0.00760]
Low_Stakes_Test3	1.79E-05	0.00625
	[0.00631]	[0.00871]
Low_Stakes_Test4	-0.000187	0.00423
	[0.00492]	[0.00666]
Low_Stakes_Test5	-6.54E-05	0.00734
	[0.00588]	[0.00800]
Low_Stakes_Test6	0.000534	-0.000204
	[0.00644]	[0.00879]
Low_Stakes_Test7	0.000132	0.00922
	[0.0162]	[0.0223]
Low_Stakes_Test8	-0.000233	0.0231
	[0.0166]	[0.0234]
Female*Low_Stakes_Test2		0.00591
		[0.0112]
Female* Low_Stakes_Test3		-0.0123
		[0.0126]
Female* Low_Stakes_Test4		-0.00876
		[0.00982]
Female* Low_Stakes_Test5		-0.0147
		[0.0117]
Female* Low_Stakes_Test6		0.00147
		[0.0129]
Female* Low_Stakes_Test7		-0.0181
		[0.0323]
Female* Low_Stakes_Test8		-0.0463
		[0.0332]
Constant	-0.0715***	-0.0757***
	[0.0250]	[0.0245]
Observations	262,568	262,568
R-squared	0.005	0.005

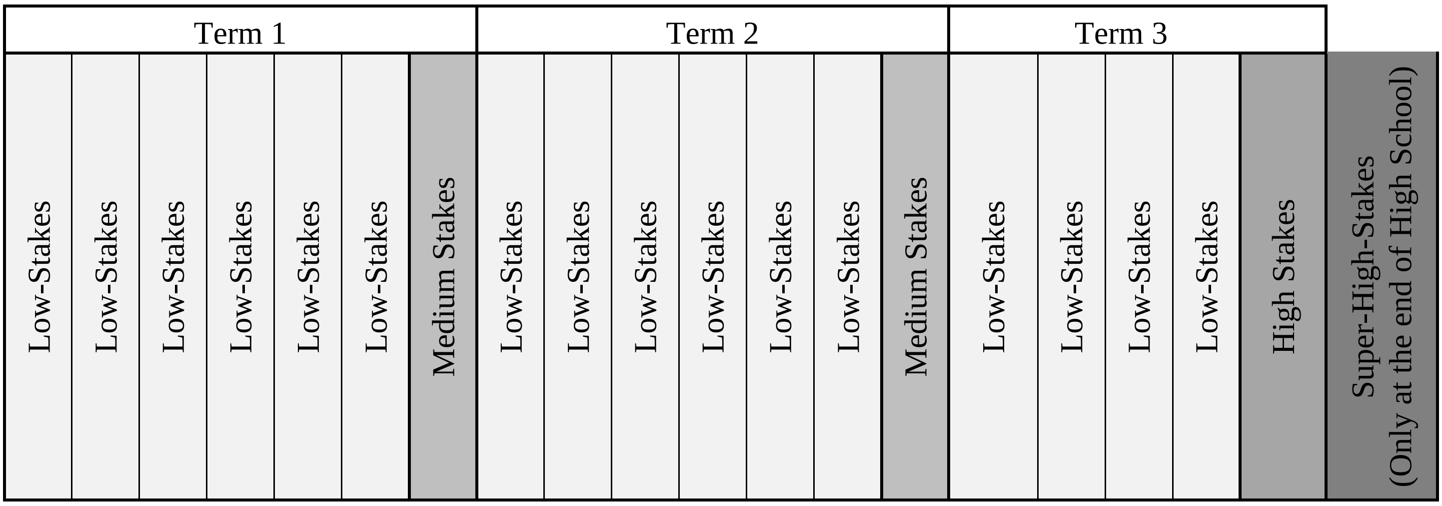
Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The regressions compare test scores on low-stakes tests taken throughout the year. *Low_Stakes_Test2* to *Low_Stakes_Test8* are the second to last low-stakes test in an academic year. The first low-stakes test is the omitted category. All test scores are standardized.

