Efficient Multitasking in Space and Time and the Ability and Gender Gaps in Cognitive Performance[*]

Preliminary and Incomplete

Victor Lavy, University of Warwick, Hebrew University and NBER

Hadar Avivi, University of California Berkeley

November_1_2018

Summary

Evidence from many countries suggest that girls outperform boys in most academic outcomes at pre-college schooling. For example, a recent OECD report (2015) about PISA Assessment (PISA), show that girls are significantly less likely than boys to score below the "proficient" threshold in math, reading and science. Similarly, in the US 2013 NAEP, girls outperform boys with similar or no gender difference in math test scores. In this paper we first show a similar pattern based on high school high stake exit exams in Israel. These exams are viewed as an assessment of the three years of secondary school and as in many OECD countries, they are used as a doorway to higher education. Viewing these tests as multi-tasking over time and space, students' performance may be affected by two parameters, the number of preparation days between exams and the chronological order of an exam in the sequence of tests. We show in this paper that the effect of these two factors vary by student gender and ability to an extent that their effect account jointly for about 40 percent of the gender gap in cognitive performance and for 30 percent of the gap between high and low ability students. The positive effect of days between exams is positive for girls and it increases with academic ability. The average effect on boys is positive as well but it is masking lack of an effect among students below median ability and positive effect for students above median ability. The effect on girls is higher than the effect on boys throughout the ability distribution. Second, we find that students' test performance improves as they advance in the sequence of exams, but this effect reflects mainly a positive effect among girls while the effect on boys is negative at all ability levels. Boys performance deteriorate as they advance in the sequence of exams while girls performance improves. This asymmetric effect by gender is so large that it explains meaningful part of the gender test score gap. We discuss We explore several explanations for this asymmetry, including that girls gain confidence and testing skills with more exam done, perhaps offsetting possible mental and physical fatigue that may increase with lengthy multitasking. We also find evidence that the effect of days (chronological test order) is larger (smaller) the higher is the chronological test order (number of days). We contrast our findings from recent evidence in the psychology literature on efficient multitasking and on related gender differences in such settings.

[*] We thank Gabriella Conti, Mirco Draca, Lucie Gadiene, Omer Moav, Roland Rathelot, and Moshe Shayo, for useful discussions and suggestions. The first author acknowledges financial support for this project from the European Research Council through ERC Advance Grant 3 3439, from CAGE at the Department of Economics at the University of Warwick and from the Falk research institute in Jerusalem.

.

1. Introduction

Evidence from many countries suggest that girls outperform boys in most academic outcomes at pre-college schooling. For example, a recent OECD report (2015) about PISA Assessment (PISA), show that girls are significantly less likely than boys to score below the "proficient" threshold in math, reading and science. Similarly, in the US 2013 NAEP, girls outperform boys with in reading and there is no gender difference in math test scores. In Israel's end of high school matriculation-exams, girls score higher in every subject and their overall composite grade is 17 percent higher than that of boys. The vast literature on the gender gap in cognitive outcomes provide however little guidance as to its origins (see ??? for a recent survey). In this paper we provide a partial explanation for this gap by exploiting the matriculation exams setup in Israel where students take multiple high-stake exams in a relatively short period of time that involve intense preparation under time pressure. We focus on the multitasking nature over space and time of the preparation towards these exams and identify two key parameters that are potential determinants of the gender gap in cognitive performance in this setup.[1] efficiency in using preparation time and level of mental ability needed for dealing with the intense and tiring effort and pressure.[2] The format and structure of these exams vary greatly across countries. These exams are taken mostly at end of high school and the number of subjects covered and number of exams vary greatly across countries, in some they include 3 exams and in other the program extends to 8- 10 exams and more. The exams take place usually during the last month of the last year of high school, in most countries from mid-May to mid or end of June. The number of exams

[1] The term *multitasking* is a loosely defined construct that covers a wide spectrum of activities and time frames. Multitasking in some conditions may require very narrow deadlines (e.g., air traffic control), whereas other types of multitasking (e.g., preparing for multiple exams) may impose lower demands on spatiotemporal processing because of more-generous time windows.

[2] Many countries administer a similar setup of exams at end of high school for the purpose of an assessment of the three years of secondary school and as a doorway to higher education. These exams have different names in different countries, *Baccalauréat* (France), *A-levels* (UK), *Abitur* (Germany), *Selectividad* (Spain), *Bagrut* (Israel) and the *Matura* in many European countries. Including Albania, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Hungary, Italy, Kosovo, Liechtenstein, Macedonia, Montenegro, Poland, Serbia, Slovakia, Slovenia, Switzerland and Ukraine. In the US the *SAT* (scholastic assessment test) and its close competitor, the *ACT* (American College Testing) are standardized college admissions tests rather than all-round final year exam qualifications – the US High School diploma tends to fulfill the latter function.

and the length of this exam period define two important parameters that affect students' performance: the length of time between adjacent exams and the order of an exam in the sequence of tests.

High exams density (ratio of number of exams to length days of the exam period) reduces the potential preparation time between exams, possibly affecting negatively cognitive performance. This effect will likely depend on how well a student studied throughout the academic year and on his ability to design and execute a study plan for the exam period. This efficiency factor may vary across students as a function of their cognitive ability and of non-cognitive skills such as organizational skills, ability to avoid temptations and postpone satisfaction. Therefore, a dense exams period may impair less the performance of the high ability students and girls who might have an advantage in terms of these non-cognitive skills. For example, girls at this age have lower time preference, and higher prioritization, concentration, and focusing skills. These skills are key in designing a study plan and sticking with it even under pressure and exam anxiety. Since high ability students may possess more of these skills, the high stake exams at end of high school may have distributional consequences beyond affecting the gender gap.

The second factor is the potential cognitive and mental fatigue that may come into play as students are down the line of exams. Consecutive tests may cause cognitive fatigue which might tax students' mental functions enough that it may cause a decrease in their performance (Ackerman and Kanfer (2009).. Moreover, multiple exams may cause perceptions of fatigue, or subjective fatigue which may lead to lead to a self-regulated withdrawal from the process and consequently lower performance (Ackerman and Kanfer (2009). However, against these factors that make it increasingly difficult to maintain motivation and effort, some students may gain confidence and skills while taking more tests, therefore performing better in late than early exams. For example, if indeed, as argued, girls underperform in high stake exams, this handicap may diminish in a lengthy exam period. Therefore, packing many exams towards the end of high school may have both efficiency and distributional consequences because of the potential heterogeneous effects it might have on students.

