

C A G E

**Gifted children
programs' short and
long-term impact:
Higher Education,
earnings, and the
knowledge economy**

CAGE working paper no. 609

March 2022

Victor Lavy

Gifted Children Programs' Short and Long-Term Impact: Higher Education, Earnings, and the Knowledge-Economy*

Victor Lavy
University of Warwick, Hebrew University, and NBER

Yoav Goldstein
Tel Aviv University

13_February_2021

Abstract

We estimate the short-run and longer-term effects of gifted children programs (GCP) in high schools in Israel. The program tracks the most talented students into gifted children classes, starting 10th grade. They receive more resources in smaller classes, a unique curriculum, access to high-quality teachers, and courses in universities. We use test scores in exams that measure cognitive achievements or intelligence and ability, measured in different ages, to select a comparison group of equally gifted students from other cities where GCP was not offered at the time. Based on administrative data, we follow 22 cohorts of GCP participants who graduated high school in 1992-2013. We measure treatment effects on outcomes, ranging from high school to the labor market in their 30s and 40s. Remarkably, the results we obtain do not vary when using alternative measures of ability or in the age, they are assessed. The evidence on the impact of GCP on academic achievements in high school is mixed, some compulsory subjects are affected negatively, and fewer are affected positively. However, these estimates are very small, implying a tiny effect size. These results stand in contrast to the abundance of educational resources enjoyed by GCP participants, in addition to better peers in terms of SES background and outcomes. We discuss in this context the objective of the program to widen the scope and area of interest of its participants beyond the regular curriculum. We also highlight the potential adverse effect of the Big-Fish-Little-Pond Effect. In the longer run, we find meaningful positive effects of GCP on higher education attainment. All gifted children achieve a BA degree, but a much higher share of GCP participants graduate with a double major. The effect of getting a Ph.D. is also positive, driven by more Ph.D. degrees in Elite Universities. GCP participants study more math, computer, and physical sciences but engage less in engineering programs. The net effect on STEM degrees is, therefore, zero. However, a much higher share of GCP participants graduated with two STEM majors. This evidence, along with the significant effect on a double major, suggests that GCP enhances the impact of “multipotentiality,” which characterizes many gifted adolescents. We find no effect of GCP on employment and earnings. Nor do we find that they work more than other equally talented children in the various sectors of the knowledge economy: hi-tech manufacturing, hi-tech services, and academic institutions. Finally, we examine marriage and family formation patterns as mediating effects and find no discerned GCP effects.

In the short-term, medium-run, and into adulthood, these comprehensive sets of results are not qualitatively different for females and males gifted children who participated in GCP. Treatment heterogeneity by giftedness level allows us to compare our results to earlier studies that used regression discontinuity designs to identify GCP effects on only marginally eligible students for such programs. We find meaningful differences in treatment effect between marginal and inframarginal gifted children, suggesting that it is essential to examine GCP's impact over the whole spectrum of Giftedness. Importantly, we find that GCP similarly affects low and high-SES students. Half of the students among the six youngest cohorts in our sample started the program in middle school, while the others did that in high school. We find no differences in GCP effect on high school and university outcomes by the length of the program.

* We thank the Central Bureau of Statistics for providing access to data we use in this study at its protected research room in Jerusalem. We also thank Dr Anat Ben-Simon, general director of The National Institute for Testing, for helpful guidance and information about the University Psychometric Entrance Test. We thank James Fensky, Emmch Duchini, and participants in seminars at Ben Gurion University, Bocconi University, CEMFI Madrid, Hebrew University, Pompeu Fabra, University of Warwick, and the CESifo Education Conference in Munich for useful comments and suggestions. Lavy acknowledges financial support from the Israel Science Foundation, from the Falk Research Institute and from CAGE.

1. Introduction

Gifted children receive special attention in many educational systems. However, despite the considerable amount of resources and time invested in this group, the evidence on the short-term effectiveness of programs targeted at it is still limited. Second, we know little about these interventions' long-term effects. Third, full-scale programs for gifted children often only begin at high school, and we do not know whether this group would also benefit from earlier interventions. Fourth, while policy circles express growing concerns regarding the persistence of gender gaps in education and labor market outcomes, there is scarce evidence about the role programs targeted at higher ability and gifted students influence these trends. This paper provides evidence on these issues by exploiting a long-existing gifted children's education program and unique administrative data that permits following children for over two decades.

We estimate the short-run and longer-term effects of gifted children programs (GCP) that started in high or middle schools in Israel. The program tracks the most talented students into gifted children classes, starting 10th grade (or 7th grade). As a result, they receive more resources, a unique and accelerated curriculum, access to high-quality teachers, and attend university courses. Based on administrative data, we follow 22 cohorts of GCP participants who graduated high school in 1992-2013. We select comparison groups from other cities where GCP was not offered at the time, based on different ability measures assessed for different cohorts at various grades. The results are very similar across ability measures and samples.

We measure treatment effects on outcomes, ranging from high school to the labor market in their 30s and 40s. We show that gifted children's academic achievements in high school, particularly in university-preparation courses, are not significantly affected by GCP participation. However, we find a small negative effect on the average composite score and math, driven mainly by adverse effects on boys.

In the long-term, we find no effects of GCP on the rate students gain B.A. degrees, as almost all (98 percent) treated and comparison gifted children achieve this degree. Still, we find a positive and relatively large impact on graduating with a double major and gaining Ph.D. degrees (driven by an increase in Ph.D. degrees from Elite Universities). We also find that GCP affect the choice of field of study, increasing academic degrees in math, computer science, and physical sciences and sharply reducing degrees in engineering programs.

GCP participants and the comparison group have similar earnings and employment rates in the knowledge economy sectors. The lack of effects on labor market outcomes persists until a late age, as seen among the oldest cohorts in our sample. We also examine marriage and family formation patterns as mediating effects and find no discerned GCP effects. However, we find that the program positively affects the "quality" of the spouse, driven by marriages of GCP participants with their

classmates. In the short-term, medium-run, and into adulthood, these comprehensive sets of results are not different for females and males gifted children who participated in GCP.

Estimating treatment effect heterogeneity by giftedness level allows us to compare our results to earlier studies that used regression discontinuity designs to identify GCP effects on only marginally eligible students for such programs. Our findings show that treatment effect differences between marginal and inframarginal gifted children can be meaningful.

Half of the GCP participants of the younger cohorts in our sample (high school graduates in 2006-2013) participated in a GCP since middle school, and the rest have participated in a GCP only during high school. This setting allows us to address an important dimension of the GCP treatment effects heterogeneity, namely whether participation in a GCP for a more extended period (7th-12th grades) affects student outcomes differently from shorter participation (10th-12th grades). Interestingly, we find no evidence for heterogeneity in this dimension, suggesting that GCP effects do not vary by the length of participation in the program. However, this evidence is limited to high school outcomes and post-secondary attainment since the sample we use in the analysis is relatively young.

Lastly, we estimated heterogeneous treatment effects by socioeconomic status (SES). The minor adverse effects on high-school test scores are more pronounced among higher SES participants, while the longer-term effects are very similar. Our results of almost no variation in treatment effect by students' background stand in contrast to the abundant literature that shows that the effect of many educational programs varies by students' socio-economic status. But our evidence is consistent with another important finding, that background characteristics of gifted children make little difference to their short-term and life-long outcomes. In other words, and in sharp contrast to the pattern among non-gifted children, there is no socioeconomic gradient in outcomes among gifted children.

Our results are highly robust to various ability measures we use to select other gifted children as a comparison group. We show that using different measures of intelligence and ability to select a comparison group yields almost identical results. Furthermore, having these measures assessed at 5th, 8th, or 10th-11th grade leave the estimated GCP effect unchanged.

The evidence we present in this paper contributes to the small number of recent studies on the causal effect of gifted children's programs (GCP) on student performance. Card and Giuliano (2014) apply a fuzzy regression discontinuity (RD) design to estimate a GCP's impact on math, reading, and writing test scores. The GCP they investigate puts gifted students together in classrooms with pupils who performed well in previous grades or high achievers and offered an enriched curriculum. This study finds little, if any, test score gains for gifted students but significant and positive improvements for their high-achieving peers. Bui et al. (2014) examine the effect of GCP on math, reading, and language test scores of middle-school students in the South-western U.S. Using either a fuzzy RD design comparing students scoring just above or below the GCP admission cut-off or exploiting a lottery in oversubscribed middle schools offering the GCP program, the authors find no significant positive effect on student performance. In a comparable experiment, Davis et al. (2013)

exploit a fuzzy RD design to estimate whether GCP can help public schools retain gifted students. Close to the admission cut-off, they find that students considered gifted are more likely to stay in public schools. Bhatt (2012) also looks at U.S. middle school students but uses an instrumental variable strategy that exploits differences in GCP admission rules between schools. She finds positive test score gains, but this may partly reflect the sorting of students to schools. Booij, Haan, and Plug (2016) examine the effect of a gifted and talented secondary education program in the form of an individualized pull-out program. Like earlier studies, they also use a fuzzy RD design to estimate the impact on those at the program's margin of acceptance. They show that participants obtain higher grades, follow a more science-intensive curriculum (most notably for girls), and report stronger beliefs about their academic abilities. They also find that these positive effects persist in university, where students choose more challenging fields of study with, on average, higher returns.

This paper makes several significant contributions to this literature. First, it is based on an experienced gifted children program, running for over three decades in separate gifted classes or schools. We present the first and most comprehensive outcomes in the medium and long run. Beyond completion of university degrees up to Ph.D., we examine the choice of field studies, especially in STEM. It takes a longer-run perspective to assess whether program effects persist. Therefore, we follow GCP participants over age 40 and examine their labor market and personal outcomes, including earnings and family formation. Uniquely, we examine GCP participants' contribution to the knowledge economy by looking at their integration in hi-tech and academic institutions. Secondly, an essential contribution in this paper is the analysis of treatment heterogeneity by giftedness level. This distinguishes this paper from earlier studies that used regression discontinuity designs to identify GCP effects on only marginally eligible students for such programs. Another significant contribution is our distinction in estimating treatment effect by the length of GCP.

We relate some of our findings to theories and hypotheses in the literature in psychology about gifted children. It includes the literature regarding the affective and personality development of the socio-emotional characteristics of gifted children. The literature on 'big fish small pond' is perhaps key in understanding our finding of no effect of GCP on test scores in high school (see Marsh et al. (2008) for a review). Of particular relevance to us are studies that study the effect of labeling (being part of a gifted program) and excessive parental expectations and pressure from teachers and social networks (E.g., Robinson et al. 2002, Pfeiffer 2003). Related literature coined the term 'the gifted paradox.' Gifted children have an ability that can be used for a meaningful process of self-exploration to form identity. Still, external pressures curtail this process and lead them to choose, for example, prestigious professions. This tends to hasten the process of identity formation and limit self-exploration. Our findings that GCP causally directs gifted adolescents to math and computer and physical sciences programs at the university are likely related to this paradox. This paradox is related to the term 'multipotentiality', which characterizes gifted children in GCP (Leung et al. 1994, Kerr

and Colngelo, 1988). The effect on double majors is possibly a manifestation of this characteristic of many gifted adolescents enhanced in a prestigious tracking program.

The paper is organized as follows. The following section describes the gifted education programs in Israel and elsewhere. Section 3 presents the data, and section 4 the empirical methodology. Next, we present the results in two sections, first the effect of GCP on high school and post-secondary schooling, including the field of study (section 5). We then present the impact on longer-term outcomes, such as advanced degrees attainment and labor market outcomes (employment and earnings), followed by evidence of the effects on career choices in knowledge-producing sectors and marriage and parenting (section 6). Finally, section 7 provides conclusions and further discussion.

2. Context and Background

2.1 Gifted Children Programs

In most countries, fostering gifted students' talent is essential in the knowledge economy crucial for securing new generations of scientists, creators, and innovators. Yet, how to deliver gifted education is at the center of a longstanding and still hotly debated topic in education policy circles. In many countries, introducing specific practices for talented children dates back to the 1960s (Boettger and Reid 2005, Vrignaud and Bonora, 2005, Monks and Pfluger, 2005). Over time, these included various interventions targeted at different age groups, from early enrolment in primary school to grade skipping, curriculum enrichment, extracurricular syllabus, and summer camps. Other countries and even different school districts within the same country also adopt different selection procedures. Early GCP used intelligence assessment (e.g., I.Q. scores) as the basis for eligibility. Still, this selection method has been strongly criticized as I.Q. Tests are argued to be ethnically or racially biased. As an alternative, researchers and practitioners have suggested that eligibility should be based on a combination of cognitive and non-cognitive measures.

Remarkably, despite this longstanding debate, there is little causal evidence on gifted education programs' relative effectiveness for different targeted groups. According to the most recent literature review (Bhatt 2011), most empirical work on this topic finds positive associations between program exposure and children's achievement. Yet, as this evidence is mainly correlational, it is unclear whether it measures these programs' true causal impact or captures selection effects. It is also unclear what is the best age to start these programs. And little is known about the mechanisms through which these interventions can benefit gifted children. For instance, these programs might help talented students by grouping them with other high-achievers or offering additional resources, including specially trained teachers or an advanced curriculum.

On the other hand, the effect of moving from an environment of ‘big fish in a small pond’ to a context of being a ‘big fish in a big pond’ may cause anxiety and decline in self-concept (defined¹ as how an individual perceives his behavior, abilities, and unique characteristics). The label of ‘gifted’ that is strongly enhanced once part of a GCP can increase pressure from family and social circles, leading to anxiety and a decline in achievements. These channels are hardly considered in the literature, even though distinguishing between them is vital given their different policy implications. Equally important, there is no evidence of the long-term effects of these programs.

2.2 Gifted Children Education in Israel

By the late 1980s, Israel had developed a separate study program for highly gifted learners throughout Grades 3-12.² This program incorporated elements of enrichment, extension, and acceleration. In parallel, some universities started to offer education and training to teachers of gifted children. By 1994, the Ministry’s Department for Gifted Education had acquired an extensive list of responsibilities, including testing children in some cities, establishing unique enrichment frameworks, or instructing teachers and field-workers. Since then, three types of GCP have been offered: (1) A weekly program, organized by a city or school district, often starting in third grade and continuing until the end of primary school (6th grade), including weekly enrichment days in pull-out sessions. The Jerusalem school district Offek is a well-known example of such a program. (2) Special classes in one of the regular city schools enable gifted learners to be taught in separate classes throughout lower (grade 7-9) and upper secondary levels (grade 10-12). The learning content is based on the standard school curriculum. Still, it incorporates advanced concepts and topics, various teaching methods, and joint teaching with university staff. (3) An afternoon enrichment program.