Table 9: Selectividad Subject versus Non-Selectividad Subject (Level 6 only).

	Final Test Score	Low-Stakes Test Score	High-Stakes Test Score	Final Test Score	Low-Stakes Test Score	High-Stakes Test Score
Female	0.170***	0.187***	0.0829**	0.316***	0.336***	0.127*
	[0.0343]	[0.0341]	[0.0344]	[0.0718]	[0.0713]	[0.0719]
Selectividad Subject	0.117***	0.112***	0.148***	0.214***	0.210***	0.177***
	[0.0409]	[0.0406]	[0.0409]	[0.0583]	[0.0580]	[0.0584]
Female*Selectividad Subject				-0.189**	-0.193**	-0.057
				[0.0817]	[0.0812]	[0.0818]
Constant	-0.173***	-0.184***	-0.141***	-0.247***	-0.260***	-0.164***
	[0.0400]	[0.0397]	[0.0400]	[0.0513]	[0.0510]	[0.0514]
Observations	3,213	3,213	3,213	3,213	3,213	3,213
R-squared	0.01	0.012	0.006	0.012	0.013	0.006

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Final Test Score* is the students' accumulated test score at the end of the academic year. *Low-Stakes Test Score* is the students' test score on the low-stakes tests, *High-Stakes Test Score* is the students' test score on the high-stakes tests. *Selectividad Subject* are subjects in which a students will do the national exams at the end of Level 6.

Figure 1: Evaluation System in the School



Notes: This is the evaluation system used in each subject. *Low-Stakes* is the test that each counts for around 2.5 percent of the final grade. *Medium-Stakes* is the test that each counts for around 11 percent of the final grade. *High-Stakes* is the test that counts for around 27 percent of the final grade. *Super-High-Stakes* is the national exams, Selectividad, taken at the end of Level 6, which counts for 50 percent of the university entry test score.

Tables and Figures in the Appendix

Table A.1: List of Subjects and their Classification as Science or Arts Subject

List of Subjects	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Arts-Science
Spanish	X	X	X	X	X	X	Arts
Catalan	X	X	X	X	X	X	Arts
English	X	X	X	X	X	X	Arts
French	X	X	X	X	X	X	Arts
Math	X	X	X	X	X	X	Science
Biology	X	X	X	X	X	X	Science
History-Geography	X	X	X	X	X	X	Arts
IT	X	X	X	X	X		Science
Chemistry-Physics			X	X	X	X	Science
Latin			X	X	X		Arts
Technical Drawing			X		X	X	Science
Art History					X	X	Arts
Contemporary Sciences					X		Arts
Economics					X	X	Science
Math Applied to Social Sciences					X	X	Arts
Philosophy					X	X	Arts
Audiovisual Culture						X	Arts
History Philosophy						X	Arts

Table A.2: Probability of Leaving School

	Overall Pr.(Staying)	Level 2 Pr.(Leaving)	Level 3 Pr.(Leaving)	Level 4 Pr.(Leaving)	Level 5 Pr.(Leaving)
Female	-0.0254	-0.0378	0.0164	0.0139	0.0119
	[0.0377]	[0.0233]	[0.0218]	[0.0314]	[0.0157]
No. Exemptions	0.0289***	-0.0141*	-0.011	-0.0306***	-0.00427
	[0.00643]	[0.00801]	[0.00737]	[0.00902]	[0.00343]
Female*No. Exemptions	0.00346	0.00596	-0.00192	0.0144	0.00241
	[0.00867]	[0.0105]	[0.00980]	[0.0100]	[0.00427]
Observations	875	875	825	771	707
Controls	Yes	Yes	Yes	Yes	Yes

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Pr. (Staying)* is the probability that the student will stay in school for all six levels. *Pr. (Leaving)* is the probability that the student will leave at the end of the previous academic year. *No. Exemptions* is the number of exemptions (at the subject level) a student has from sitting high-stakes tests. Controls include year and subject dummies.