In this paper we estimate the causal effect of the preparation time between exams, and of test order of exams on the performance of high school students in matriculation exams at end of high school in Israel. Israel is among the countries with the largest number of subjects and tests at end of high school. The exams are known as the *Bagrut* and are a critical component of Israel's college admissions system, acting as a gatekeeper for the most selective schools, similar to the role played by high-stakes exams in other countries, such as the aforementioned SATs or A-levels in England. In Israel, access to college majors is also determined by *Bagrut* performance, with many lucrative professional programs requiring minimum overall average scores for admission, such as law and medicine. Furthermore, admission decisions in Israel are based almost entirely on concrete measures of student performance, with no weight assigned to extra-curricular activities or student essays. As a consequence, *Bagrut* scores can affect an individual's entire academic career, and subsequent labor market outcomes.

We explore two dimensions of heterogeneity, by student ability and by gender, which allow to assess the distributional consequences of the system of large number of exams in a relatively short period. The *Bagrut* study program includes 7 compulsory subjects and 2-3 elective subjects, mounting to at least nine exams, some done at end of 10^{th} and 11^{th} grade. Therefore, the time schedule of exams vary across students depending on their number of subjects. Combining data for six cohorts, 2000- 2005, generate large variation across and within students in the distribution of number of days between each two adjacent exams and in the order of exams. Using this within pupil variation allows neutralizing the selection of students into study programs which largely determine the number of exams and the test order, and also the number of preparation days between exams. This selection is negative since low ability students study fewer subjects and therefore have fewer exams and more time between each two exams while the study program of high ability students' incudes more subjects and therefore more exams and denser exams' period. However, students choose their study program in the beginning or end of 10^{th} grade, at least two years before the schedule of *Bagrut* exams at end of 12^{th} grade is announced. Therefore, the natural experimental

within pupil variation in the two treatments of interest is not endogenous, and can be viewed as 'quasi random' permitting causal identification and inference.

The results presented in this paper show two important regularities. First, number of days between each two adjacent exams have a negative OLS association with scores even after controlling for test and student's characteristics, including lagged test scores in the same subject. However, the negative effect is reversed once we control for pupil fixed effects. The positive effect of days between exams is positive for girls and it increases with academic ability. The average effect on boys is positive as well but it is masking zero effect for students below median ability and positive effect for students above median ability. The effect on girls is higher than the effect on boys throughout the ability distribution. Second, we find that test order is positive on average, but this effect reflect mainly a positive effect on girls while the effect on boys is negative and at all ability level. Boys performance deteriorate with test order why girls improve their test achievement as they progress towards the end of the exam period. This asymmetric effect by gender is interesting and large enough to explain meaningful part of the gender test score gap. A possible explanation is that girls gain confidence and test writing skills with more exam done and this factor offsets any mental and physical fatigue that naturally accompany an intensive testing season. We also find evidence that the effect of days (test order) is larger (smaller) the higher is the test order (number of days).

The rest of the paper is organized as follows. Section 2 present literature review focusing on studies in economics and psychology. Section 3 discusses the background information on high school exit exams in Israel and in other countries and how they are used for college admission. Section 4 describes the data and section 5 describes the empirical strategy. Section 6 presents the results on the short-term effects on high school outcomes and section 7 concludes.

2. Literature (Incomplete)

Time to Learn

The theoretical hypothesis that learning takes time has been considered in at least three related fields — educational (Bloom, 1974), cognitive (Ericsson and Charness, 1994; Simon, 1990), and neural (McClelland, McNaughton and O'Reilly, 1995). These fields concluded that learning takes place within an extended time period during which information is encoded, rehearsed, elaborated, and consolidated. Both in economics and psychology past research has found support for this hypothesis and has shown that school length, instruction time and perpetration time have effects on student's achievements in tests. However, while most of the previous literature have largely been devoted to identifying the effect of increased time spent in class or in school, the evidence on the effect of the preparation time high-school students have prior the exam does not exist and evidence of that effect on college students are sparse. Taraban, Rynearson and Stalcup (2001) is an example of a study in psychology that investigates the direct relationship between preparation time and exams scores. They used computer records from an undergraduate course website that contain most of the course materials and lectures and found a positive and significant correlation between the students' success in the course exams and the time that they devoted surfing in the website and using its materials.

The effect of instruction time on students' performance was first investigated by Card and Kruger (1992). Using between-state variations, they find that students who grew up in states with longer terms had a higher return to education, but when using within-state variation the effects disappear. However, their identification strategy suffers from a serious potential for endogeneity as it might be that term length is correlated with other unobservable measures of school quality. Later studies have found similar results regarding the relationship of school year length and students' achievements (Wößmann 2003, Pischke 2007).

One approach to identify the effect of instruction time on students' tests scores was exploiting unscheduled school closing days due to extreme weather as a source of exogenous variation in instruction time. Using this approach Marcotte and Hemelt (2008) and Hansen (2011) find that more instructional time prior to test administration increases student performance and the effects are robust.

Another approach employed in two recent studies is using student panel data with multiple observations of both test scores and instruction time. This approach was first introduced by Lavy (2015), using within-student variation in PISA data. He finds that instructional time has a positive and significant effect on test scores and that the effect is much lower in developing countries. This first study was followed up by Rivkin and Schiman (2015) who employed the same method to a newer set of PISA data, finding similar results.

Previous research has also shown that there appears to be some heterogeneity between students on the extent to which they benefit from extra instruction time. In contrast to our results, where students in the upper part of the ability distribution gain more from an increase in the number of preparation days before the test than students in the lower parts of the distribution, Eren and Millimet (2008), using US data, find that students with the highest test scores benefit from a shorter school year, while students in the bottom half benefit from a longer school year.

In our study, we also investigate how the test order influence students' performances. To our knowledge, there are no previous contributions on the effects of test order.

Explaining gender differences

In our study, we find that there are significant gender differences in the effect of the number of preparation days and test order on exams' score. We find that girls benefit much more than boys in terms of final exam grades, having extra days to study before the exam. Also, while boys' grades deteriorate as their tests are taken later, girls gain better score in later exams. The psychological literature can give several possible explanations for these result; First, it is documented that both in college and in high-school women report greater use of strategies for coping and time management than men (e.g Misra and McKean, 2000; Xu, 2006 and 2007), and these strategies are positively correlated with better academic performance and grades (Claessens, Van Eerde, Rutte and Roe, 2007). There is also evidence that girls have more self-discipline (e.g., Duckworth and Seligman, 2006) than boys. Using cash incentive to high-school students in Israel to increase matriculation certification rate, Angrist and Lavy (2009) find that such an intervention leads to

substantial increase in girls' certification rate and in likelihood of college attendance, but had no effect on boys. They suggest that the female matriculation rates increase was partly because treated girls devoted extra time to exam preparation.