This paper focuses on upper secondary gifted children programs (type 2 above) because they are numerous, offer a meaningful sample size for analysis, and resemble many of the GCP in Europe and the U.S., offering more external validity to this paper’s findings. Admission to these programs is based on an intelligence test undertaken during the year preceding the program. We focus on the gifted classes for students in their 10th to 12th grades from 1992-2013. During the 1990s, there were gifted classes in 11 high schools in 10 localities in Israel, most in the major cities. In later years, GCP programs were offered for middle-school students, too. We use this variation in our sample to estimate how the effects of a GCP vary by the age students started the program, namely length of exposure.

¹ Self-concept tends to be more malleable at a younger age and is still going through self-discovery and identity formation (Bailey, 2003). Research in human psychology suggests that the self-concept is made up of three different parts. **Ideal self**: the person you want to be. **Self-image**: How an individual sees himself, including physical characteristics, personality traits, and social roles. **Self-esteem**: how much an individual likes, accept, or values himself, which can be impacted by several factors, including how others see him, how he thinks he compares to others, and his role in society.

² The material presented in this section draws details from <https://giftedphoenix.wordpress.com/2012/11/15/gifted-education-in-israel-part-one/> (retrieved on 06-09-2021).

Finally, a 2004 reform consolidated the country's GCP into a national program to develop Israeli gifted education. It embraced the two-morning frameworks – weekly enrichment days and special classes in schools. In addition, it added a residential school for the gifted that serves mainly children from all over the country. As a result, the number of special classes operating in secondary schools has expanded to over 20. In our analysis, we focus on the eleven older programs.

2.3 Israel's High School and Higher Education Systems

When entering high school (10th grade), students enroll in the academic or non-academic track. Students enrolled in the academic track receive a matriculation certificate (*bagrut*) if they pass a series of national exams in core and elective subjects taken between 10th and 12th grade. Students choose to be tested at various proficiency levels, with each test awarding one to five credit units per subject, depending on difficulty. Advanced level subjects award students more credit units (5 relative to 4 for an intermediate level and 3 for a basic level); a minimum of 20 credit units must qualify for a *bagrut* certificate. About 52% of all high school seniors received a *bagrut* in the 1999 and 2000 cohorts (Israel Ministry of Education, 2001).³ The *bagrut* is a prerequisite for university admission, and receiving it is an economically important educational milestone. For more details on the Israeli high school system, see Lavy (2020, 2021).

Israel's post-high school academic schooling system includes ten universities (one of which confers only graduate and Ph.D. degrees) and 50 colleges that confer undergraduate degrees (some also give master's degrees).⁴ All universities require a *bagrut* diploma for enrolment. Most academic colleges also require a *bagrut*, though some look at specific *bagrut* diploma components without requiring full certification. It is typically more difficult for a given field of study to be admitted to a university than college. The national university enrolment rates for the cohort of graduating seniors in 1995 (through 2003) was 27.6 percent, and the rate for academic colleges was 8.5 percent.⁵

3. Data

We use several panel datasets from Israel's Central Bureau of Statics (CBS). CBS allows restricted access to this data in their protected research lab. The underlying data sources include the following. The population registry data consists of a fictitious individual national I.D. number that appears in all the data sets described below and enables matching and merging the files at the personal level. It also contains information on marital status, number of children, and birth year. In addition, administrative

³ Bagrut rates are much higher among gifted children, almost 100% among the gifted students in our sample.

⁴ A 1991 reform sharply increased the supply of postsecondary schooling in Israel by creating publicly funded regional and professional colleges.

⁵ These data are from the Israel Central Bureau of Statistics, Report on Post-Secondary Schooling of High School Graduates in 1989–1995 (available at: http://www.cbs.gov.il/publications/h_education02/h_education_h.htm, retrieved on 06-09-2021).

records of the Ministry of Education on Israeli high schools' universe during the 1992-2016 school years provide the following student's family-background variables: parental schooling, number of siblings, country of birth, ethnicity, student's detailed study program by subject and level, a variety of high school achievement measures, and test scores in all national matriculation exams in 10th-12th grades. Another source is Higher Council of Education records of post-secondary completed degrees (B.A., MA, and Ph.D.), the institution of study (colleges and universities), and majors (one or two), and completion date; (4) Israel Tax Authority (ITA) information on income and earnings of employees and self-employed individuals for 2000-2018, and three-digit code of industry of employment.

CBS matched and merged these files using the individual-level national I.D. number. The matching is perfect without the loss of observations. We had restricted access to this data in the CBS research lab at its headquarters in Jerusalem. The primary analysis includes twenty-two cohorts who completed high school (12th grade) in 1992-2013. However, we focus on the older cohorts (1992-2005) when examining long-term outcomes. In 2018, the last year in the labor market data we use (we observe post-secondary education data for 2019, too), the youngest cohort in this sample is 31 years old, while the oldest is 44. Choosing this age range ensures that individuals who have completed their higher education - including Ph.D. degrees - are well integrated into the labor market and have usually taken marriage and fertility decisions.

Definitions of Outcomes in Adulthood

This subsection describes the outcomes in adulthood for students in the sample.

Education. University schooling: obtaining a B.A., M.A., or Ph.D. degree. We also use it as an outcome of getting a degree from an Elite University. To study how GCP affects decisions regarding the field of study in university, we create dummy indicators for individual or grouped majors, for example, STEM. We follow the grouping definition of the CBS and focus mainly on areas of study that lead to employment in the knowledge economy.

Labor Market Outcome. *Employment:* An indicator with value 1 for individuals with non-zero number months of work in a given year and a non-zero income, 0 otherwise. *Earnings:* The primary outcome is total annual earnings. We also use earnings from salaried employment and earnings conditional on employment as outcomes. Twenty percent of individuals have zero earnings at age 30-42 in our baseline sample. We dropped from the sample all observations that are six or more standard deviations away from the mean to account for earnings data outliers. Significantly few observations are dropped from the sample each year, and the results are not qualitatively affected by this sample selection procedure. The exact earnings data is also available for our sample's parents for the same years.

The sector of employment: Using a three-digit sector code, we focus on the following sector of work of the knowledge economy:

High-tech Manufacturing industries: Pharmaceutical products for human and veterinary uses, Office and accounting machinery and computers, Electronic components, Electronic communication equipment, Industrial equipment for control and supervision, medical and scientific equipment, Aircraft.

High-tech Services industries: Telecommunications, Computer and related services, Research and development.

Academic: Colleges of education, other educational extensions of foreign universities, Universities.

Knowledge: any of the above.⁶

Personal Status Outcomes. *Marriage:* is an indicator for being married, and we also use an indicator for marriage before the age of 30. Additionally, we measure age at first marriage. *Children:* is an indicator for having at least one child, and again we use indicator for the first child before 30, and we measure age when having the first child, and we also measure the number of children. Having the same data for marital partners, we measure their “quality” based on the following outcomes: participation in GCP, test scores, and income.

4. Methodology: Identification of GCP Short and Long-Term Effects

Previous studies used fuzzy R.D. designs to estimate GCP programs’ effect in the U.S. (Card and Giuliano, 2014 and 2016; Bui, Craig and Imberman, 2014). This design exploits the admission cut-off to GCP. It yields a local average treatment effect of providing gifted education services to students on the margin of gifted child qualification.⁷ However, to understand how GCP affects achievement for infra-marginal gifted children, we use an alternative identification strategy in this paper. We chose as a comparison group gifted children from cities where GCP was not offered at the time. Under the assumption that the timing of starting a GCP in a locality is independent of potential outcomes of gifted children, this sample restriction will permit to overcome the potential selection bias that may arise when comparing cities offering GCP with those that did not.

We use two domains for the selection of the comparison group. The first is the high school study program, which is determined at the beginning of 10th grade. The gifted children’s study program includes several subjects at an advanced level (where the minimum compulsory *bagrut* program requirement is only one subject). These courses, which award five credits (where the basic and intermediate levels reward 3 and 4 credits, respectively), are equivalent to advance placement courses in the US high school system. The most common subjects are English (taken by almost 100percent of gifted students), math (86 percent of boys, 73 percent of girls), physics (64 percent of

⁶ We further validate the reliability of the labor market outcomes that we use by comparing their means to the respective statistics based on labor survey data available for a sub-sample of individuals in our sample. However, we do not use these data in our analysis because the sample is relatively small.

⁷ Abdulkadiroglu, Angrist, and Pathak (2014) used the same identification strategy to estimate the effect of elite schools in Boston and NY.

boys, 31 of girls), chemistry (32 percent by girls and 27 by boys) and computer science (57 percent by boys, 34 percent by girls). A study program that includes several subjects at an advanced level is challenging and demanding, and only very talented or gifted children follow it. Since a student's study program is pre-determined (at the beginning of 10th grade), we can use it to match GCP students to students in localities where GCP was not offered.

The second domain includes test scores that measure general intelligence and ability. Since screening exams for gifted children were administered primarily in cities with GCP, no such systematic test scores are available for selecting a comparison group from other towns. We, therefore, opt for other ability measures. Our data includes two different kinds of exams that measure ability and intelligence. The first is the national exams taken in four fields (science, math, English, Hebrew) during primary school (5th grade) and middle school (8th grade), and the second is the University Psychometric Entrance Test (UPET).^{8 9}

The clear advantage of using the first is that its timing is before the participation in a GCP. However, the limitation is that these national exams were introduced in 2002, allowing us to observe test scores only for the eight younger cohorts (high school graduates of 2006 and later). Thus, when analyzing GCP effects on labor market outcomes (based on cohorts 1992-2005), we rely on the UPET scores. Nevertheless, we argue that matching the UPET scores is a valid approach and provides a rich set of evidence supporting the plausibility of viewing it as pre-determined to the GCP we study.

First, using the eight youngest cohorts (2006-2013), we could implement the matching four times: using middle-school test scores only, using UPET scores only, using primary-school test scores only, and using both primary-school and middle-school test scores. Remarkably, we find and show that the results are almost identical in all four matching models/specifications. Second, we compare the middle-school test scores in English, Hebrew, math, and science, of GCP participants and students of the comparison group when the latter is selected based only on UPET scores. We show that the two groups' density function of these pre-program test scores and the respective means are similar. Third, we show evidence that the UPET scores of gifted children vary very little by the age of taking the exam. This finding is consistent with recent evidence from other contexts, and it supports the idea of viewing these tests as pre-treatment in our context. Fourth, using the older cohorts (1992-2005), where we observe the UPET only, we will first show results based on a sample of only 10th-11th UPET early takers (about 60% of the gifted children) who took the exam even before midway into the program. We will then expand the sample gradually to include 12th-grade UPET takers and late takers and show that GCP treatment estimates are remarkably identical across these samples.

⁸ The UPET is required for applicants to universities in Israel, and it is administered by The National Institute for Test and Evaluation (NITE). According to NITE, the UPET is a tool for predicting academic success at higher education institutions in Israel.

⁹ Table A1 in the online appendix reports the correlations between 8th-grade Metzav test scores and UPET among GCP participants and students from other cities (before matching).

We also discuss three potential concerns regarding our identification. The first is that there could be systematic differences between the potential outcomes of gifted children in cities with a GCP, which are typically larger cities, and the potential outcomes of gifted children in other cities. First, although cities with GCP are larger, students in these cities have similar characteristics, including parental background (education and income) and academic outcomes, like those in other cities (see tables A2 and A3 in the online appendix for these comparisons). Second, we show that restricting the comparison group to include students from larger cities without a GCP yields the same results.

The second potential concern regarding our research design is that families may relocate based on access to the GCP program in the locality. Therefore, we examined whether families with GCP participants had a higher mobility rate before 10th grade than families with gifted children that did not participate in a GCP. We find no such differential mobility rate. These results are shown in the online appendix Table A4.

The last potential concern regarding our research design could be that students are selected into GCPs according to an unobserved, non-cognitive characteristic. For example, highly motivated students probably enroll in GCPs at higher rates than less-motivated students, even conditional on their ability. We account for this unobserved factor by including the high-school study program of each student in the matching model. As discussed earlier, students choose their high-school study program before enrolling in high school. This study program, unique for each individual, measures each subject's scope and level of study in the *bagrut* program. Therefore, accounting fully for the detailed study programs is a good proxy for the student's motivation before participation in the GCP.

4.1 Evidence on UPET Scores by Testing Age

Remarkably, even though most students in Israel start college education at age 22-23,¹⁰ the vast majority of GCP participants take this test while in high school and even before midway into the program (see Figure A1 in the online appendix). Figures 1.A and 1.B, present the UPET score distributions for two samples of GCP participants, those tested in 10th or 11th grade versus those in 12th grade. We present this evidence separately for the sample of younger and older cohorts. The first (upper left) panel of each figure shows the total score distributions, and the other three panels show the distributions for each of the three domains of UPET. The Kolmogorov–Smirnov test (K.S. test) for equality of the two sample distributions is shown in the figure.¹¹ The P values of this test for the

¹⁰ This late age of starting college education results from the compulsory military service that starts immediately after high school graduation. Men serve for three years, and women do two years.

¹¹ The Kolmogorov–Smirnov test is a nonparametric test of the equality of continuous probability distributions from two samples. It quantifies a distance between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. The distribution considered under the null hypothesis is a continuous distribution but is otherwise unrestricted. The KS test is one of the most useful and general nonparametric methods for

composite score and the quantitative and verbal (Hebrew) reasoning indicate that the pair-wise sample densities are not statistically different. The P-value for the total score is 0.47 for the younger cohorts (0.45 for the older cohorts), for the quantitative reasoning 0.76 (0.79), and verbal reasoning, it is 0.70 (0.25). The p-value (0.07) rejects that the two densities are statistically different only for English for the older cohorts (0.25 for the younger). Online appendix figures A2.A and A2.B show the respective evidence for non-gifted children in our baseline sample with slightly larger differences.

Online appendix figures A3.A and A3.B show the densities for those tested in 12th grade versus those tested after high school. The figures show that those tested later have slightly higher English and Hebrew scores. However, the quantitative score densities of these two groups are still very similar. Therefore, we will also check the robustness of our main results to exclude the Hebrew and English UPET scores from the matching, as will be discussed later.

Perhaps the finding that UPET test scores do not vary among highly talented individuals by the age of testing is not surprising because its structure and content are very similar to the SAT and CAT used in the U.S. for the same purpose. In addition, these tests were shown to be highly correlated with I.Q. and other ability measures (Koenig, Frey, and Detterman 2004, Beaujean et al. 2006, Koenig, Frey, and Detterman, 2008), and we should not expect them to vary much by age. So is the UPET, but the evidence for its correlation with IQ test scores is more limited.¹² Another evidence supporting the relative invariability of the UPET score with age is that gifted children who take this test more than once achieve a very modest improvement in test scores. This gain seems particularly low compared to the gain experienced by non-gifted children.¹³

4.2 Constructing the Comparison Group

Based on the quasi-natural experiment that GCP is not offered in many cities in Israel, we use propensity score matching to choose the comparison group.¹⁴ That is, we use the following three-step algorithm:¹⁵

1. Estimating the propensity score using a Logit specification

comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

¹² Evidence of a high correlation between UPET scores and IQ is presented in a NITE study of these relationships among students who submitted requests for extra time due to ADHD.