Table A.3: Probability of being Exempt from High-Stakes Test

	Overall Pr.(Exempt)	Level 1 Pr.(Exempt)	Level 2 Pr.(Exempt)	Level 3 Pr.(Exempt)	Level 4 Pr.(Exempt)	Level 5 Pr.(Exempt)	Level 6 Pr.(Exempt)
Female	0.00268	0.138**	0.0931*	-0.133	-0.111	-0.0928	0.0482
	[0.0351]	[0.0542]	[0.0556]	[0.0934]	[0.0776]	[0.0958]	[0.0426]
Score_LS_Test	0.0646***	0.0806***	0.0718***	0.0567***	0.107***	0.0558***	-0.00128
	[0.00401]	[0.00665]	[0.00576]	[0.0111]	[0.00824]	[0.0120]	[0.00328]
Female* Score_LS_Test	0.00103	-0.0214**	-0.0114	0.0227	0.0185	0.014	-0.00581
	[0.00532]	[0.00841]	[0.00866]	[0.0138]	[0.0117]	[0.0136]	[0.00557]
Observations	38,409	8,824	8,013	7,079	6,555	4,565	3,373
R-squared	0.144	0.201	0.175	0.229	0.237	0.235	0.198
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. *Pr. (Exempt)* is the probability that the student is exempt from a subject's high-stakes test. *Score_LS_Test* is the average test score in the low-stakes tests in the subject from which the student is exempt. Controls include year and subject dummies.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1313	Saul Estrin Ute Stephan Sunčica Vujić	Do Women Earn Less Even as Social Entrepreneurs?
1312	Nicholas Bloom Renata Lemos Raffaella Sadun John Van Reenen	Does Management Matter in Schools?
1311	Erling Barth Alex Bryson James C. Davis Richard Freeman	It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the US
1310	Christos Genakos Pantelis Koutroumpis Mario Pagliero	The Impact of Maximum Markup Regulation on Prices
1309	Gianmarco I.P. Ottaviano Filipe Lage de Sousa	Relaxing Credit Constraints in Emerging Economies: The Impact of Public Loans on the Performance of Brazilian Manufacturers
1308	William Fuchs Luis Garicano Luis Rayo	Optimal Contracting and the Organization of Knowledge
1307	Alex Bryson Richard B. Freeman	Employee Stock Purchase Plans – Gift or Incentive? Evidence from a Multinational Corporation
1306	Andrew E. Clark Sarah Flèche Claudia Senik	Economic Growth Evens-Out Happiness: Evidence from Six Surveys
1305	Jorge De la Roca Gianmarco I.P. Ottaviano Diego Puga	City of Dreams

1304	Jan-Emmanuel De Neve George W. Ward Femke De Keulenaer Bert Van Landeghem Georgios Kavetsos Michael I. Norton	Individual Experience of Positive and Negative Growth is Asymmetric: Evidence from Subjective Well-being Data
1303	Holger Breinlich Anson Soderbery Greg C. Wright	From Selling Goods to Selling Services: Firm Responses to Trade Liberalization
1302	Esteban Aucejo Teresa Foy Romano	Assessing the Effect of School Days and Absences on Test Score Performance
1301	Gianmarco I.P. Ottaviano	European Integration and the Gains from Trade
1300	Antoine Dechezleprêtre Ralf Martin Myra Mohnen	Knowledge Spillovers from Clean and Dirty Technologies
1299	Stephen Machin Richard Murphy	Paying Out and Crowding Out? The Globalisation of Higher Education
1298	Iain Cockburn Jean O. Lanjouw Mark Schankerman	Patents and the Global Diffusion of New Drugs
1297	David W. Johnston Grace Lordan Michael A. Shields Agne Suziedelyte	Education and Health Knowledge: Evidence from UK Compulsory Schooling Reforms
1296	Fabio Pinna Stephan Seiler	Consumer Search: Evidence from Path-tracking Data
1295	Matt Dickson Paul Gregg Harriet Robinson	Early, Late or Never? When Does Parental Education Impact Child Outcomes?

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7404 0612
Email info@cep.lse.ac.uk Web site http://cep.lse.ac.uk