Another possible explanation is that boys have a higher discount rate that girls. People with high discount rate are likely not to work on a project with a deadline in the far future since they discount the future outcomes. Higher return to preparation days might be a signal to low discount rate and therefore a better use of the days far before the exam day. There is literature suggesting that adolescent girls are more likely to delay gratification (e.g., Silverman, 2003; and Duckworth and Seligman, 2006) than adolescent boys. Moreover, Warner and Pleeter (2001) find that among young adults, boys behaved as if they have higher discount rates than girls in the same group.

A third additional explanation can be driven by the gender differences in the reaction to pressure. The literature in psychology shows that increasing the pressure beyond a certain level can lead to a decline in performance, commonly referred to as "choking under pressure" (Baumeister 1984) and one of the mechanisms that produce "choking under pressure" can be personal traits such as competitiveness and ego-relevant traits like the belief that a task is diagnostic of an inherent characteristic (such as intelligence) that one cares about (Ariely et al. 2009). Though females have more effective time management behaviors than males, they experience higher academic stress and anxiety (Misra and McKean, 2000). Our results show that when students have severe time pressure of less than three preparation days, the effect of the number of days on boys' grade in higher than girls, and when they have more days, girls usually benefit more than boys. Azmat, Calsamiglia and Iriberri (2016) find similar heterogeneous results across gender as a result of increasing pressure. They exploit the variation in the stakes of tests, which range from 5% to 27% of the final grade and find that female students outperform male students in all tests—but to a relatively larger degree when the stakes are low. The gender gap disappears in the highest stakes tests which account for 50% of the university entry grade.

Time Pressure

The literature in psychology suggest that time pressure is detrimental for decision quality (Diederich 1997, Busemeyer and Diederich 2000, Diederich and Busemeyer 2003). The mechanism driving this effect is the worsening of reasoning processes and the tendency of individuals to ignore important information and rely on heuristics (Kruglanski and Freund 1983, Gigerenzer et al. 1999, Rieskamp and Hoffrage 2008). In spite of the importance that time pressure has in many economic decisions, only recently economists paid attention to this issue. Kocher and Sutter 2006 present evidence from a lab experiment using the beauty-contest game to study the effect of time pressure on quality of decision-making and on performance in time-dependent incentive schemes. They show that, in absence of a time-dependent incentive scheme, the depth of reasoning decreases under time pressure even in interactive contexts. Similarly, Sutter et al. 2003 examine the effects of time pressure on bargaining behavior in an ultimatum game, showing that it has high efficiency costs by leading to significantly higher rejection rates of offers, despite the effect vanishing with repetition. Kocher at al. (2003) and Bollard et al. (2007) find that time pressure changes individual attitudes toward risk. In addition, time pressure can change individual behavior by rising physiological stress, which in turn increases risk taking (Starcke et al., 2008; Putman et al., 2010; Buckert et al., 2004) and inhibits strategic thinking (Leder et al., 0 3). Even less is known on gender differences in response to time pressure. Some physiological studies (see Voyer, 0 0 for a review) show that in some cognitive tests (such as mental rotation tasks) gender differences in favor of men are significantly larger when the task is administered with time constraints compared to when such constraints are absent. Shurchkovy (forthcoming) shows that among factors that make women less effective than men in certain competitive environments (Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007) a crucial role is played by the ability to handle time pressure. In a laboratory experiment, she finds that gender inequality in performance is due to men and women reacting differently to time pressure: women perform significantly better than men in competitive verbal tasks without time constraints. In this paper, we provide new evidence both of a negative impact of time pressure on individual performance and

of gender differences in handling time pressure. Understanding whether males and females react differently to time pressure is relevant in trying to explain why women, even if as educated as men, continue to be heavily under-represented in many professions involving risky and high-pressure activities, such as executives, financial traders, entrepreneurs etc. An increasing literature documents how gender differences in preferences have a role in explaining gender differences in economic and social outcomes (Croson and Gneezy, 2009; Bertrand, 0 0; Niederle and Vesterlund, 2007).[3]

Adding to this literature, in our work we focus on gender differences in the ability to face time pressure: these differences, as those in attitudes towards competition, risk aversion, time and social preferences, already widely investigated by the economic literature, might help at explaining job sorting and labor market outcomes. However, compared to many of the existing works on gender differences in preferences (and all the few papers dealing with time pressure) that rely on laboratory experiments, we run a field experiment allowing us to observe individuals in a real life environment, in which they have strong incentives to perform well.

[2] For studies that argue that gender inequality persists due to the innate inability of women to compete, see also Baron-Cohen (2003), Lawrence (2006), and the citations in Barres (2006). Babcock and Laschever (2003) provide yet another possible explanation for gender di/erences in income and status that stem from the suppressed relative willingness and ability of women to engage in negotiations.

[3] Society generally perceives men to be better than women at following directions and reading maps, while women supposedly tend to follow landmarks when driving (Rahman et al. 2005). When it comes to solving mazes, men are thought to be overwhelmingly superior to women (Pease and Pease 2000, p. 107). Similarly,

[3] Bertrand and Hallock (2001) document this fact by gathering data on the ve highest-paid executives of a large group of U.S. rms over the period of 19921997, where they nd that only 2.5 percent of the executives in the sample are women. A similar under-representation of women is found among CEOs at Fortune 500 companies (CNNMoney 2006), tenured faculty at leading research institutions (MIT 1999), and top surgeons in New York City (New York Magazine 2010).

men are perceived to have higher math abilities than women, while women are perceived to have superior verbal skills. In particular, Pajares and Valiante (2001) note that differences in achievement of middle school students lie in the stereotyped beliefs about gender differences rather than gender itself. Girls report stronger motivation and condence in writing and receive higher grades in language arts. Boys report stronger performance-approach goals (Pajares and Valiante 2001).