¹³ Goodman, Gurantz and Smith (2020) provide similar evidence for SAT scores of second takers. Students with first scores nearest to the 700–1500 thresholds had almost five times the largest SAT score improvements than those nearest the 1600–2300 thresholds. Low-scoring students' second take total scores are 91 points higher than their first, relative to an only 22-point gain for higher-scoring students. The increase for students at the very top of the score distribution, say over 2000, is even smaller.

¹⁴ Rosenbaum and Rubin (1983) proposed an approach that circumvents the curse of dimensionality when using selection on observables for the identification of causal effects. They provide proof that states that if treatment assignment can be ignored given x , then it can be ignored given any balancing score that is a function of x , in particular the propensity score.

¹⁵ See Abadie and Cattaneo (2018) for a survey of econometric methods for program evaluation and a useful comparison of matching/propensity score models with other methods.

$$(1) \quad P_i(X_i) \equiv \Pr(T_i = 1 | X_i)$$

We include in the logit regression the following covariates: parental schooling, number of siblings, three dummy indicators for whether the father, mother, and child were born in Israel, dummy indicator for each *bagrut* subject at five credits, test scores (in the three domains of the UPET or the four domains of the middle-school tests), and cohort fixed effects.

2. Matching GCP participants to the comparison group using the nearest neighbor. We match without replacement. We include in our sample only matches in a caliper of 0.1 standard deviations of the propensity score, and with the same sex and same religious status of the school.¹⁶

3. Estimating the following controlled regression

$$(2) \quad Y_i = \alpha_i + \beta' X_i + \tau \cdot T_i + \varepsilon_i$$

This propensity score matching and regression combination allow for enhanced robustness to misspecification. As long as the parametric model for either the propensity score or the regression functions are specified correctly, the resulting estimator for the average treatment effect is consistent. A notion discussed and termed ‘double robustness’ in Robins and Ritov (1997) and Imbens (2004).¹⁷ The standard errors of the program effects estimates were clustered at the school level.¹⁸

During the first part of our analysis, we focus on eight cohorts of 2006-2013 high school graduates. We implement four matching specifications: (1) Matching I, where we use 8th-grade test scores as a measure for ability (available for 2006-2010 high school graduates); (2) Matching II, where we use UPET scores as a measure for ability (2006-2010 high school graduates); (3) Matching III, where we use 5th-grade test scores as a measure for ability (available for 2009-2013 high school graduates); (4) Matching IIII, where we use both 5th and 8th-grade test scores as a measure for ability (2006-2013 graduates).¹⁹

Note that 2006-2010 high school graduates are at ages 27-31 in 2019, allowing us to study short- and medium-term effects on high-school academic achievement and postsecondary education attainment and field of study decision. However, when interpreting the results based on a sample of 2009-2013 high school graduates, it is essential to remember that these cohorts were at ages 24-28 in 2019. Thus, we do not observe their final postsecondary outcomes (almost half of them did not complete their B.A. studies in 2019).

Figure 2.A presents the propensity score distribution before and after matching I, and figure 2.B presents the distribution before and after matching II. In matching I (II), we match 477 (371) of the sample of 551 (452) GCP participants. The unmatched are mainly from the top of the propensity score

¹⁶ We further validate that our main results are not sensitive to the specification by running alternative matching specifications. The results are shown in the online appendix Table A5, and they are identical to our main results.

¹⁷ See Abadie and Imbens (2002) for details regarding the use of OLS with the matching procedure weighting.

¹⁸ We also provide an alternative calculations for the standard error in the online appendix Table A6. The table shows that the clustered standard errors are almost identical to the correction specified by Abadie and Imbens (2008), and to clustered bootstrapped standard errors.

¹⁹ We observe either 8th grade or 5th-grade test scores for most of the students. However, for students with both 5th and 8th-grade test scores, we used the 5th-grade test scores in Matching IIII.

distribution. The propensity score density graphs of the GCP participants and their matched counterparts are perfectly aligned and not distinguishable in both cases.

Figure 3 presents the density distributions of the 8th school Metzav scores before and after matching I. As expected, there are substantial differences between GCP participants and the sample of students from other cities (before matching) since most students are not gifted students. However, the matching allows us to eliminate most of these differences in test scores, as the distributions in math and Hebrew test scores become statistically indistinguishable (the distributions in science and English are still statistically distinguishable, but the differences are much smaller).

In Figure 4.A, we present the density distributions of the UPET scores (total score, quantitative reasoning, verbal reasoning, and English) before the matching (I). Again, unsurprisingly The KS test p-value shows that the UPET total score distributions are statistically distinguishable before the matching with a clear advantage for GCP participants. In Figure 4.B, we present the density distributions after the matching, where all distributions are not distinguishable statistically; For the total score, the K.S. test p-value is 0.71. Similar results are evident for the pair-wise distributions of the quantitative reasoning, verbal reasoning, and English parts of UPET.

Table 1, columns 1-2, presents detailed summary descriptive statistics for the main sample (based on matching I) for the variables that we use in the propensity score estimation and other background variables by GCP participants and their matched counterparts we use as a comparison group. In column 3, we present the mean differences, and in column 4, the p-values for the test that the differences are statistically different from zero. In columns 5-8 (9-12), we present the same evidence for the girls (boys) sample. Panel A presents evidence for students' demographic and socio-economic characteristics. We also present evidence on the 8th grade Metzav scores in the four subjects in panel B.

First, it is interesting to note that gifted children, either GCP participants or others, come from higher socio-economic backgrounds than regular Israel students. For example, mean mother and father years of schooling are around 15.5 years for these two groups, higher than among non-gifted children (where years of education of the father is 14.4 and of the mother is 14.5). Moreover, the difference between the gifted and non-gifted samples in the father's income is 25 percent, and the mother income is 34 percent.

Thirty-six parameter estimates are shown in the table. Note that all variables are included in our propensity score estimation, except for father and mother income. All indicate a perfect balance between treatment and comparison group, with only one out of the 12 entire sample estimates statistically significant at the 90% level. Overall, the differences between GCP participants and other gifted children, the differences in parental education and number of siblings, are very small and not statistically different from zero.

Note that the close similarity between the two groups is also evident in parental annual earnings, which we do not use as controls in estimating the propensity score regression. Therefore, evidence on

balancing parental income is particularly important. For example, the Father's (Mother's) average yearly earnings are 170,000 (92,000) NIS in the treated group and 186,000 (98,000) in the comparison group. These differences of less than 20,000 NIS yearly (about 4,000 US\$ in 2003) are not statistically different from zero. However, there is a small and statistically significant difference of about 21,000 NIS for mothers' annual earnings in the girls' sample when the average yearly earnings are 80,000 NIS in the treated group and 101,000 in the comparison group.

The four Metzav mean scores are very balanced, as expected, given the evidence shown in Figure 3, with only one marginally significant difference in English. Table 2 presents descriptive statistics for the *bagrut* Study Program, with none out of 11 outcomes showing a statistically significant (90%) difference between treatment and comparison group in the entire sample. The table shows that gifted children have much higher participation rates in science advanced study programs. For example, in Math (86% of the boys and 73% of the girls), Physics (64% and 31%), Computer Sciences (57% and 34%). Gifted children also have many *bagrut* study credits (30.3 and 29.6, while the requirement is for 21).

Before discussing the results of estimation GCP effects on outcomes, we note that matching only on UPET scores eliminates almost all the differences in 8th-grade test scores. Figure A4 in the online appendix presents the test score distribution in the national exams ("Metzav") in English, Hebrew, math, and science, before and after matching II. Although the KS test is still statistically significant after the matching, the figure shows that most differences between score distributions before matching are eliminated. This result supports the validity of our research design. The main caveat for our application could have been that the UPET score is positively affected by participation in GCP. However, this would imply a negative difference in their pre-treatment abilities between our treatment and comparison group. Instead, we find a very small positive difference in a pre-treatment proxy for abilities, alleviating these concerns.²⁰

4.3 Simple Treatment-Comparison Differences in Outcomes

Table 3 shows the treatment-comparison contrast in mean test scores in all *bagrut* compulsory subjects. The unconditional differences in the entire sample are mostly very small, none larger than 3 percent relative to the comparison group mean, where half of the estimates are negative and half positive. However, only two are statistically significant at the 90% level (Hebrew and English). Among the girls' sample, we find a statistically significant difference in the mean composite score and three subjects (Hebrew, English, and literature). However, these differences are positive and very small in magnitude (below 4%). We do not find a statistically significant difference in the mean

²⁰ We also show in the online appendix figures A5.A and A5.B, the distribution of the UPET scores before and after matching I (where we use only 8th-grade test score in the logit specification). We find that the matching partly eliminates the difference in UPET scores between the groups before the matching.

composite score among the boys, and there is only one statistically significant difference in English. Thus, this table suggests very tiny effects of GCP on *bagrut* test scores. The following section will show that estimating GCP effects in a controlled regression yields mostly very small negative or zero effects.

Table 4 present the treatment-comparison simple differences in university outcomes. In contrast to the impression of little gender difference in the effect of GCP on 12th-grade test scores, the impact on academic degree attainment and choice of field of study is very similar for men and women. The table shows a statistically significant difference in university degree attainment between GCP participants and the comparison group of 7p.p. However, note that in this sample, there are students who are still studying towards their B.A., so this effect might represent that GCP participants finish their studies earlier than the comparison group, rather than attaining more B.A. degrees.²¹ We also find a largely positive and statistically significant difference in completing BA degrees with a double major and a small positive difference in attaining Ph.D. degrees. We again caution about the estimated effect on advanced degrees because of the young age of the students in this sample).

The university's choice of studies, also presented in Table 4, reveals an interesting pattern. Among GCP participants, boys and girls have a much higher enrollment rate in math, computer science, physical and biological sciences than gifted students in regular classes. Against these meaningful increases, GCP participants significantly decline in engineering degrees relative to non GCP participants. The net difference in STEM degrees is positive and statistically significant for both genders.

5. Results: Short- and Medium-Term Effect of Gifted Children Programs (2006-2013 Sample)

In this section, we present results using the sample of GCP participants and comparison group's students from the cohorts of 2006-2013 high-school graduates, by the following outcome groups: (a) high school, (b) university schooling, (c) fields of university studies. The following section will present the results on these outcomes among the sample of older cohorts (1992-2005). Then, we will extend the analysis to (d) employment and earnings, (e) employment in sectors of the knowledge economy (f) marriage, divorce, and parenthood. First, we will show the results based on matching I while using the middle-school standardized Metzav test scores to measure ability. Then, in subsection 5.3, we will show that the results are identical when using other ability measures.

5.1 Effect on High School Outcomes

Figure 5 presents the estimates and confidence intervals (10 percent significance level) for effect on matriculation certification and *bagrut* exam scores obtained from estimating equation (2). We show

²¹ This is indeed the case, as we do not find GCP effects on B.A. rate in the older samples, where the rate of B.A. is 98%, but we do find effects on studies timing (see a discussion in section 6).

estimates based on the entire sample and separately for females and males. We present the effect on compulsory subjects (civic studies, Bible studies, English, Hebrew, history, literature, and math) and the four elective subjects most popular among gifted children, computer science (CS), physics, biology, and chemistry. We do not present the effect on obtaining a matriculation diploma because this outcome for both treatment and comparison groups is almost 100 percent. We also show the comparison group means and standard errors in the figures. These are also presented for females and males separately.

Overall, the point estimates on test scores are tiny in all subjects, with significant positive effects on English (1.77 relative to a baseline mean of 90.90) and significant adverse effects on math and civic studies (the largest estimated effect, -1.83 relative to a baseline of 85.66). The difference in GCP effects between male and female are small, where it seems that male GCP students have slightly more adverse effects. The online appendix shows that the results are very similar if we use scores on a 1-6 scale (presented in Figure A6²²) and standardized Z-scores (presented in Figure A7). Again, for the entire sample of boys and girls, we find minor GCP effects on *bagrut* scores range between an adverse effect of 0.06SD (math and civic studies) and a positive effect of 0.07SD (English).

Discussion of mechanisms: gifted children program learning environment

The pattern of tiny effects of GCP on test scores in *bagrut* exams, with small adverse effects on math and civic studies, is interesting. It is especially intriguing given the abundance of educational inputs that GCP participants enjoy relative to the comparison group that we use. In the online appendix Tables A7 and A8, we present various school and class level inputs for the two groups. GCP classes are much smaller, the socio-economic background of peers in these classrooms is much higher, and the averages of outcomes are much better. We also based on details of GCP in Israel that their teachers are more qualified and receive additional training and that the budget per student is higher. So, what can explain the lack of positive effect of GCP on achievements at the end of high school exit exams?

We offer two mechanisms based on evidence in recent studies in the economics of education. The first is the change in ordinal ranking in terms of the ability and achievements of GCP participants. When academically gifted students are placed in self-contained programs, they usually experience a new environment with equally competent peers, more challenging materials, and more rigorous requirements. One reality they inevitably must encounter is a more talented peer group than they are used to in a regular classroom. This could be beneficial or harmful at the same time. This is beneficial because a peer group of equal academic caliber gives personal validation to one's identity and serves to reinforce each other's talents and interests mutually. This can be harmful because individuals, particularly those who might already feel insecure, are likely to think that the very talented people

²² This is important, since for the older cohorts we observe only scores in 1-6 scale. However, as we show here, the results are not affected by whether we use the full score or the 1-6 score.

have touted about them. They may also find that the top student status they have enjoyed in the regular classroom is no longer a sure thing, as there are potentially more talented people in the new peer group. When two students of the same ability or achievement level are placed in different classrooms or programs, the one with the high ability or achievement group tends to temporarily lower self-concept in respective domains than the one with the less able group. This effect has been labeled the Big Fish Little Pond Effect (BFLPE; Marsh, Chessor, Craven, & Roche, 1995; Marsh & Parker, 1984, Preckel et al. 2010. Herrmann et al. 2016.).²³ A big fish that used to be in a little pond may reassess their competence when put into a larger pond with even bigger fish.

Although the BFLPE model is not specific to gifted programs, facets of the BFLPE have been examined with gifted and high-ability students ranging in grade from the early elementary years (Tymms, 2001) to the college years (Rinn, 2007). The practical implications are obvious and have already produced repercussions in the gifted education community (e.g., Dai & Rinn, 2008; Plucker, Robinson, Greenspon, Feldhusen, McCoach, & Subotnik, 2004).