Olga Shurchkovy "Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints" Forthcoming: The Journal of the European Economic Association

Abstract: Gender gaps in the work place are widespread. One explanation for gender inequality stems from the effects of the interaction between competition and two pressure sources, namely, task stereotypes and time constraints. This study uses a laboratory experiment to find that the gender gap in performance under competition and preferences for competition can be partly explained by the differential responses of men and women to the above pressures. In particular, while women underperform the men in a high pressure math-based tournament, women greatly increase their performance levels and their willingness to compete in a low-pressure verbal environment, such that they actually surpass the men. This effect appears largely due to the fact that extra time in a verbal competition improves the quality of womens work, reducing their mistake share. On the other hand, men use this extra time to increase only the quantity of work, which results in a greater relative number of mistakes. A labor market study suggests that the nature of the job and stress levels seem to be correlated with the gender gap in the labor market in a manner consistent with the results of my experiment.

Jensen JL, Berry DA, Kummer TA (2013) Investigating the Effects of Exam Length on Performance and Cognitive Fatigue. PLoS ONE 8(8): e70270. doi:10.1371/journal.pone.0070270

Department of Biology, Brigham Young University, Provo, Utah, United States of America

Abstract: This study examined the effects of exam length on student performance and cognitive fatigue in an undergraduate biology classroom. Exams tested higher order thinking skills. To test our hypothesis, we

administered standard- and extended length high-level exams to two populations of non-majors biology students. We gathered exam performance data between conditions as well as performance on the first and second half of exams within conditions. We showed that lengthier exams led to better performance on assessment items shared between conditions, possibly lending support to the spreading activation theory. It also led to greater performance on the final exam, lending support to the testing effect in creative problem solving. Lengthier exams did not result in lower performance due to fatiguing conditions, although students perceived subjective fatigue. Implications of these findings are discussed with respect to assessment practices.

Citation: Jensen JL, Berry DA, K

3. High-Stakes High School Examinations in Israel and in Other Countries

Since the Scholastic Aptitude Test's (SAT) first administration in 9 6, it has been taken by millions of test-takers and has been used to rank students applying for college in the United States, and similar tests are used around the globe. The great weight placed on such exams has the benefit of being a cost-effective way of comparing students across schools with a similar metric, but may also represent a noisy measure of student quality. Many factors can affect student performance that are unrelated to cognitive ability, including how a student slept the previous night, whether the testing room has a comfortable temperature, and potentially, exposure to ambient air pollution. In light of the great weight placed on test scores in admissions processes at many elite schools, it is worth knowing whether (a) these scores are sensitive to systematic factors that are derived from the structure of the exams and whether these factors have also distributional consequences by students' background. Since this would be an extremely challenging question to address in the US, where SAT score data is fiercely guarded and generally not available for matching to adult outcomes, the Israeli *Bagrut* represents a novel opportunity to examine this question.

The *Bagrut* exams take place over a number of days, and are predominantly administered at the

conclusion of the academic school year following 0th, th, and th grades.[4] The exams focus on seven mandatory subjects and one or more elective subjects, and are held at the student's high school without opportunity for rescheduling or changing the testing site. Since students take between 8- 5 separate exams, there is significant variation in number of free days that can be used for preparation across the same student's different tests, enabling us to estimate models with student fixed effects. Our design is also aided by the fact that retaking *Bagrut* exams is costly. Since most exams are given at the end of th grade, and Israelis begin a period of compulsory military service (3 years for boys and years for girls) after high-school graduation, retaking the exam is only possible for most students several years after the relevant coursework and would require many additional days of testing. A negative *Bagrut* outcome during a student's first attempt is likely to have a significant effect on a student's post-secondary academic options.

Passing the *Bagrut* exams awards a student a *Bagrut* (matriculation) certificate, which is a pre-requisite for study at universities and most academic and teachers' colleges.[5] Students are admitted to university programs on the basis of their average *Bagrut* scores and a separate psychometric examination. Each university ranks applicants according to the same formula, thus producing an index based on a weighted average of the student's average score on all his or her *Bagrut* exams and the psychometric examination. This ranking determines students' eligibility for university admission, and even which major they can choose within the university. Therefore, number of preparation days before each exam and the order of the exam can affect students' university schooling by affecting their probability of passing *Bagrut* exams, and also by affecting the average score on these exams. In summary, the mechanisms by which these two factors can affect long-term economic outcomes is through their effect on (a) the probability of pursuing higher education (b) affecting the type of higher education pursued and (c) the quality of higher-

[4] A small number of exams are taken near the end of the first term in January (less than % of our sample).
[5] The post-secondary education system in Israel consist of eight universities that grant PhDs (as well as other degrees), approximately 50 academic colleges which offer undergraduate degrees (of which a very limited subset which offer masters degrees), and a set of non-university institutions of higher education that confer teaching and vocational certificates. Practical engineering colleges run two-year programs awarding degrees (or certificates) in fields like electronics, computers, and industrial production. An additional two years of study in an engineering school is required in order to complete a BSc in engineering.

education institution ultimately attended.

b. Data

Our data set is generated by combining three primary data sources: Israeli test score data from 000-005, information on the exact dates of all exams, and students' demographic and socio-economic characteristics. The *Bagrut* exam information and demographic information for each test taker is provided by the Israeli Ministry of Education. These files also contain rich demographic information on the student and the student's family, such as parental education level, number of siblings, country of origin, and ethnicity. For each exam, we also know the date of the test.

The summary statistics for our sample are presented in Table in two panels; Panel A reports sample means of our exam-level data, and Panel B reports sample means of our student-level data. The sample is composed of 4 5, 9 examinations taken by 55,796 students at 6 6 schools throughout Israel between 000 and 00 . In columns ( ) and (3) we stratify the sample by sex, and in columns (4) and (5), we stratify by a measure of achievement known as the *Magen* score. The *Magen* score is calculated for each exam using the student's performance over the course of the school-year, and on an exam similar to the *Bagrut*, making its composite average over all exams taken by the student a natural candidate for stratifying the sample by student quality.[6] In Panel B, we report our student-level means, which includes demographic information on the student, the education of both parents, and the student's earnings in 0 0. The sample means also reveal several interesting patterns, including the higher achievement of girls: roughly 7 % of girls receive a matriculation certificate, compared to only 64% of boys.

[6] The date on which the *Magen* exam is given is usually up to few weeks before the *Bagrut* exam but the exact date is unavailable, precluding a direct analysis of these scores.

3. Empirical Strategy - Examination Performance and *Bagrut* Structure

In first section of our analysis, we examine the partial correlation between the two variables of interest, number of days between each two adjacent exams and the order number of an exam, and test scores in our sample of exam-level data. For identification, we rely on the panel structure of the data and the repeated nature of the *Bagrut* exam. Since we observe the exact location of the test, we can include city or school fixed effects. Since we observe the students taking multiple exams, we can include student fixed effects. Formally, the models we estimate are of the following form:

$$Y_{ist} = \beta D_{ist} + \delta O_{ist} + X_{it}\Pi + E_{it}\Theta + C_t + P_l + I_{i+}\varepsilon_{ist} \quad (1)$$

where Y_{ist} is the test score (z-score mean zero and standard deviation) of student i at school s at time t; D_{ist} is our measure of number of days between adjacent exams; O_{ist} is the order of the exam; X_{it} is a vector of observable individual characteristics possibly related to test outcomes, in which we include parental education in years, a dummy for sex, number of siblings and ethnic origin indicators; E_{it} is a vector of test characteristics in which we include the length of the exam (minutes), and an indicator for compulsory subjects, E_{it} is the school grade in the subject of the test, C_t and P_l are cohort and exam proficiency level fixed effects respectively; I_i is our fixed effect for the individual; and ε_{ist} is an idiosyncratic error term. Note that in specifications with individual fixed effects our individual-level controls are obviously dropped.

The key identifying assumption for inferring a causal relationship between number of days between exams (and test order) and test scores estimated by equation ( ), β and δ , is that unobserved determinants of student's test scores are uncorrelated with these two structural parameters of the exams schedule. Without the pupil fixed effects to absorb unobserved variation in individuals, this assumption is likely violated since the number of exams is, and therefore number of days between exams, is correlated with student ability. For example, lower ability students will have more time space between exams, therefore the OLS estimate of β will likely have the opposite than expected sign negative instead of positive. This negative selection may be eliminated by adding student's background controls that are included in X_{it} or by controlling for

student's school score E_{is} which is exam specific but as will be shown below these variables account only partially for the selection embodied in D_{is}.

3. Empirical Results

In Table 4 we present estimates of the two treatment of interest from five different specifications of equation (1). In column 1 the estimated regression includes school and cohort fixed effects and the test characteristics. Focusing first on the full sample estimates presented in panel i, the estimated effect of days is negative, as expected due to negative selection in the number of exams that a student has. Adding students' characteristics as controls (column 2) leaves the parameter estimate unchanged, -0.0016 (se=0.0004). Standard errors are clustered at the school level. Adding the school score, which is exam specific, reverse the sign of effect of days to positive and it is also statistically significant, 0.0031 (se=0.0003). Controlling for the school score apparently eliminates most of the selection bias since, as we see from the estimate presented in column 4, adding a pupil fixed effect as a control changes only marginally the estimate to 0.0037 (0.004), leaving the standard error unchanged.

The effect size of adding one potential day for preparation between any two exams is therefore not very large but since some exams have one day only between exams, changing this gap to the mean in the sample (7.3 days) will increase the test score by 0.02 sd which is not small relative to the effect size of costly educational interventions such as reducing class size or adding instructional time. In appendix Table A1 we report estimates where we allow for non-linearity in the effect of days by adding to equation (1) a square term of days. The effect of this second term is negative and significant, suggesting that the marginal effect of an additional free day for exam preparation declines as the number of days increases. The effect of number of days is larger in this specification, 0.008 and therefore the gain from spacing more generously exams will be larger on average for students who have more exams and fewer days on average in between any two exams.

The effect of the test order follows an opposite pattern to that of number of days. The estimate in Table 4, panel i, column 1, is positive, large and highly significant. Adding students' characteristics (column 2) leaves it unchanged and adding the school score halves it but it is still positive. Adding however the student's fixed effect reverses the sign to negative, -0.0027 (se=0.0014). This negative estimate indicates that test score declines with test order even though the initial estimate in column 1 was positive, likely due to positive selection of students with higher orders exams. The positive selection is eliminated only when we add the pupil fixed effects. The size of the estimates in column 4 implies a modest effect size. For example, the test score in the 5^{th} exam is, holding everything else constant, 0.014 sd lower than the test score in the first exam, and the test score of the last exam for a median student (7th exams) is lower by 0.019 sd relative to the score in the first exam.

In column 5 we report estimates from a regression with the same specification of column 4 but with an addition of allowing the effect of days (test order) to vary with the test order (days). The estimate of the interaction term between the two treatments is positive and significant, 0.004 (se=0.002), implying that potential preparation days between exams are more important downstream the exam order. Symmetrically, the negative effect of exam order declines when more days are available for preparation between exams.

The conclusion from estimates presented in panel i of Table 4 is that performance of students in the high school high stake exist exams varies meaningfully with number of potential preparation days and the test order. These efficiency effects might be very detrimental for students who apply to highly selective university programs where the average *bagrut* scores or scores in specific subjects such as in math and science subjects are critical for admission. Further, heterogeneity in these treatment effects by student's characteristics may amplifies the problem, carrying perhaps also distributional consequences that might have long term implications much beyond the short term average effect on *Bagrut* test scores. In the next section we examine heterogeneity effect by gender and by student's ability and obtain very different results across these groups, implying potential distributional consequences.

***Heterogeneity in Treatment Effects by Gender*.** I panel ii of Table 4 we present the results for boys and in panel iii the results for girls. The 'naïve' estimated effect of days among boys is negative, large and significant but it becomes positive and significant when pupil fixed effects are added. The pattern for girls is different, as it is positive even in the first specification in column 1, and it gets much larger and more precise when pupil fixed effects are added. The effect of days among girls is twice the effect on days among boys, 0.0046 versus 0.0023 and the difference between the two estimates is statistically significant. Girls benefit more than boys from available time for preparation before exams. In Table A1 we see that the quadratic term of days has a negative coefficient for both boys and girls, suggesting that the gain from an additional day of potential preparation is declining.

The effect of test order reflects positive selection both for boys and girls but more so for girls (column 4 of Table 4). This estimate declines for both gender with the addition of controls. For boys it turns negative, -0.0159, when adding the pupil fixed effect and for girls it remains positive, 0.008. Boys' performance deteriorates as they progress in the exams order while girls' performance improves. Naturally mental and physical fatigue can explain the deterioration in the performance of boys' and this mechanism may apply also for girls. However, it could be that gaining confidence and motivation during the exam period may offset any such fatigue among girls. The declining performance of boys with test order may also reflect lower effort and time of study as boys has a higher time preference[7] and at some point the desire for 'leisure activity' overtake long term considerations such as having a better *Bagrut* scores. More pronounced fatigue among boys may be also caused…..