The influence of change in students' ordinal ranking may ultimately impact beliefs and outcomes through its effects on an individual's actions and investment decisions or others around them. Change in ordinal ranking can affect non-cognitive skills in some or all subjects, such as grit, resilience, and perseverance (Valentine et al., 2004). In our context, GCP participants moved from an environment in middle school where they were most likely at the very top of the ranking in their class to a class with peers who were, on average, their equal. As a result, their rank order most likely declined. Elsner and Isphording (2017) show that students' ordinal rank significantly affects educational outcomes later in life, such as finishing high school, attending college, and completing a 4-year college degree. Exploring potential channels, these authors find that students with a higher rank have higher expectations about their future career, a higher perceived intelligence, and receive more support from their teachers. Murphy and Weinhardt (2020) show that ordinal academic rank during primary school has lasting impacts on secondary school achievement independent of underlying ability. In addition, they find significant effects on test scores, confidence, and subject choice during secondary school, even though they have a new set of peers and teachers unaware of their previous ranking in primary school.²⁴

Earlier studies have shown that gifted students who move from heterogeneous classes to a homogenous classroom where all students are gifted are also subject to BFLPE. Studies have shown that this change lowers their academic self-concept and increases their anxiety (Marsh and Parker 1984, Marsh 2005, and Marsh and Craven 2002). In addition, with the increasing ability level of the reference group (the gifted class), students in gifted classes compare themselves with high-ability

²³ See Marsh et al. (2008) for a review.

²⁴ Recent papers have shown that an individual's rank impacts their well-being and job satisfaction, conditional on their cardinal relative position (Brown et al., 2008; Card et al., 2012).

peers, as do their teachers. These dynamics are especially harmful if students are transferred from a school or class where they have ranked high academically to a lower rank.

Zeidner and Schleyer (1999), and Praekelt et al. (2008) report evidence-based on Israeli data that extended Marsh et al. (2008). They examined the effect of BFLPE on academic self-concept, test anxiety, and school grades in a sample of 1020 gifted Israeli children participating in two different educational programs: (a) special homogeneous classes for the gifted and (b) regular mixed-ability classes. The hypothesis was that gifted students enrolled in special gifted classes would perceive their academic ability and chances for academic success less favorably than students in regular mixed-ability classes. These negative self-perceptions, in turn, will serve to deflate students' academic self-concept, elevate their levels of evaluative anxiety, and result in depressed school grades. Overall, the data supports reference group theory, with the big-fish-little-pond effect supporting all three variables tested. Also, academic self-concept and test anxiety were observed to mediate the impact of reference groups on school grades. Though this evidence is consistent with the findings in economics, we note that they are correlational, and we should be cautious in interpreting them as causal.

A second and related mechanism is the gender peer effect. There is growing evidence that a higher proportion of female students in the classroom improves the learning environment and raises boys' and girls' achievements (Lavy and Schlosser 2012). However, the ratio of female students in a gifted children's classroom is lower on average than in regular classes. Therefore, in this case, a gender peer effect will lower the learning outcomes of both genders.²⁵

Lastly, as mentioned earlier, the GCP's studies program incorporates advanced concepts and topics not directly relevant for the *bagrut*. Finally, note that gifted students' *bagrut* test scores are typically very high and allow them to enter most university degrees. Thus, it is very plausible that GCP participants get educational benefits that are not manifested in higher test scores.

5.2 Effect on University Schooling

Figure 6 presents the estimates for university outcomes, focusing on BA and MA degrees. We find a small and statistically significant GCP effect on BA attainment, but as discussed earlier, it might be due to students who did not complete their BA studies yet. We also find that male GCP participants tend to increase their enrolment in BA programs in non-elite universities. One possible explanation that we discuss later is that GCP participants enroll in special IDF programs that include post-secondary studies (mostly in non-elite universities).

²⁵ The literature on tracking students by ability is of limited relevance to the context of this study. This literature refers mostly to grouping students based on median ability, the upper half having a much lower mean than gifted children. However, it is important to note that several recent studies provided evidence that the above-median ability students benefit from tracking (see, for example, Duflo, Dupas, and Kremer 2011). Other relevant studies include Betts, and Shkolnik. 2000, Epple, Newton, and Roman (2002), Figlio and Page (2002).

Notably, the program has a significant positive effect on getting a double major. The comparison group average rate is 35 percent, and the estimated impact is 0.27, implying a statistically significant 77 percent increase. The estimated effect on men is larger than on females though the two estimates are not different statistically.

Fifty-three percent of gifted children choose a STEM subject for their university schooling. The effect on double major can be divided between those whose double major include two STEM subjects versus others. Almost half of all double majors have two STEM subjects (16 out of 35 percent). GCP enhances this ratio. Its overall effect is to increase double major with two STEM subjects by 16 percentage points relative to a baseline of 20 percent. Both genders drive this effect with a similar relative increase, and when estimated separately by gender, both estimates are statistically different from zero.

Graduating with a double major, especially with two STEM subjects, can be related to the multipotentiality of gifted children. This concept has been defined as “the ability to select and develop any number of career options because of a wide variety of interests, aptitudes, and abilities” (Kerr, 1991, p. 1). Multi-potentiality is widely cited as a characteristic of the most gifted individuals who have the ability and interest to pursue various activities and goals, especially related to career choice (Sajjadi et al. 2001, Sampson et al. 2008). This effect may be activated and enhanced in an environment where giftedness status is ‘formally’ recognized as in a GCP environment. GCP has no statistically significant effect on MA degree attainment among girls and boys, but the students in these samples are still below age 30; thus, they might still study toward advanced degrees.

Figure 7 presents results regarding the choice of GCP participants of university field of study (major). We show estimated effects for the following majors: medicine, a STEM subject, and the primary majors included in STEM: math, computer science, engineering, physical sciences, and biological sciences, and also business & management, social sciences, humanities, and law. The evidence shows an interesting pattern: a zero effect on STEM that masks two opposite direction changes in some of these majors. First, GCP participation increases majoring in math, computer, and physical and biological sciences, with a substantial effect on men and women. Second, it causes a sharp decline in engineering degrees, especially for women.

The increase in math/computer science is 60 percent for men (16 percent point increase relative to comparison group mean of 26 percent) and almost 90 percent among women (7 percent point increase relative to comparison group mean of 8 percent). The absolute decline in engineering degrees is of similar magnitude. It is a 44 percent decline for men (14 percent point decline relative to the comparison group mean of 32 percent) and a 45 percent (statistically insignificant) decrease among women (8 percent point relative to 18 percent). Another dramatic effect on studying science is the increase among women majoring in physics and related subjects. The counterfactual is 4 percent, and the GCP effect is a significant 13 percentage points increase, implying a massive 325 percent increase.

The dramatic result from Figures 6 and 7 is how significant GCPs' effect shapes adolescents' university choices. The realization of academic potential is often perceived as acquiring higher education, impressive academic achievements, or pursuing a prestigious profession. But what motivates gifted adolescents to make future professional choices and the themes that guide them? To what extent does the environment impact these choices? Studies in educational psychology on the formation of gifted adolescents' identity (ex. Zeidner et al. 2005, 2015, Shani et al. 2009) provide insights to these relevant questions for understanding and interpreting our results. They argue that the desire to realize their potential and the concern not to choose areas considered "potential waste" is a central theme among gifted adolescents, especially those enrolled in gifted classes. The label 'gifted' impacts their choices; they are affected by their expectations to make the most of their high abilities, i.e., their potential, and exhibit a future focus that does not characterize non-gifted adolescents.²⁶ They feel obligated to realize their potential in its conventional sense. This leads to an interesting paradox - precisely, those with high abilities who can choose any field of study are those who feel that they have only a limited range of options. In their experience, they are limited to the same possibilities that will be considered to realize the potential.²⁷

We should interpret these results cautiously since not all students have finished their university studies. However, as we will discuss in section 6, the effects' absolute magnitudes are similar to our results when analyzing the older cohorts who completed their studies by 2018.

5.3 Using Different Measures of Ability in the Matching Specification

We test the sensitivity of our results to the ability measure we use in our matching model. As discussed in section 4, we implement four versions of the matching specification, with the national standardized 8th-grade test scores for the cohorts of 2006-2010 (I), with UPET scores for cohorts 2006-2010 (II), with the national standardized 5th-grade test scores for the cohorts of 2009-2013 (III), and with both 5th and 8th-grade test scores for the cohorts of 2006-2013 (IV).

Table 5 shows the estimated GCP effects on outcomes by specification and sample. Overall, the table shows minimal variation in estimated effects across the specifications. For example, the

²⁶ In an attempt to identify additional themes related to the development of self-identity in gifted adolescents, Shani-Zinovich, I., (2007), and Zeidner and Shani-Zinovich (2013) show that gifted adolescents in gifted classes in Israel were characterized by the following profile: a strong commitment to academic achievement; serious consideration of prestige and other external factors in choosing a vocation; a strong commitment to their professional future at a relatively young age; and a heightened fear of academic failure.

²⁷ Some gifted students experience pressure from their parents to choose an occupation based on prestige instead of values or interests (Colangelo & Assouline, 2000). Kerr and Colangelo (1988) found that 50% of intellectually gifted college-bound students in their study selected majors from only three areas, engineering, health profession, and physical science, even though they were presented with almost 200 possibilities had identified broad extracurricular interests. Fredrickson (1982) noted that multi-potential students showed less variability in occupational choice than students who were not identified as multi-potential. Parents who pressure their gifted and talented children to consider only prestigious occupations can cause these students to foreclose potentially viable options prematurely.

effect on math and civic studies scores, which were negative and statistically significant in the main specification, remains negative and statistically significant in all other specifications. Still, there is some variation in GCP effects estimates on *bagrut* test scores. However, the impression that GCP effects on *bagrut* test scores are tiny in magnitude, with negative effects on math and civic studies and a positive effect on English, is very similar in all model specifications.

The similarity of the results across the four matching models is manifested even more strongly in estimating GCP effects on higher-education outcomes. The main results discussed earlier remain the same in every matching specification. For example, the substantial positive effects on a double major BA and the double major BA in STEM are pretty stable across specifications, with estimates ranging between 0.27-0.17 and 0.20-0.13 (all estimates are statistically significant). Similarly, the positive effects on math, computer sciences, and physical sciences and the adverse impact on studying engineering are apparent in all matching specifications.

Our interpretation of the stability of the results across matching specifications is that all the ability measures we use, based on cognitive or intelligence outcomes, pick up similar fundamental ability traits unique to gifted children. This interpretation is enhanced further given that these alternative ability measures are assessed at different ages (11, 14, 16-17), yet they yield the same estimates of GCP effects when used in the propensity score regression. Considering the evidence we discussed earlier in sections 4.1 and 4.2, and the fact that outcomes of gifted children are unrelated to their background characteristics (e.g., parental education and income), allow concluding that the intelligence of gifted children is pretty constant during their primary to secondary schooling. Their ability is not significantly affected by events they experience during primary, middle, and high school (such as participation in a GCP). Importantly, this finding allows us to treat the UPET scores as pre-determined measures of ability and use them in the propensity score regression for estimating GCP effects for the older cohorts (1992-2005). We will first use this sample to validate that the short-medium GCP effects are similar to those reported in this section. Then, we will extend the analysis into estimating GCP effects on longer-term outcomes, such as advanced academic degrees attainment and employment in the knowledge economy and income.

5.4 Heterogeneity of the Effects by the age of Joining a GCP

What is the value-added of attending a GCP from a young age? Despite its importance, this question was not addressed yet in the literature. Our unique setting allows studying the GCP effect by its length.²⁸ About half of the GCP participants in our sample started the program from middle school. The others began the program in high school. To study this issue, we define *M* as an indicator of a GCP that started in middle school and continued through high school and estimate the following equation:

²⁸ We use in this estimation the younger cohorts sample because for the older cohorts we do not have information about middle school GCP.

$$(3) \quad Y_i = \alpha^0_i + \beta^0 \cdot X_i + \gamma \cdot M_i + \tau^h \cdot T_i + \delta \cdot (T_i \cdot M_i) + \varepsilon^0_i$$

The coefficients of interest in the above equation are τ^h , the effect of a shorter GCP (only in high school), and δ , the value-added impact of a GCP that started in middle school. The overall effect of a GCP that began in middle school is $\tau^h + \delta$.

Table 6.A and Table 6.B report results when we include the 8th and 5th-grade test scores in the propensity score regression, respectively.²⁹ The results presented in the two tables are identical. This finding is consistent with our earlier evidence and conclusion that the cognitive outcomes in 5th and 8th-grade tests measure the same ability traits, which do not vary from age 11 to 14, among gifted children. We find no statistically significant heterogeneity in GCP effects on bagrut test scores by the length of the program, except for English test scores in the 2009-2013 sample; a slightly larger positive effect on students in the GCP high school only program. Another difference is that these students also experienced a small decline in the Bible and civic studies test scores. Overall the interpretation of these results is that GCP's little effects on high-school academic achievement do not vary by length of participation.

Similarly, for higher-education outcomes, we find similar patterns in both program types, with positive effects on double major BA, double major BA in STEM, math, computer sciences, and physical sciences, and a negative effect on engineering. Again, the only statistically significant difference is for effect on a double major BA in STEM, when using the 2009-2013 sample. However, we note that this sample includes students who did not yet complete their BA schooling; therefore, we view evidence from the 2006-2010 sample as more reliable.

To conclude, the evidence we present in Table 6.A and 6.B suggests that GCP effects do not vary by program length. This result is consistent with the finding that GCP does not affect the participants' *bagrut* exams academic achievement but does affect university study program choices. The mechanisms for these large effects on higher-education outcomes are probably psychological, as discussed in section 5.2, and they are not affected by the length of participation.

6. Results: Short-, Medium- and Long-Term Effect of Gifted Children Programs (1992-2005 Sample)

This section presents the estimated GCP effects using the older cohorts of 1992-2005 high school graduates. The students in this sample are at ages 31-44 in 2018, allowing us to extend the analysis to long-term outcomes. We focus only on students who took the UPET (92% of the gifted children in this sample) and use these test scores in the propensity score logit regression. As our primary sample, we include students who took the UPET during their 10th or 11th grade. Figure 8 shows the propensity score distributions before and after the matching. We matched 1,735 of the 1,949 GCP participants in this sample, and the propensity scores density functions of the treated and

²⁹ Table A9 in the online appendix compares the characteristics of GCP participants by the two types of programs, those that began in middle school versus those that began in high school.

comparison align perfectly after the matching. Figure 9 presents these distributions before and after the matching. We obtain perfect alignment for the total UPET score and also for each of its sub-fields. In the online appendix, Tables A10 and A11 provide descriptive statistics (corresponding to Tables 1 and 2 of the younger sample) showing the balance between treatment and comparison group in all variables included in the matching regression.