In Table A2 in the online appendix we present estimates using the full sample and allowing interactions of the two treatment effect with an indicator for boys. This approach is an alternative to

[7] It is worth noting, however, that there is a literature suggesting that adolescent girls have more self-discipline (e.g., Angela L. Duckworth and Martin P. Seligman 2006) and are more likely to delay gratification (e.g., Irwin W. Silverman 2003) than adolescent boys. Among young adults, John T. Warner and Saul Pleeter (2001) find that male enlisted personnel behave as if they have higher discount rates than women in the same group.

estimating the effect separately based on stratified samples by gender. This regression yields very similar estimates for the two parameters of interest. The estimate of days for girls is 0.045 versus 0.046 in Table 4 and for boys it is 0.0025 versus 0.0023. The estimate of exam order for girls is 0.009 versus 0.008 in Table 4 and for boys it is -0.0085 versus -0.0059. In columns 1-3 we also report the main effect the boys' indicator which is positive and significant.

The estimates from the regression when we allow the effect of days to vary with test order and vice versa (column 5 of Table 4) are positive for both boys and girls but it is larger and more significant for girls. We will return to these differences when discussing results by gender within ability sub-samples. However, we can already conclude from these results that the difference in the effect of days and of test order by gender is significant and large and therefore it may explain a meaningful part of the average gender difference in overall *Bagrut* score. The gender difference in 12th grade *Bagrut* exams predicted average test score is 0.25 sd in favor of girls. Using the gender differences in the parameter estimates of days and exam order, and multiplying them respectively by the girls' mean of days and exam order yield a gender difference of 0.104, equivalent to 40 percent of the gender *Bagrut* test score gap.

***Heterogeneity in Treatment Effects by Ability*.** There are several alternative lagged measures of ability that we can use, including lagged test scores in *Bagrut* exams in 10th-11th grade, number credit units in the *Bagrut* program and number of subjects in the *Bagrut* program chosen during 10th grade. All three measures are highly correlated but we prefer to use the third measure because the first two are endogenous while the third is determined in the beginning of high school. We divide the sample into quintiles based on the number of subjects in the *Bagrut* program. We note again that the composition of the quintiles overlaps largely with the composition of quintiles obtained when using one of the other measures of ability. Given the heterogeneity we found by gender, we also present the results by gender within ability groups. We first stratify the sample by ability and then by gender within quintiles. We present in Table 5 the estimates for each quintile, for the full sample (columns 1-2, and by gender (columns 3-6), based on the fuller

specification that includes students fixed effects and secondly based on the specification that includesalso the interaction term between days and test order.

In the full sample the estimated effect of days is positive and significant and increasing with ability, in the 5th quintile it is *nine* times larger than in the 1st quintile. The effect of test order has an opposite trend by ability, positive in the first two quintiles, though not significantly different from zero, and declining and becoming negative in the 4th quintile and also significantly negative in the 5th ability quintile. However, these patterns that we obtain from the full sample mask sharp differences by gender in the effect of days and test order.

The effect of days among boys is practically zero in the first quintile of ability and it increases monotonically with ability, at the 5th quintile it is 0.0087 (se=0.0009). The same pattern is observed in the sample of girls though the effect is positive and significant throughout from the 1st quintile and it is larger than that of boys in each quintile. The difference in the effect of days is smallest in the 5th quintile, 0.010 for girls and 0.0087 for boys but the gap is significantly different from zero because the two estimates have very small standard errors. This pattern implies that low ability boys (the lowest 40 percent in the ability distribution) do not benefit at all from potential preparation days before exams while girls in the same ability range do. These results suggest that number of preparation days before exams contributes to girls' test score advantage among students below median ability but not at the upper end of the ability distribution.

The effect of test order is negative throughout the ability distribution of boys. The parameter estimate is of similar size from the 1st to 4th quintile of ability and it jumps -0.021in the 5th quintile. Among girls it is positive and highest in the 1st quintile; it declines gradually but remain positive until the 4th quintile and in the 5th quintile it turns negative (-0.0053, se=0.0019); however, it is much smaller than the respective estimate among boys in this upper ability quintile. The opposite signs by gender of the estimated effect of test order clearly imply that this effect explains to some extent why girls have a test score advantage in the average *Bagrut* score. Given that the number of exams of boys and girls in each quintile is very similar, the

main factor is the gender difference in the estimated effect of test order. The size of this parameter suggest that its contribution to the gender gap is much larger than that of the number of preparation days.

The decline in the estimated effect of test order as we move up in the ability ladder is interesting. We noted above that this effect is most likely a combination of an increase in mental and physical fatigue and compensating gains in confidence and in experience and skills in tests performance, which, for example, can come from improved management of time, writing skills and performance under pressure. Perhaps the former effect might increase with ability because the sharp increase in level and scope of the material covered, while the later effect is larger for low ability female students because they have much more room for improvement in test taking skills. A third relevant mechanism is the effect of the number of remaining tests which declines as we move up in the test order and along with it the need to study for multiple exams during the time spell between each two exams. It could also be that girls are better in multitasking (here it means preparing simultaneously for multiple exams).[8] To shed some lights on these potential explanations, we present in Tables A5-A6 estimates from a sample that includes on the first 5 exams of every student. We expect that the estimated effect of number of days will be similar to that obtained from the full sample but that the estimated effect of test order will be different depending which of the above explanations is more relevant during the first five exams. In Table A5 the estimates of days are identical to those in Table 4. The effect of test order for boys is only marginally higher than what is expected due to the lower values of this variable (a mean of 2.5 versus a mean of almost 4 in the full sample). The effect of test order on girls is negative, but still lower than the effect on boys, suggesting that

[8] Recent lab experiments have shown that women find it easier than men to multitask and switch between tasks, set priorities, and adapt to changing conditions. For example, Kuptsova et al 2015 found in a lab experiment that men require more brainpower than women when multitasking and that women find it easier than men to switch attention and their brains do not need to mobilize extra resources in doing so, as opposed to male brains.

until the fifth exam, girls also are mainly affected by fatigue factors and the offsetting effects of gaining confidence and exams skills are evident only in later exams.