Then, we also estimate GCP effects while varying the sample's definition based on the age at which students took the UPET: (1) the baseline sample (2) adding to the baseline sample UPET 12th grade test takers (3) adding to the baseline sample UPET test-takers at all ages (4) the baseline sample, where we restrict the comparison pull to include students from larger cities.³⁰ We will first discuss the results based on the baseline sample, and section 6.6 shows that the results are identical in all extended samples.

6.1 Effect on High School Outcomes

Figure 10 shows the effects of GCP on compulsory and on the most selected elective subjects (Table A13 in the online appendix shows the unconditional differences). The results are very similar to those estimated when using the younger sample with small and mostly insignificant estimated effects. Again, we find a small negative impact on math, civic studies, Bible studies, Hebrew, and physics. We also find positive effects on history and literature. Still, all estimated effects are tiny. This sample holds the overall impression of the insignificant and mostly negligible adverse effects of GCP on high-school academic achievements.

6.2 Effect on University Schooling

Figure 11 shows the effects of participation in a GCP on degrees attainment (Table A14 in the online appendix provides unconditional mean differences). Again, the results are similar to those estimated for the younger sample, with two crucial differences. First, we do not find any effect on BA degree attainment because the mean in the comparison group is 98%. This is consistent with the interpretation that the positive impact on BA attainment obtained while using the younger sample was due to GCP participants finishing their BA degrees earlier. Second, we find a positive estimated effect on Ph.D. attainment of 3 percentage points, relative to a baseline of 13 percent. The impact on Ph.D. degrees is driven by an increase in the rate of Ph.D. degrees in elite universities (40%).

Figure 12 shows the effects on the field of study (unconditional mean differences appear in Table A13, too). As we found based on the younger cohort sample, here as well, the estimates show

³⁰ Table A12 in the online appendix shows results based on alternative matching specifications. First, we only use the UPET quantitative scores for matching (dropping the English and verbal scores). Second, we omit the *bagrut* program variables from the matching regression. The results obtained from these two alternative matching regression specifications are similar when we use all ability measures together, but there are some differences. Therefore, we prefer the evidence that we obtain when we use to match all the ability measures.

positive effects on the tendency to study math, computer, and physical sciences and a negative for engineering. However, this effect's magnitude is somewhat smaller, a 13% increase in math and computer sciences, a 31% increase in physical sciences degrees, and a 17% decline in engineering degrees. Additionally, we see a large increase in social sciences degrees, MD degrees among girls, and a smaller decrease for boys.

Figure 13 presents the estimated GCP effect on studies' timing. The graph shows a statistically significant difference in the age at which GCP participants started and finished their B.A. degrees. Concerning M.A. degrees, there is again a negative effect on the age at which GCP participants started and finished. However, the estimated effect on the age when completing M.A. is statistically significant only for boys. These effects are mainly driven by GCP's students that finish their B.A. degree before 21 (the age of finishing regular services in the Israeli army for boys). When we define an indicator variable for completing a B.A. degree before 21, we find a large increase of 2.1 percentage points relative to a baseline of 6% in the comparison group. This effect is statistically significant (with a standard error of 0.9p.p.).

This finding indicates that participation in GCP is interpreted by the IDF (Israel Defense Forces) as a signal of exceptional talent. This signal is beyond the evidence-based tests done by the military. It leads to enlistment in special military units and enrollment for some in university schooling during the compulsory service (men three years and women two years).

6.3 Impact on Labor Market Outcomes

Seventy-eight percent of gifted women and 74 percent of gifted men were employed in 2018 (see Table 7 for labor market outcomes' means and unconditional mean differences). These proportions are almost unchanged relative to the rates in 2016. Figure 14 shows the estimated effect of GCP on these rates, zero. We also estimate the effect of GCP on the likelihood of being self-employed. Eight percent of GCP men and ten percent of GCP women are self-employed in 2018. Again, the impact of GCP on these rates is zero.

The mean income of GCP participants and gifted children in the comparison group is 50 percent higher than those of non-gifted students from the cities with no GCP program. However, the income differences between the two gifted children groups are minor and not statistically different from zero (As shown in Figure 15, there is a difference of -0.04 relative to a comparison group mean of 2.34 hundred thousand New Israeli Shekels (NIS)). The results for salaried earnings for the entire sample and males and female sub-samples are very similar. Restricting the sample to those employed only leaves the results unchanged. This result is expected because GCP does not significantly affect the employment rate. Figure A8 shows the estimated effect on the average annual income for 2016-2018, and the results are identical. Figures A9 and A10 show the estimated GCP effect on the natural log and the income rank, providing similar results.

To further validate these estimates, we estimated the effects on the labor market outcomes with three additional samples: (1) the baseline sample, with exact matches on the year of graduating high school (2) A subset of the baseline sample, including the 1992-2000 cohorts only. (3) The baseline sample was restricted to those living in Israel in 2018.³¹ The first sample allows us to validate that our results are not due to each group's different mix of cohorts. The second sample allows focusing on older cohorts that in 2018 are 36-44 years old, already on a much more stable stage of their professional career. The third exclude students for whom we do not observe their labor market outcomes in 2018. The results are shown in the online appendix Table A15. They are very similar to our main results. Figures A11 and A12 also show the estimated effects in the sample restricted to those who live in Israel in 2018. We, therefore, conclude that GCP has no significant positive effects on earnings.

6.4 Does GCP Direct More Talent to the Economy's Knowledge-Producing Sectors?

Twenty-one percent of gifted children were employed in 2018 in high-tech services industries. This rate is 37 percent among men and 20 among women (Figure 14). These include telecommunications, computer and related services, and research and development. The effect on these rates is a small negative and not statistically different from zero.³²

Six percent of the women men are employed in high-tech manufacturing industries. These include pharmaceutical products, office machinery and computers, electronic components, electronic communication equipment, medical and scientific equipment, and aircraft. The effect on these rates is negative, and even though they are small in absolute magnitude, they are still statistically significant. The impact on women is a three-percentage-point decline relative to 6 percentage points in the comparison group. The effect on men is statistically insignificant.

When focusing on academic institutions where 4 percent of men and 8 percent of women are employed, we still find a (statistically insignificant) negative effect of GCP on this outcome among women and zero effect on men (Figure 14). Finally, estimating the impact on employment in the knowledge economy (in any of these sectors), we find a negative effect of 4p.p. relative to a baseline of 43 percent. The conclusion is that GCP does not enhance talented individuals' employment in the economy's knowledge-producing sectors. Note that this does not imply that gifted children do not contribute to the knowledge economy in Israel but that GCP does not affect the probability that gifted children will work in the knowledge economy.

6.5 Effect on Family Formation

³¹ Twelve percent of the gifted students in our main sample lived outside of Israel in 2018. The GCP effect on this outcome is 0.1 percentage points, and is statistically insignificant ($p=0.39$). Thus, analyzing the subset of students who lived in Israel is involves no selection bias.

³² Table A16 in the online appendix provides details on the relationship between studying for a STEM degree and working in the knowledge economy.

Figure 16 presents estimates on six personal outcomes based on data in 2016: married, married before age 30, married to GCP participant, having children, having a first child before age 30, and divorced. Sixty-five percent of women and 61 percent of men were married within the analysis period, but GCP did not affect these rates. However, GCP has a large positive effect on marrying a GCP participant, both for men and women. This effect is driven by matches within the program, and it increases the “quality” of the match, measured by the UPET score of the partner (see Table 7). The effect on marriages with the same GCP participants is fascinating in light of the recent work by Mogstad et al. (2021).³³

6.6 Analyzing Extended Samples

Next, we estimated the effects of GCP while varying the sample’s definition based on the age at which students took the UPET. The estimation with four different samples yields dozens of estimates. The overall riding and striking pattern emerging from these estimates are that the results regarding the impact of GCP on all outcomes are similar regardless of which sample we use. We, therefore, present the point estimates for only a few outcomes that we chose either because they had a statistically significant effect estimated in the baseline sample or because of their economic importance. These results are shown in Table 8 and Figures A14-A35 in the Online Appendix. The estimates based on separate boys’ and girls’ samples also yield the same results as those obtained from the benchmark sample.

6.7 Heterogeneity in Effect of GCP by Level of ‘Giftedness.’

We estimated a model where we allowed for heterogeneity of GCP impact by the level of Giftedness. We divided the sample into thirds for sample size considerations based on the UPET overall score distribution. We then estimated equation (2) while interacting the treatment variable (participation in GCP) with dummy indicators of the level of Giftedness. We also included the main effects of these three indicators. The estimated equation is:

$$(4) \quad Y_i = \alpha^l_i + \beta^l \cdot X_i + \tau_l \cdot T_i + \gamma_1 \cdot \text{Med}_i + \gamma_2 \cdot \text{Hi}_i + \delta_m \cdot \text{Med}_i \cdot T_i + \delta_h \cdot \text{Hi}_i \cdot T_i + \varepsilon^l_i$$

Where Med and Hi are indicator variables for the medium and the higher thirds of giftedness (between 672 and 716 in the UPET, and above 716). In this equation, τ_l represent the GCP effect on the lower third, and $\tau_l + \delta_m$ and $\tau_l + \delta_h$ represent the effects on the medium and higher thirds.

We pooled together in the baseline sample (including students from cohorts 1992-2005 who took UPET during 10th-11th grade), both males and females. Not allowing the estimates to vary by gender is not a limitation in this case because, as documented above, the impact (or lack of) of GCP is evident for boys and girls for almost all the outcomes.

³³ Figure A13, in the online appendix show the results for the sample restricted to those living in Israel in 2018.

We present these results in Table 9. We show results for outcomes for which we found a GCP significant average effect or based on their importance. There are some outcomes for which the estimated effect is evident for all three thirds. But for other outcomes, we see some important heterogeneity, either an effect only in the lowest third or some effect only in the upper two-thirds of Giftedness. For example, the estimated negative effect on the math *bagrut* score is large for the lower Giftedness third. In contrast, the estimated effect on the higher two-thirds is not statistically significant.³⁴ On the other hand, the negative effect on Hebrew *bagrut* test scores is large and significant only for the upper two-thirds.

Another substantial heterogeneous treatment effect is on the attainment of Ph.D. degrees that is evident only for the upper two-thirds of the UPET score distribution, with a very large effect on the middle third (3.4 percentage points increase relative to a baseline of 11%, implying a relative increase of about 30%).

More treatment effect heterogeneity is seen on the effects on choice of field of study. For example, the effect on math and computer sciences university degrees is large for the middle third, with an insignificant effect on the other two-thirds. In contrast, the positive estimated effect on physical sciences university degrees is evident for all thirds.

There is also heterogeneity in the effects on labor market outcomes. For example, the negative and significant treatment effect on employment in the tech manufacturing sector is significant only for the lowest two-thirds. At the same time, there is a statistically significant effect on employment in the tech services sector for the higher third. Interestingly, we find a statistically significant negative effect on income for the lower third, with insignificant effects for the higher two-thirds. This is consistent with the fact that the lower third of GCP participants had a high negative effect on math achievement and high school and had no positive effects on studying two STEM majors, or math and computer sciences, in the university.

The heterogeneity in treatment effect that we show in Table 8 implies that it might be very misleading to evaluate the contribution of GCP programs by using only marginal gifted children. As a result, we will be missing much of the most exciting impact of GCP on the life course outcomes of gifted children.

6.8 Heterogeneity in Effect of GCP by Students' SES

Another potentially important source of heterogeneity in GCP treatment effects is the participants' socio-economic status (SES). To explore that, we estimated separately the GCP effects on those with lower SES backgrounds, proxied by father education of fewer than 15 years (the

³⁴ Booij, Haan, and Plug (2017) conducted an experiment with third-grade gifted students and find that all participating students do better because of the program. Students near the admission cutoff experience a 0.2 standard deviation gain in their grade point average. Students further away from the admission cutoff experience larger gains.

minimal number of years required to attain a BA degree), and those with higher SES backgrounds (father education of 15 years or more). Again, we pulled boys and girls together.³⁵ We also stratified the sample by father’s income. The following table presents the means of these two SES variables for the treatment and comparison group below and above the median. The mean of father’s years of schooling among GCP participants above the median is 17.31, and below the median, it is 12.05, a huge difference of 5.26 years. The respective means in the comparison group are very similar. The table also shows the mother’s year of schooling, revealing equally significant differences between students above and below the median. The comparison based on parental income statistics reveals a similar pattern (with a little smaller difference of about 10-15%). These significant gaps in parental education and parental income typically lead to large and significant gaps in schooling outcomes of non-gifted children. However, innate talent makes these two SES indicators irrelevant for gifted children's outcomes. The question is whether there is an interaction effect between SES and the impact of GCP.

		Years of Education		UPET Score		
		Father	Mother	Quantitative	Verbal	English
GCP	Father education => 15	17.31	16.24	135.42	129.81	136.50
	<15	11.50	12.17	134.18	128.59	133.68
Comparison	=>15	16.96	15.96	136.31	130.07	136.60
	<15	12.05	12.70	133.13	127.65	132.93
Non-Gifted	=>15	16.81	15.68	120.14	111.31	122.34
	<15	11.85	12.58	114.52	106.74	115.09

Table 10 shows the results. Overall, GCP effects, particularly on high-school outcomes, do not vary by student’s SES background. For example, the estimated negative effect on math test scores is 0.09 and 0.06 for the lower and higher SES students (both estimates are statistically significant). The effects on university schooling and field of study choice are similar, too. Both groups significantly increase their enrolment in double major degrees, in a double major in STEM, and in physical sciences. Both groups also decrease their enrolment in engineering programs. The insensitivity of the estimated effects of GCP to SES variation is in sharp contrast to the effects of many other schooling inputs, which vary by student’s background.³⁶

³⁵ We also checked these heterogeneity estimates separately for boys and girls and found very similar patterns. We prefer presenting here the joint sample due to sample size limitations and implication for the power of estimating treatment effect by gender.

³⁶ For example, the effect of length of the school week and instructional time in each subject (Lavy 2015 and Lavy 2020a), and peer effects (Lavy, Silva, and Weinhardt, 2012 and Lavy, Paserman and Schlosser, 2012), class size (Angrist and Lavy 1999), Remedial Education (Lavy, Kott, and Rachkovski 2021).

Still, we find interesting differences between the effects on low and high SES GCP participants, mainly concerning long-term outcomes. The increase in Ph.D. degrees is driven entirely by the lower SES GCP participants, with an increase of almost 50%. While the impact on most of the labor market and personal outcomes is almost entirely statistically insignificant, there is heterogeneity in the effects on employment and working in the knowledge economy. We find a negative impact on the high SES participants with no effect on low SES participants.