***Potential Non-Linearity of Treatment Effects by Gender and Ability*.** We focus in this section on non-linear estimation of the treatment effects in the two highest ability groups because the two effects of interest seem to be highest in these two sub-samples. In Table 8 we model non-parametrically the non-linearity of the effect of days as series of dummy indicators for the number of days. We keep in this model the effect of test order as a continuous linear effect. Columns 1-3 present the estimate for the 4th ability quintile and columns 4-6 present the estimates for the 5th ability quintile. First we note that the estimated effects of test order in all 6 columns are almost identical to the respective estimates in Table 6, an indication that the treatment variables are practically uncorrelated and therefore the estimates of effect of test order are not sensitive to how we model the effect of days.

Regarding the pattern of the effect of days, it is seen from columns 3 and 6 that the effect of days among girls is constant in the range of up to 5 days and then it increases sharply to a higher level where it is also constant in the range of 6- 5 days with another smaller increase for 5+ days. Among boys the pattern is similar except that there is a decline in the effect from day (4th quintile) or from day (5th quintile) until day 6 where the pattern in the increase in the effect resemble that among girls. In Table 9 we allow also for the effect of test order to be non-linear allowing for a separate effect for each order of test. For estimation efficiency and based on the similarity in the estimates presented in Table 8, we pool the 4th and 5th quantile samples. The effect of days on boys and girls are very similar to the average of the estimates presented in Table 9, again not a surprise given the low or zero correlation between the days and test order variables. The test order estimates reveal clearly the striking different pattern between boys and girls in the effect of this variable: for boys the effect is negative from start (3rd versus 2ed test) and then increases with a small gradient until the 6th test and then declines from the 10th test. The precision of the estimates declines with the order of the test, most likely because there are fewer students as one moves up with the test order. The

pattern among girls is strikingly different, until the 6th test the effect is negative relative to the 2 ed test but from the 7th test the estimate turns positive and it increases monotonically to the 3th test.

4. Robustness and Sensitivity of the Treatment Effect Estimates

The *Bagrut* program varies across schools, especially in the elective subjects offered. While all schools follow the same national curriculum with respect to the compulsory subjects in the *Bagrut* program, some schools offer study tracks in humanities, social science, and sciences, while other schools offer also or only technical tracks for electives subjects (such as electronics, optometry and so on). In average in our sample, the number of compulsory *Bagrut* exams is 5 and the average of elective subjects and is 2.5. Boys have 4.8 compulsory and 2.6 electives and girls 5.5 compulsory and 2.5 electives.

In Table 8 we present estimates based on a sample that includes only compulsory subjects. All students in this sample are tested in exactly the same subjects, therefore the maximum number of exams and the highest test order are smaller than in the full sample. Table 8 follows the same format as Table 4. The changes in sign and size of the estimates in each row across columns are very similar to the pattern observed in Table 4. Overall, the estimates in Table 8 are close to those in Table 4, especially in the sub-sample by gender. The effect of number of days on boys in Table 8 is 0.0018 (se=0.005) versus 0.0023 (se=0.0004). The effect of test order on boys in Table 8 is -0.0107 (se=0.0024) versus -0.0159 (se=0.0015) in Table 4. The effect of number of days on girls in Table 8 is 0.0050 (se=0.0006) versus 0.0046 (se=0.0005) in Table 4. The effect of test order on girls in Table 8 is 0.0 05 (se=0.0205) versus 0.0082 (se=0.0016) in Table 4.

Table 9 present the estimated effect by quintile of ability based on the compulsory subjects tests and it has the same format as Table 6. The estimates presented in Table 9 are of similar order of magnitude and same pattern over quintile as those presented in Table 6 but there are still some interesting differences. The effect of days among the lowest ability boys is still zero and it becomes positive and significant only

in the 3rd quintile. The effect of days among girls is positive and increasing throughout the ability distribution and it is almost twice the effect on boys in the 5th quintile. The effect of test order is negative and increasing with boys ability and in the 5th quintile it is -0.0167 (0.0036) versus -0.0 0 (se=0.00) in Table 6. The effect of test order is positive and declining with ability in the girls' sample but it does not turn negative in the 5th quintile as it does in the sample that includes both compulsory and elective subjects. In Table 0 it is 0.00 8 (se=0.00 7) versus -0.0053 (0.00 9) in Table 6.

The conclusion based on the results from the sample of compulsory subjects is similar to the one we drew from the full sample of all subjects: girls benefit more from days between exams and they are not negatively affect by the length of the exam season, though these two advantages vary by ability of girls, the benefit from days of preparation is particularly helpful for girls above median ability and the increase in performance with test order is particularly evident for girls below the median ability.

Results by regular and vocational schooling

To Be Completed

6. References

Ackerman PL, Kanfer R (2009). "Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions." J Exp Psychol: Appl 15: 163–181.

Ackerman PL, Kanfer R (2006). "Test length and cognitive fatigue". Final report to the College Board. Atlanta, GA: Author.

Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar (2009). "Large Stakes and Big Mistakes." Review of Economic Studies, 76, 45 469.

Babcock, Linda, and Sarah Laschever (2003). Women Dont Ask: Negotiation and the Gender Divide. Princeton University Press.

Barres, Ben (2006). Does Gender Matter? Nature, 44 (3), 33- 36.

Baron-Cohen, Simon (003). The Essential Di*ff*erence. Men, Women, and the Extreme Male Brain. Allen Lane, p. 56.

Becker, Gary (957). The Economics of Discrimination. The University of Chicago Press.

Bertrand, Marianne, and Kevin Hallock, Kevin (00). The Gender Gap in Top Corporate Jobs. Industrial and Labor Relations Review, 55( ), 3- .

Rahman, Qazi, Davinia Andersson, and Ernest Govier (2005). "A Specific Sexual Orientation-Related Difference in Navigation Strategy@. Behavioral Neuroscience, 9( ), 3 -3 6.

Black, Sandra, and Philip Strahan (00). The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry. American Economic Review, 9 (4), 8 4-83 .

Bollard, A., Liu, R., Nursimulu, A.D., Rangel, A., Bossaerts, P. 007. Neurophysiological evidence on perception of reward and risk: Implications for trading under time pressure. Working paper, CalTech, Pasadena, CA

Booth, Alison, and Patrick Nolen (2009). Gender Differences in Risk Behaviour: Does Nurture Matter? IZA Discussion Papers No. 40 6.

Buckert, M., Schwieren, C., Kudielka, B., and Fiebach, C. J. 0 4. Acute stress affects risk taking but not ambiguity aversion. Front.