7. Conclusions

Gifted children receive special attention in many educational systems. With the growth of the knowledge economy, governments are becoming aware that nurturing gifted students is crucial for securing new generations of scientists, creators, and innovators. Yet, the vast majority of published research on the impact of GCP has only examined their effects on short-run outcomes, primarily by looking at their impact on standardized test scores and educational attainment. While important, a possibly more profound question of interest to society is the effect of such interventions on long-run life outcomes. We address this important question using Israel's unique setting, offering both wide-scope GCP and rich administrative data to follow program participants over their life-cycle, from teenagerhood to adulthood, for some up to age 42.

We report several exciting and unique findings. First, no discernible effect of GCP on high school outcomes. If any, we find primarily adverse effects on test scores in high school high stake exit exams. Of particular interest is the negative impact on math test scores and the positive impact on history and literature test scores. This mixed pattern is perhaps consistent with the objectives of GCP, aiming primarily to widen the scope and diversity of the human capital of gifted children. These talented adolescents already have significant interest in STEM subjects and less interest in humanities. GCP intends to increase their curiosity and interest in non-science subjects.

Secondly, the large and significant effect on a double major, including mainly two STEM subjects (and the combination of STEM and non-STEM subjects), reveals perhaps the multipotentiality of gifted children and their difficulty selecting one area of interest which to focus. The focus on prestigious and highly regarded fields of study, such as math and physical sciences, is consistent with the view that gifted children are under social pressures by parents and social circles to 'maximize' their potential and not to 'waste' it on areas that are not too challenging intellectually. As a result, we should not be surprised by our findings of no effect on earnings at adulthood as the choice of the career path of gifted children is not necessarily guided by consideration of maximizing the financial return to their ability.

Perhaps surprising is the 'no' effect of GCP on integrating gifted children in work in sectors that produce 'new' knowledge. One explanation could be that gifted children are directed into these industries in advance. Thus, the GCP plays no important role in these decisions. However, it is also

important to note that our sectoral classification might be too coarse for capturing this potential effect. Therefore, we should instead use occupational categories within these sectors.

Against the benefit and gains accruing in gifted children's programs, we should note the potential loss to other students in the education system. Some evidence suggests that non-gifted children benefit from having high achievers and gifted children as peers (Lavy, Olmo, and Weinhardt, 2012, Balestra, Aurélien, and Stefan Wolterc. 2021). Thus, there is a concern that excluding gifted children from regular classes might negatively affect their peers.

8. References

- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10, 465-503.
- Abadie, A., & Imbens, G. 2002. Simple and bias-corrected matching estimators for average treatment effects. Working Paper 0283, National Bureau of Economic Research.
- Abadie, A., & Imbens, G. W. 2008. On the failure of the bootstrap for matching estimators. *Econometrica*, 76(6), 1537-1557.
- Angrist Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement." *Quarterly Journal of Economics*, May.
- Abdulkadiroglu, Atila, Joshua D. Angrist and Parag A. Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82(1): 137-196.
- Bailey JA 2nd. "Self-image, self-concept, and self-identity revisited." *J Natl Med Assoc.* 2003;95(5):383-386.
- Balestraa Simone, Aurélien Sallinb and Stefan C. Wolterc. 2021 "High-Ability Influencers? The Heterogeneous Effects of Gifted Classmates." Forthcoming, *Journal of Human Resources*.
- Betts, Julian, and Jaimie Shkolnik. 2000. "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools." *Economics of Education Review* 19, no. 1: 1-15.
- Bhatt, Rachana. "The Impacts of Gifted and Talented Education." SSRN Working Paper No. 09-11 (2009).
- Boettger, Eva Reid–Heiner, and Eva Reid. "Gifted education in various countries of Europe." *Journal of the Academy of Marketing Science* 33.3 (2005): 275-294.
- Beaujean, A.A.; Firmin, M.W.; Knoop, A.J.; Michonski, J.D.; Berry, T.P.; Lowrie, R.E. 2006. "Validation of the Frey and Detterman (2004) I.Q. prediction equations using the Reynolds Intellectual Assessment Scales." *Personal. Individ. Differ.* 41, 353–357.
- Bhatt, R. (2011). A review of gifted and talented education in the united states. *Education Finance and Policy*, 6(4):557–582.
- Bhatt, R. (2012). The impacts of gifted and talented education. Technical report, Georgia State University.

- Booij Adam S., Ferry Haan, Erik Plug. “Enriching Students Pays Off: Evidence from an Individualized Gifted and Talented Program in Secondary Education.” IZA DP No. 9757: February 2016.
- Booij Adam S., Ferry Haan, Erik Plug. “Can Gifted and Talented Education Raise the Academic Achievement of All High-Achieving Students?” IZA DP No. 10836 June 2017.
- Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does wage rank affect employees’ well-being?. *Industrial Relations: A Journal of Economy and Society*, 47(3), 355-389.
- Bui, S. A., Craig, S. G., and Imberman, S. A. (2014). Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy*, 6(3):30 – 62.
- Card, D. and Giuliano, L. (2014). Does gifted education work? For which students? Technical Report 20453, National Bureau of Economic Research.
- Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences*, 113(48), 13678-13683.
- Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6), 2981-3003.
- Colangelo, N., & Assouline, S.G. (2000). Counseling gifted students. In *International Handbook of Giftedness and Talent* (2nd). Heller, K. A., Monks, F. J., Sternberg, R. J., & Subotnik, R. F. (Eds.). New York, NY: Elsevier Applied Science Publishers/Elsevier Science Publishers.
- Davis, B., Engberg, J., Epple, D. N., Sieg, H., and Zimmer, R. (2010). Evaluating the gifted program of an urban school district using a modified regression discontinuity design. Working Paper 16414, National Bureau of Economic Research.
- Dai, D. Y., & Rinn, A. N. (2008). The big-fish-little-pond effect: What do we know and where do we go from here?. *Educational Psychology Review*, 20(3), 283-317.
- Davis, Billie, John Engberg, Dennis Epple, Holger Sieg, & Ron Zimmer. (2013). “Evaluating the Gifted Program of an Urban School District using a Modified Regression Discontinuity Design” *Annals of Economics and Statistics*, Vol.111/112, 2013.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101(5):1739-74.
- Elsner, B. & Isphording, I. E. (2017). A Big Fish in a Small Pond: Ability Rank and Human Capital Investment. *Journal of Labor Economics*, 35(3), 787–828.
- Epple, Dennis, Elizabeth Newton, and Richard Romano. “Ability Tracking, School Competition, and the Distribution of Educational Benefits.” *Journal of Public Economics* 83, no. 1 (2002): 1-48.
- Figlio, David, and Marianne Page. 2002. “School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?” *Journal of Urban Economics* 51, no. 3: 497-514.

- Fredrickson, R. (1982). A multipotential approach to career planning. In *Career information* (pp. 42–47). Englewood Cliffs, NJ: Prentice-Hall.
- Goodman, Joshua, Oded Gurantz and Jonathan Smith. 2020. “Take Two! SAT Retaking and College Enrolment Gaps.” *American Economic Journal: Economic Policy*, 12(2):115-58.
- Imbens Guido, 2004. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review”. *The Review of Economics and Statistics*, vol. 86, issue 1, 4-29.
- Herrmann J, Schmidt I, Kessels U, Preckel F. 2016. “Big fish in big ponds: Contrast and assimilation effects on math and verbal self-concepts of students in within-school gifted tracks.” *Br J Educ Psychol*. Jun; 86(2):222-40.
- Kerr, B. A., & Colangelo, N. (1988). The college plans of academically talented students. *Journal of Counselling & Development*. 67(1), 42–48.
- Kerr, B., and Erb, C. (1991). Career counseling with academically talented students. Effects of a value-based intervention. *Journal of Counseling Psychology*, 38(3), 309–314.
- Koenig, K.A.; Frey, M.C. 2008. Detterman, D.K. ACT and general cognitive ability.” *Intelligence*. 36, 153–160.
- Meredith C. Frey and Douglas K. Detterman. 2004. “Scholastic Assessment or g? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability.” *Psychological Science*, Volume 15 - Number 6: 373:378
- Lavy Victor and Analia Schlosser. 2011. “Mechanism and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics*, April.
- Lavy Victor, D. Paserman and A. Schlosser. 2012. “Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom”, *Economic Journal*, March.
- Lavy Victor, Olmo Silva and Felix Weinhardt, 2012. “The Good, The Bad and The Average: Evidence on Ability Peer Effects in Schools.” *Journal of Labor Economics*, April.
- Lavy Victor, 2015. “Do Differences in School’s Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries,” *Economic Journal*, November 125 (November), 397–424.
- Lavy Victor. 2020a. “Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior” *Journal of the European Economic Association*, February, 18(1):232–265.
- Lavy Victor. 2020b. “Teachers’ Pay for Performance in the Long-Run: The Dynamic Pattern of Treatment Effects on Students’ Educational and Labor Market Outcomes in Adulthood.” *The Review of Economic Studies*, Volume 87, October: 2322–2355.
- Lavy Victor. 2021. The Long-Term Consequences of Free School Choice.” *Journal of the European Economic Association*. Volume 19, Issue 3, June, Pages 1734–1781.

- Lavy Victor, Assaf Kott, and Genia Rachkovski. 2022. "Does Remedial Education at Late Childhood Pay Off After All? Long-Run Consequences for University Schooling, Labor Market Outcomes and Inter-Generational Mobility." *Journal of Labor Economics*, Volum 40, 1, pp. 239–282.
- Leung , S. A. , Conoley , C . W. , and Sfheel , M.J . 1994. The career and educational aspirations of gifted high school students: A retrospective study. *Journal of Counseling and Development*, 72, 298-303.
- Marsh, H. W. (2005). *Self-concept theory, measurement and research into practice: the role of self-concept in educational psychology*. Leicester, UK: Education Section of the British Psychological Society.
- Marsh, H., Seaton, M., Trautwein, U., Ludtke, O., Hau, O'Mara, A., & Craven, R. (2008). The Big Fish Little Pond Effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research. *Educational Psychology Review*, 20(3), 319–350.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366–380.
- Marsh, H. W., and Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213–231.
- Marsh H. and Rhonda G. Craven. (2002). "The Pivotal Role of Frames of Reference in Academic Self-Concept Formation: The "Big Fish-Little Pond" Effect." *Developmental Psychology*, January, Volume 2.
- Mogstad M., L. Kirkeboen, and E. Leuven, "College as a Marriage Market", March 2021, Draft.
- Monks, F.J. & Pfluger, R. (2005). *Gifted Education in 21 European Countries: Inventory and Perspective*. Radboud University Nijmegen. Retrieved from https://www.bmbf.de/pub/gifted_education_21_eu_countries.pdf
- Murphy, Richard and Felix. Weinhardt. 2020. "Top of The Class: The Importance of Ordinal Rank" *The Review of Economic Studies*, vol. 87(6), pp. 2777–2826.
- Pfeiffer, S. I. 2003. Psychological considerations in raising a healthy gifted child. In P. Olszewski-Kubilius, L. Limburg-Weber, & S. I. Pfeiffer (Eds.), *Early gifts: Recognizing and nurturing children's talents* (pp. 173–185). Waco, TX: Prufrock Press.
- Plucker, J. A., Robinson, N. M., Greenspon, T. S., Feldhusen, J. F., McCoach, D. B., & Subotnik, R. F. (2004). It's Not How the Pond Makes You Feel, but Rather How High You Can Jump. *American Psychologist*, 59(4), 268–269.
- Praekelt, F., Zeidner, M., Goetz, T., & Schleyer, E. (2008). Female big fish swimming against the tide: The BLUE and gender ratio in special gifted classes. *Contemporary Educational Psychology*, 33, 78-96.

- Preckel F, Götz T, Frenzel A. 2010. "Ability grouping of gifted students: effects on academic self-concept and boredom." *Br J. Educ Psychol.* Sep; 80 pp:451-72.
- Rinn, Anne N. "Effects of programmatic selectivity on the academic achievement, academic self-concepts, and aspirations of gifted college students." *Gifted Child Quarterly* 51.3 (2007): 232-245.
- Robinson, N. M., Lanzi, R. G., Weinberg, R. A., Ramey, S. L., & Ramey, C. T. (2002). Factors associated with high academic competence in former Head Start children at third grade. *Gifted Child Quarterly*, 46, 281–294.
- Rosenbaum Paul R. and Donald B. Rubin, 1983. "The central role of the propensity score in observational studies for causal effects", *Biometrika*, Volume 70, April 1983, Pages 41–55.
- Robins JM, Ritov Y. 1997. "Toward a curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models. *Stat Med.*" Jan 15-Feb 15;16(1-3):285-319.
- Sajjadi, Seyed Hossein, F. Gillian Rejsikind and Bruce M. Shore "Is Multipotentiality a Problem or Not? A New Look at the Data", June 2001, *High Ability Studies* 12(1):27-43.
- Sampson, James P. Jr. and Ashley K. Chason "Helping Gifted and Talented Adolescents and Young Adults: Psychoeducational Theory, Research, and Best Practices." In *Handbook of Giftedness in Children*, Edited by Steven I. Pfeiffer, 2008, Springer.
- Shani-Zinovich, I., (2007). Personal-Emotional Profile and 'Ego Identity Status among Gifted Adolescents." Ph.D. Thesis, University of Haifa, Faculty of Social Sciences, Department of Psychology.
- Shani-Zinovich, I., & Zeidner, M. (2009). 'On being a gifted adolescent: Developmental, affective and social issues." In R. Leikin, A. Berman, & B. Koichu (Eds.), *Creativity in mathematics and the education of gifted students* (pp. 195–220). Rotterdam, The Netherlands: Sense.
- Tymms, Peter. "A test of the big fish in a little pond hypothesis: An investigation into the feelings of seven-year-old pupils in school." *School Effectiveness and School Improvement* 12.2 (2001): 161-181.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004) "The relations between self-beliefs and academic achievement: A systematic review" *Educational Psychologist* (39), pp.111– 133.
- Vrignaud, Pierre, Denis Bonora, and Annie Dreux. "Counselling the gifted and talented in France: minimizing gift and maximizing talent." *International Journal for the advancement of counselling* 27.2 (2005): 211-228.
- Zeidner, Moshe, and Inbal Shani-Zinovich. (2013) ."Research on Personality and Affective Dispositions of Gifted Children: The Israeli Scene." *Gifted and Talented International*, 28:1-2, 35-50.
- Zeidner, Moshe, and Inbal Shani-Zinovich . A comparison of multiple facets of self-concept in gifted vs. non-identified Israeli students. *High ability studies*, 2015

- Zeidner, Moshe, Inbal Shani-Zinovich, Gerald Matthews, and Richard Roberts. 2005. "Assessing emotional intelligence in gifted and non-gifted high school students: outcomes depend on the measure." *Intelligence* 33 (4):369–391.
- Zeidner, Moshe, and E. J. Schleyer. "The Big-Fish-Little-Pond Effect for Academic Self-Concept, Test Anxiety, and School Grades in Gifted Children." *Contemp Educ Psychol.* 1999 Oct;24(4):305-329.

Table 1: Descriptive Statistics, Demographics and 8th Grade Tests Scores, 2006-2010 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Family Background:												
Father education	15.70	15.68	0.01	0.95	15.91	15.56	0.36	0.27	15.57	15.76	-0.19	0.47
Mother education	15.63	15.65	-0.02	0.91	15.71	15.55	0.15	0.63	15.58	15.71	-0.13	0.59
Born in Israel	0.86	0.86	0.01	0.83	0.83	0.89	-0.06	0.12	0.88	0.84	0.04	0.14
Father Born in Israel	0.63	0.63	0.01	0.85	0.60	0.64	-0.03	0.51	0.65	0.62	0.03	0.44
Mother born Israel	0.67	0.68	-0.01	0.74	0.69	0.67	0.01	0.83	0.65	0.68	-0.02	0.54
Father income 2003	1.70	1.86	-0.16	0.26	1.75	1.94	-0.19	0.48	1.68	1.82	-0.14	0.37
Mother income 2003	0.92	0.98	-0.06	0.40	0.80	1.01	-0.21	0.05*	0.99	0.95	0.04	0.69
Siblings	1.64	1.67	-0.03	0.74	1.84	1.84	0.00	1.00	1.53	1.57	-0.05	0.61
B. 8th Grade Test Scores:												
Math	1.21	1.20	0.01	0.71	1.13	1.18	-0.05	0.31	1.26	1.20	0.05	0.19
Science	1.04	1.03	0.01	0.67	1.01	0.96	0.05	0.23	1.06	1.07	-0.01	0.70
Hebrew	0.97	0.94	0.03	0.35	1.11	0.99	0.12	0.01*	0.88	0.90	-0.02	0.60
English	0.91	0.88	0.04	0.09*	0.89	0.91	-0.02	0.52	0.93	0.86	0.07	0.00*
Number of Observations	477	477			178	178			299	299		

Notes: This table presents descriptive statistics for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. Income is measured in 100K NIS. The 8th-grade test score is normalized (z score). The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. * represents statistical significance at the 90% level.

Table 2: Distribution of *Bagrut* Subjects at Advanced Level, 2006-2010 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Math	0.81	0.77	0.04	0.12	0.73	0.69	0.03	0.48	0.86	0.82	0.04	0.14
Physics	0.52	0.49	0.03	0.32	0.31	0.27	0.04	0.42	0.64	0.62	0.03	0.50
Computer Sciences	0.48	0.47	0.01	0.66	0.34	0.24	0.10	0.04*	0.57	0.61	-0.04	0.37
Chemistry	0.29	0.28	0.00	0.95	0.32	0.30	0.01	0.81	0.27	0.27	0.00	0.93
Biology	0.19	0.18	0.01	0.84	0.31	0.24	0.07	0.12	0.12	0.15	-0.04	0.19
Arabic	0.11	0.13	-0.02	0.37	0.13	0.16	-0.03	0.45	0.10	0.12	-0.01	0.62
Literature	0.10	0.10	0.00	0.92	0.14	0.19	-0.06	0.15	0.08	0.05	0.03	0.13
Social Sciences	0.11	0.10	0.01	0.48	0.16	0.18	-0.02	0.57	0.09	0.05	0.04	0.08*
History	0.04	0.03	0.01	0.36	0.05	0.02	0.03	0.07*	0.03	0.04	0.00	0.79
Art	0.08	0.09	-0.01	0.74	0.12	0.16	-0.03	0.37	0.05	0.04	0.01	0.54
Total credits	30.00	29.64	0.36	0.13	29.57	28.91	0.66	0.09*	30.26	30.08	0.19	0.53
Number of Observations	477	477			178	178			299	299		

Notes: This table presents the proportions of advance level bagrut study subjects in the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. Total credits refer to the total credits accumulated in the study program. All other variables are indicator variables for five credits bagrut in the subject. The sample includes only students who participated in the Meitzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. *represents statistical significance at the 90% level.

Table 3: Bagrut Test Scores in Compulsory Subjects, 2006-2010 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean composite score	89.51	88.86	0.66	0.10	90.33	89.03	1.30	0.03*	89.03	88.76	0.27	0.61
English	93.45	90.90	2.55	0.00*	93.38	89.88	3.51	0.00*	93.49	91.51	1.98	0.00*
Math	86.79	87.50	-0.72	0.44	86.05	86.82	-0.78	0.62	87.23	87.91	-0.68	0.55
Bible	88.15	88.28	-0.13	0.81	89.85	89.50	0.35	0.66	87.14	87.55	-0.42	0.56
Literature	83.98	83.32	0.66	0.32	88.04	85.99	2.05	0.04*	81.56	81.73	-0.17	0.83
History	86.13	86.18	-0.05	0.93	87.32	86.92	0.40	0.68	85.42	85.74	-0.32	0.67
Civil Studies	85.06	85.66	-0.60	0.39	86.57	86.60	-0.02	0.98	84.16	85.10	-0.95	0.26
Hebrew	87.97	87.08	0.89	0.07*	89.85	87.86	1.99	0.00*	86.85	86.62	0.22	0.74
Number of Observations	477	477			178	178			299	299		

Notes: This table presents the mean test scores of compulsory *bagrut* subjects for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. All scores are on a scale of 0-100. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. * represents statistical significance at the 90% level.

Table 4: University Degrees Attainment, and Field of Study, 2006-2010 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Degrees:												
BA	0.94	0.87	0.07	0.00*	0.94	0.90	0.04	0.15	0.94	0.85	0.09	0.00*
BA, Double Major	0.60	0.35	0.26	0.00*	0.52	0.35	0.17	0.00*	0.66	0.34	0.31	0.00*
MA	0.22	0.19	0.03	0.26	0.19	0.19	-0.01	0.89	0.24	0.19	0.05	0.14
PHD	0.02	0.00	0.02	0.02*	0.01	0.01	0.00	1.00	0.03	0.00	0.03	0.01*
B. Field of Study:												
STEM	0.64	0.51	0.12	0.00*	0.47	0.33	0.14	0.01*	0.74	0.62	0.11	0.00*
Math, Computer Sciences, Statistics	0.34	0.19	0.15	0.00*	0.16	0.08	0.08	0.02*	0.45	0.26	0.19	0.00*
Engineering	0.19	0.27	-0.08	0.00*	0.14	0.18	-0.05	0.25	0.22	0.32	-0.10	0.01*
Physical Sciences	0.24	0.12	0.12	0.00*	0.14	0.04	0.10	0.00*	0.30	0.17	0.13	0.00*
Biological Sciences	0.20	0.11	0.09	0.00*	0.18	0.10	0.08	0.02*	0.21	0.11	0.09	0.00*
Business and Management	0.05	0.07	-0.02	0.24	0.07	0.11	-0.03	0.26	0.04	0.05	-0.01	0.56
Social Sciences	0.18	0.18	0.01	0.81	0.24	0.27	-0.03	0.46	0.15	0.12	0.03	0.28
Humanities	0.13	0.08	0.05	0.01*	0.18	0.10	0.08	0.02*	0.11	0.07	0.04	0.11
Law	0.06	0.05	0.01	0.64	0.07	0.10	-0.03	0.35	0.05	0.02	0.03	0.07*
Medicine	0.03	0.03	0.00	0.86	0.05	0.03	0.02	0.27	0.02	0.03	-0.01	0.40
Number of Observations	477	477			178	178			299	299		

Notes: This table presents means of university degree attainment and field of study for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. Therefore, * represents statistical significance at the 90% level.

Table 5: GCP Treatment Effect Estimates, 2006-2013 Samples, by Matching Model

Panel A.		Bagrut Exam Scores						
Outcome:	Mean Score	Math	Hebrew	Bible	English	History	Literature	Civil Studies
Matching on:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I, 8th Grade Test Scores (2006-2010) N: 510 (510)	0.21 (0.37)	-1.50* (0.86)	0.17 (0.38)	-0.74 (0.55)	1.77* (0.33)	-0.05 (0.52)	-0.37 (0.60)	-1.83* (0.65)
II, UPET Scores (2006-2010) N: 409 (409)	-1.03* (0.33)	-3.52* (0.90)	-0.56 (0.37)	-1.02* (0.58)	0.17 (0.26)	-0.32 (0.53)	0.24 (0.62)	-2.49* (0.64)
III, 5th Grade Test Scores (2009-2013) N: 483 (483)	0.01 (0.40)	-1.48* (0.89)	0.58 (0.48)	-2.06* (0.90)	1.00* (0.36)	-0.64 (0.58)	0.00 (0.78)	-1.65* (0.67)
IIII, 5th&8th Grade Test Scores (2006-2013) N: 999 (999)	0.18 (0.26)	-1.04* (0.59)	0.03 (0.28)	-1.38* (0.45)	1.56* (0.24)	-0.14 (0.38)	0.51 (0.48)	-1.88* (0.47)

Panel B.		Higher Education						
Outcome:	BA	MA	BA Double Major	Double Major in STEM	Double Major, STEM & Other	Math, CS, Stats	Engineering	Physical sciences
Matching on:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I, 8th Grade Test Scores (2006-2010) N: 510 (510)	0.03* (0.02)	-0.01 (0.03)	0.27* (0.04)	0.20* (0.03)	0.05* (0.03)	0.13* (0.03)	-0.12* (0.03)	0.12* (0.03)
II, UPET Scores (2006-2010) N: 483 (483)	0.02 (0.02)	0.01 (0.03)	0.17* (0.04)	0.13* (0.03)	0.05* (0.03)	0.08* (0.03)	-0.09* (0.03)	0.07* (0.03)
III, 5th Grade Test Scores (2009-2013) N: 409 (409)	0.12* (0.04)	0.01 (0.02)	0.20* (0.04)	0.14* (0.03)	0.02 (0.02)	0.15* (0.03)	-0.09* (0.04)	0.10* (0.03)
IIII, 5th&8th Grade Test Scores (2006-2013) N: 999 (999)	0.09* (0.02)	0.03 (0.02)	0.20* (0.03)	0.17* (0.02)	0.04* (0.02)	0.14* (0.02)	-0.08* (0.02)	0.12* (0.02)

Notes: This table presents the treatment effect estimated by the matching model. Each row represents a different specification, with another type of exam used as a proxy for ability (the cohorts included in each sample are also shown in parentheses). Each column represents another outcome variable. * represents statistical significance at the 90% level.

Table 6.A: Treatment Estimated Effect Heterogeneity, by Participation in a GCP during Middle-School, 2006-2010 Sample, Matching I (by 8th Grade Test Scores)

Panel A. Outcome:	Mean	Math	Hebrew	Bible	English	History	Literature	Civil Studies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Only in High-School								
Comparison Group Mean	88.257	87.086	86.510	88.358	90.319	85.716	83.346	85.868
GCP Effect	0.675 (0.497)	-1.793 (1.170)	0.602 (0.521)	-1.225* (0.724)	1.800* (0.465)	-0.144 (0.689)	0.272 (0.872)	-2.517* (0.915)
Number of Observations	257 (257)							
Since Middle School								
Comparison Group Mean	89.561	87.986	87.755	88.186	91.573	86.718	83.286	85.418
GCP Effect	-0.318 (0.542)	-1.120 (1.361)	-0.317 (0.569)	-0.150 (0.803)	1.731 (0.507)	0.069 (0.766)	-1.036 (0.948)	-0.963 (1.007)
Number of Observations	220 (220)							
Difference	-0.993 (0.741)	0.673 (1.833)	-0.919 (0.767)	1.075 (1.116)	-0.069 (0.655)	0.213 (1.075)	-1.308 (1.336)	1.554 (1.451)
Panel B. Outcome:	BA	MA	BA Double Major	Double Major in STEM	Double Major, STEM & Other	Math, CS, Stats	Engineering	Physical sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Only in High-School								
Comparison Group Mean	0.868	0.202	0.311	0.113	0.082	0.144	0.257	0.082
GCP Effect	0.028 (0.021)	-0.056 (0.041)	0.237* (0.049)	0.182* (0.040)	0.019 (0.034)	0.117* (0.040)	-0.125* (0.042)	0.126* (0.037)
Number of Observations	257 (257)							
Since Middle School								
Comparison Group Mean	0.873	0.182	0.386	0.209	0.073	0.250	0.286	0.173
GCP Effect	0.028 (0.023)	0.055 (0.044)	0.301* (0.053)	0.227* (0.045)	0.074* (0.039)	0.144* (0.045)	-0.104* (0.048)	0.113* (0.042)
Number of Observations	220 (220)							
Difference	0.000 (0.033)	0.111 (0.062)	0.064 (0.072)	0.045 (0.063)	0.055 (0.054)	0.027 (0.063)	0.021 (0.065)	-0.013 (0.061)

Notes : This table presents the treatment effect heterogeneity by the length of participation in the GCP. Each column represents a different outcome variable. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. The effects are calculated based on the estimation of equation three described in the text (Section 5.4).* represents statistical significance at the 90% level.

**Table 6.B: Treatment Estimated Effect Heterogeneity, by Participation in a GCP during Middle-School, 2009-2013
Sample, Matching III (by 5th Grade Test Scores)**

Panel A. Outcome:	Mean	Math	Hebrew	Bible	English	History	Literature	Civil Studies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GCP During High-School								
Comparison Group Mean	88.324	86.683	85.239	88.790	91.062	85.066	83.465	86.498
GCP Effect	0.323 (0.598)	-0.964 (1.432)	1.096 (0.724)	-3.119* (1.148)	1.764* (0.546)	-0.760 (0.864)	0.167 (1.115)	-2.525* (1.131)
Number of Observations	240 (240)							
GCP During Middle and High School								
Comparison Group Mean	90.116	89.075	87.362	88.929	92.742	86.862	83.983	86.996
GCP Effect	-0.249 (0.557)	-1.856 (1.245)	0.128 (0.681)	-1.040 (1.355)	0.299 (0.513)	-0.464 (0.822)	-0.281 (1.137)	-0.661 (0.998)
Number of Observations	243 (243)							
Difference	-0.572 (0.811)	-0.892 (1.861)	-0.968 (0.988)	2.079 (1.913)	-1.465* (0.702)	0.296 (1.200)	-0.448 (1.758)	1.864 (1.786)
Panel B. Outcome:	BA	MA	BA Double Major	Double Major in STEM	Double Major, STEM & Other	Math, CS, Stats	Engineering	Physical sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GCP During High-School								
Comparison Group Mean	0.490	0.025	0.148	0.086	0.021	0.107	0.173	0.074
GCP Effect	0.101* (0.052)	0.019 (0.027)	0.144* (0.058)	0.129* (0.050)	-0.031 (0.023)	0.113* (0.048)	-0.115* (0.052)	0.106* (0.047)
Number of Observations	240 (240)							
GCP During Middle and High School								
Comparison Group Mean	0.512	0.042	0.167	0.108	0.017	0.138	0.171	0.092
GCP Effect	0.138* (0.051)	-0.007 (0.027)	0.243* (0.054)	0.156* (0.047)	0.072* (0.027)	0.187* (0.047)	-0.067 (0.048)	0.089* (0.043)
Number of Observations	243 (243)							
Difference	0.037 (0.073)	-0.026 (0.038)	0.099 (0.079)	0.027 (0.070)	0.103* (0.044)	0.074 (0.071)	0.048 (0.070)	-0.017 (0.063)

Notes : This table presents the treatment effect heterogeneity by the length of participation in the GCP. Each column represents a different outcome variable. The sample includes only students who participated in the Metzav primary school test during their 5th grade, about half of the students in cohorts of high-school graduates in 2009-2013. The effects are calculated based on the estimation of equation three described in the text (Section 5.4).* represents statistical significance at the 90% level.