CNNMoney.com (2006). Women CEOs for FORTUNE 500 companies. http://money.cnn.com/magazines/fortune/fortune500/womenceos/

Dreber, Anna, and Magnus Johannesson (2008). Gender Di*ff*erences in Deception. Economics Letters, 99, 97-99.

Duckworth, Angela Lee, and Martin P. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." Journal of Educational psychology, 98(1): 198–208.

Kocher, M. G., & Sutter, M. 006. Time is money–Time pressure, incentives, and the quality of decision-making. Journal of Economic Behavior & Organization, 6 (3), 375-39

Kocher, M. G., Pahlke, J. & Trautmann, S. T. 0 3. Tempus Fugit: Time Pressure in Risky Decisions, Management Science, 59(0), 380- 39

Lawrence, Peter (2006). Men, Women, and the Ghosts in Science.PLOS Biology, 4( ), 3- 5.

Leder, J., Häusser, J. A., & Mojzisch, A. 0 3. Stress and strategic decision-making in the beauty contest game. Psychoneuroendocrinology, 38(9), 503- 5

Macpherson, David, and Barry Hirsch (1995). Wages and Gender Composition: Why Do Women Job Pay Less? Journal of Labor Economics, 3(3), 4 6-47 .

Massachusetts Institute of Technology (999). A Study on the Status of Women Faculty in Science at MIT.

New York Magazine (0 0). The Top , 9 Physicians. http://nymag.com/bestdoctors/

Neurosci. 8:8 Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. Quarterly Journal of Economics. 8, 049– 074

Niederle, Muriel, and Lise Vesterlund (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? Quarterly Journal of Economics, (3), 067- 0 .

Niederle, Muriel, Carmit Segal, and Lise Vesterlund (008). How Costly is Diversity? A¢ rmative Action in Light of Gender Differences in Competitiveness. NBER Working Paper No. 39 3.

Pajares, Frank, and Giovanni Valiante (00). Gender Differences in Writing Motivation and Achievement of Middle School Students: A Function of Gender Orientation? Contemporary Educational Psychology, 6(3), 366-38 .

Paserman, M. Daniele (007). Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players. CEPR Discussion Paper No. 6335.

Pease, Allan, and Barbara Pease (000). Why Men Do Not Listen and Women Cannot Read Maps: How Were Di/erent and What to Do About It. Welcome Rain.

Polachek, Solomon(1998). Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure. Review of Economics and Statistics, 63( ), 60-69.

Putman, P., Antypa, N., Crysovergi, P., and van der Does, W. A. 0 0. Exogenous cortisol acutely influences motivated decision making in healthy young men. Psychopharmacology 08, 57– 63

Rolfe, John, and Peter Troob (2000). Monkey Business: Swinging Through the Wall Street Jungle. Warner Books.

Silverman, Irwin W. 2003. "Gender Differences in the Delay of Gratification: A Meta-Analysis." sex Roles, 49(9–10): 451–63.

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

Shurchkov, Olga, "Gender Differences in Output Quality and Quantity under Competition and Time Constraints: Evidence from a Pilot Study" (Dec. 2009), Fondazione Eni Enrico Mattei Working Papers. Working Paper 356. http://www.bepress.com/feem/paper356.

Spencer, Steven, Claude Steele, and Diane Quinn (1999). Stereotype Threat and Womens Math Performance. Journal of Experimental Social Psychology, 35( ), 4 8.

Steele, Claude, and Joshua Aronson (1995). Stereotype Threat and the Intellectual Test Performance of African-Americans. Journal of Personality and Social Psychology, 69, 797-8 .

Starcke, K., Wolf, O. T., Markowitsch, H. J., and Brand, M. 008. Anticipatory stress influences decision making under explicit risk conditions.

Behav. Neurosci. , 35 – 360 Sutter, M., Kocher, M. & Strauß, S. 003. Bargaining under time pressure in an experimental ultimatum game. Economics Letters, 8 (3), 34 -347

Voyer, D. 0 . Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. Psychonomic bulletin & review, 8( ), 67- 77

Warner, John T., and Saul Pleeter. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." American Economic Review, 91(1): 33–53.

Angrist, J., & Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. The American Economic Review, 99(4), 1384-1414.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. The Review of Economic Studies, 76(2), 451-469.

Azmat, G., Calsamiglia, C., & Iriberri, N. (2016). Gender differences in response to big stakes. Journal of the European Economic Association, 14(6), 1372-1400.

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. Journal of personality and social psychology, 46(3), 610.

Bloom, B. S. (1974). Time and learning. American psychologist, 29(9), 682.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. Journal of political Economy, 100(1), 1-40.

Claessens, B. J., Van Eerde, W., Rutte, C. G., & Roe, R. A. (2007). A review of the time management literature. Personnel review, 36(2), 255-276.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of educational psychology, 98(1), 198.

Eren, O., & Millimet, D. L. (2008). Time to learn? The organizational structure of schools and student achievement. In The economics of education and training (pp. 47-78). Physica-Verlag HD.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. American psychologist, 49(8), 725.

Hansen, B. (2011). School year length and student performance: Quasi-experimental evidence.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. The Economic Journal, 125(588), F397-F424.

Marcotte, D. E., & Hemelt, S. W. (2008). Unscheduled school closings and student performance. Education, 3(3), 316-338.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological review, 102(3), 419.

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. American Journal of Health Studies, 16(1), 41.

Pischke, J. S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. The Economic Journal, 117(523), 1216-1242.

Rivkin, S. G., & Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. The Economic Journal, 125 (588), F425–F448.

OECD, The ABC of Gender Equality in Education Aptitude, Behaviour, Confidence

Silverman, I. W. (2003). Gender differences in delay of gratification: A meta-analysis. Sex roles, 49(9-10), 451-463.

Simon, H. A. (1990). Invariants of human behavior. Annual review of psychology, 41(1), 1-20.

Taraban, R., Rynearson, K., & Stalcup, K. A. (2001). Time as a variable in learning on the World-Wide Web. Behavior Research Methods, Instruments, & Computers, 33(2), 217-225.

Warner, J. T., & Pleeter, S. (2001). The personal discount rate: Evidence from military downsizing programs. American Economic Review, 33-53.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. Oxford bulletin of economics and statistics, 65(2), 117-170.

Xu, J. (2006). Gender and homework management reported by high school students. Educational Psychology, 26(1), 73-91.

Xu, J. (2007). Middle-School Homework Management: More than just gender and family involvement. Educational Psychology, 27(2), 173-189.