Table 7: Descriptive Statistics, Labor Market and Personal Outcomes, 1992-2005 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Labor:												
Income Salaried Worker	2.19	2.22	-0.03	0.68	1.65	1.70	-0.05	0.62	2.47	2.50	-0.03	0.77
Income Salaried Worker (employed=1)	2.93	2.95	-0.02	0.79	2.16	2.15	0.01	0.93	3.35	3.41	-0.06	0.59
Income (employed=1)	3.09	3.10	-0.02	0.86	2.31	2.26	0.05	0.66	3.52	3.59	-0.07	0.52
Income (employed=1), with outliers	3.34	3.24	0.10	0.51	2.35	2.26	0.09	0.46	3.88	3.79	0.09	0.69
Self employed	0.09	0.09	0.00	0.72	0.10	0.10	0.00	1.00	0.09	0.08	0.01	0.40
Employed	0.75	0.75	-0.01	0.58	0.76	0.78	-0.02	0.41	0.74	0.74	0.00	1.00
B. Employment Sectors:												
Knowledge economy	0.39	0.43	-0.04	0.02*	0.28	0.34	-0.06	0.02*	0.45	0.47	-0.02	0.34
High tech services	0.29	0.31	-0.02	0.21	0.18	0.20	-0.02	0.38	0.35	0.37	-0.02	0.32
High tech manufacturing	0.04	0.06	-0.02	0.02*	0.03	0.06	-0.03	0.01*	0.05	0.06	-0.01	0.30
Academic	0.05	0.06	0.00	0.94	0.07	0.08	-0.01	0.51	0.05	0.04	0.01	0.25
C. Personal:												
Married	0.60	0.61	-0.01	0.70	0.65	0.67	-0.02	0.46	0.57	0.57	0.00	1.00
Married Before Age 30	0.39	0.40	-0.01	0.63	0.53	0.54	-0.01	0.73	0.32	0.33	-0.01	0.61
Divorced	0.02	0.02	0.00	0.90	0.03	0.03	0.00	1.00	0.02	0.02	0.00	1.00
Number of children	1.24	1.23	0.01	0.77	1.45	1.43	0.02	0.78	1.13	1.12	0.01	0.85
First Child Before 30	0.28	0.27	0.01	0.55	0.39	0.40	-0.01	0.72	0.22	0.21	0.01	0.56
Age at marriage	29.62	29.67	-0.05	0.74	28.45	28.62	-0.17	0.42	30.34	30.33	0.01	0.95
Age at first child	31.01	30.94	0.07	0.62	30.19	30.15	0.04	0.85	31.54	31.47	0.07	0.69
GCP Partner	0.10	0.02	0.07	0.00*	0.13	0.04	0.09	0.00*	0.07	0.01	0.06	0.00*
Same School Partner	0.11	0.10	0.01	0.47	0.13	0.09	0.04	0.10	0.10	0.10	0.00	1.00
Same Class Partner	0.05	0.00	0.05	0.00*	0.07	0.00	0.07	0.00*	0.04	0.00	0.04	0.00*
Partner's UPET Score	626.55	618.23	8.32	0.06*	650.95	637.30	13.65	0.05*	613.11	607.51	5.60	0.30
Partner's Income 2018	1.01	1.10	-0.97	0.16	1.57	1.78	-0.21	0.18	0.71	0.75	-0.03	0.55
Number of Observations	1,735	1,735			602	602			1,133	1,133		

Notes: This table presents means of labor market and personal outcomes for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. Income is measured in 100K NIS. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. * represents statistical significance at the 90% level.

Table 8: GCP Treatment Effect Estimates, 1992-2005 Cohorts, by Sample

Panel A.											
Outcome:	Bagrut Exam Scores						Higher Education				
	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	MA	PHD	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Main	-0.04*	-0.07*	-0.08*	-0.09*	0.02	0.03*		0.06*	0.04*	-0.05*	0.06*
N: 1,735 (1,735)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Extend 12	-0.03*	-0.09*	-0.07*	-0.09*	0.01	0.03*	0.10*	0.06*	0.06*	-0.07*	0.07*
N: 2,128 (2,128)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Extend all	-0.02*	-0.11*	-0.06*	-0.10*	0.02	0.03*	0.11*	0.04*	0.05*	-0.07*	0.07*
N: 2,920 (2,920)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Main, larger cities	-0.04*	-0.11*	-0.09*	-0.10*	0.01	0.03*	0.09*	0.04*	0.05*	-0.07*	0.05*
N: 1,717 (1,717)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)

Panel B.											
Outcome:	Labor market outcomes									Personal	
	Employed 2018	Self Employed 2018	Salaried Income 2018	Salaried Income (employed=1) 2018	Income (employed=1) 2018	HT services 2018	HT manufacturing 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Main	-0.01	0.00	-0.05	-0.07	-0.06	-0.02	-0.02*	-0.04*	0.00	-0.01	0.00
N: 1,735 (1,735)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Extend 12	-0.03*	0.00	-0.08	-0.03	-0.02	-0.02	-0.02*	-0.04*	-0.01	0.00	0.00
N: 2,128 (2,128)	(0.01)	(0.01)	(0.07)	(0.07)	(0.07)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Extend all	-0.03*	0.00	-0.10*	-0.04	-0.04	-0.01	-0.02*	-0.03*	0.00	-0.02	-0.01
N: 2,920 (2,920)	(0.01)	(0.01)	(0.06)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Main, larger cities	-0.01	0.01	-0.08	-0.10	-0.09	-0.02	-0.03*	-0.05*	-0.01	0.00	0.02
N: 1,717 (1,717)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	0.02	(0.01)	(-0.02)	(0.01)

Notes: This table presents the treatment effect estimated on girls by sample. Each row represents a different sample, and each column represents a different outcome variable. Main denotes the main sample of 1992-2005 graduates, described in text. Extend12 is adding to the main sample UPET 12th grade test takers. Extend all is adding to the main sample UPET test takers at all ages. Main, larger cities is restricting the comparison group to include only students from larger cities (with 200 or more students in the cohort).
 * represents statistical significance at the 90% level.

Table 9: Treatment Estimated Effect Heterogeneity, by Lower, Middle, and Upper Thirds of Giftedness Levels, 1992-2005 Sample

Panel A. Outcome:	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	MA	PHD	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Lower:											
Comparison Group Mean	5.121	5.109	5.093	5.035	0.462	0.075	0.495	0.189	0.203	0.278	0.113
GCP Effect	-0.037 (0.029)	-0.186* (0.063)	-0.032 (0.034)	-0.080* (0.048)	0.016 (0.029)	0.013 (0.016)	0.091* (0.030)	0.033 (0.023)	0.001 (0.023)	-0.054* (0.024)	0.057* (0.021)
Middle:											
Comparison Group Mean	5.480	5.468	5.442	5.367	0.553	0.112	0.527	0.255	0.253	0.321	0.155
GCP Effect	-0.058* (0.025)	-0.048 (0.050)	-0.109* (0.034)	-0.160* (0.044)	0.031 (0.029)	0.034* (0.019)	0.126* (0.029)	0.083* (0.026)	0.092* (0.026)	-0.046* (0.025)	0.084* (0.023)
Upper:											
Comparison Group Mean	5.681	5.757	5.624	5.565	0.619	0.190	0.630	0.403	0.428	0.299	0.309
GCP Effect	-0.014 (0.019)	0.011 (0.036)	-0.104* (0.030)	-0.032 (0.038)	0.023 (0.028)	0.038 (0.024)	0.080* (0.027)	0.064* (0.027)	0.025 (0.028)	-0.042* (0.026)	0.043* (0.026)
Panel B. Outcome:											
	Employed 2018	Self Employed 2018	Salaried Worker Income 2018	Salaried Worker Income (employed=1) 2018	Income (employed=1) 2018	High-tech services 2018	High-tech manufacturing 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Lower:											
Comparison Group Mean	0.799	0.089	2.071	2.576	2.690	0.262	0.075	0.377	0.040	0.323	0.287
GCP Effect	-0.016 (0.024)	0.004 (0.017)	-0.220* (0.111)	-0.231* (0.115)	-0.186* (0.111)	-0.022 (0.025)	-0.024* (0.014)	-0.053* (0.027)	0.004 (0.017)	-0.010 (0.027)	0.019 (0.027)
Middle:											
Comparison Group Mean	0.750	0.105	2.208	2.936	3.114	0.304	0.070	0.431	0.058	0.258	0.255
GCP Effect	0.008 (0.025)	-0.005 (0.018)	0.160 (0.133)	0.131 (0.142)	0.138 (0.141)	0.002 (0.027)	-0.026* (0.013)	-0.025 (0.029)	-0.005 (0.018)	0.035 (0.026)	0.036 (0.026)
Upper:											
Comparison Group Mean	0.713	0.079	2.383	3.389	3.558	0.372	0.038	0.478	0.068	0.236	0.186
GCP Effect	-0.014 (0.027)	0.013 (0.016)	-0.084 (0.144)	-0.108 (0.153)	-0.120 (0.150)	-0.050* (0.027)	-0.001 (0.012)	-0.046 (0.029)	0.013 (0.016)	-0.011 (0.025)	0.012 (0.024)
Number of Observations	3,468										

Notes: This table presents the treatment effect heterogeneity by gifted levels. Gifted levels are "Lower", "Middle", and "Higher". Lower includes all gifted students with a psychometric score below 671 (569 GCP participants and 577 comparison students). Middle includes all gifted students with a psychometric score between 672 and 716 (580 and 576 students). Higher includes all gifted students with a psychometric score above 716 (582 and 579 students). Each column represents a different outcome variable. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. The effects are calculated based on the estimates heterogeneity estimation equation described in the main text (section 6.7). * represents statistical significance at the 90% level.

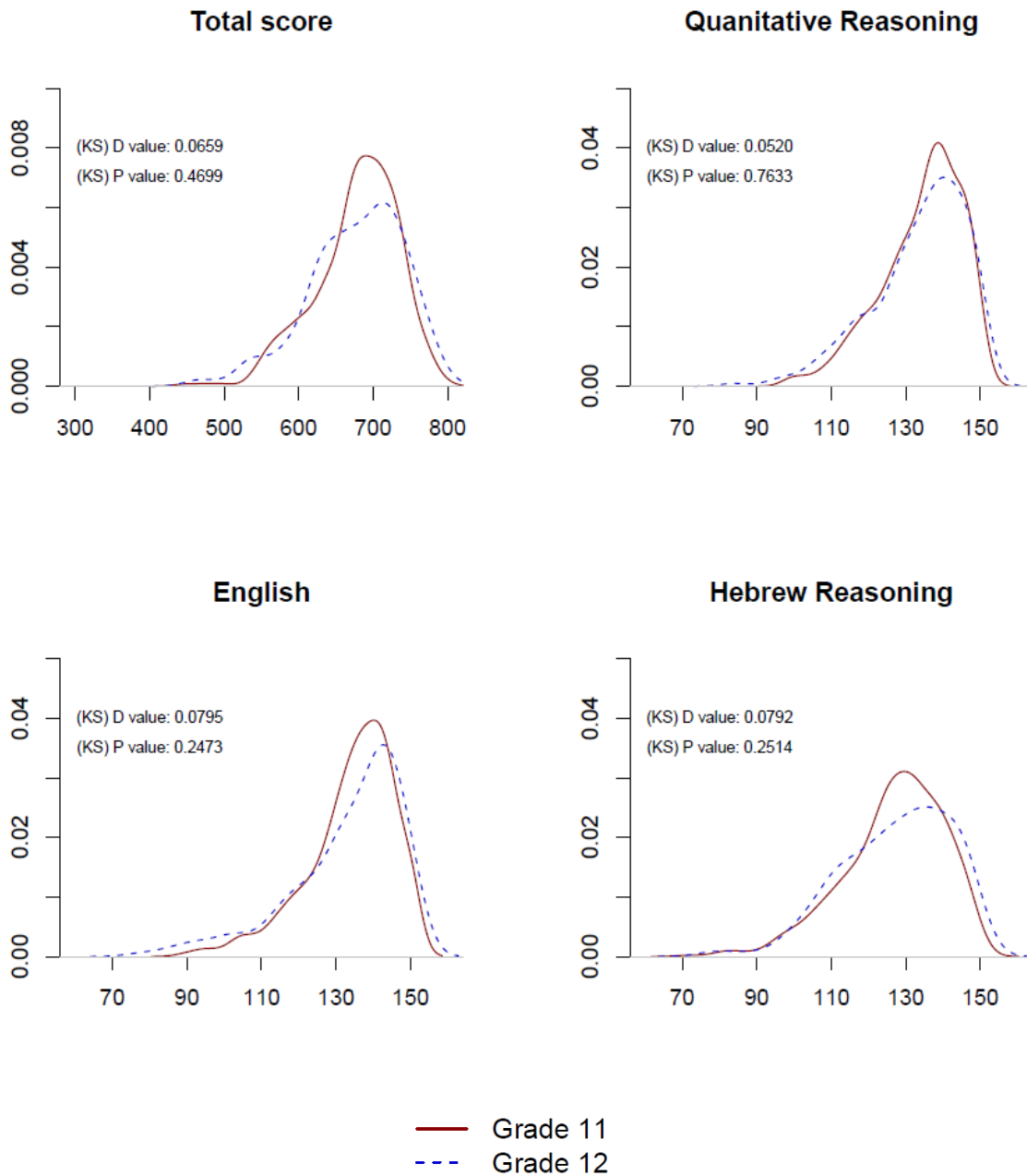
Table 10: GCP Treatment Estimated Effect Heterogeneity, by Socio-Economic Background, 1992-2005 Sample

Panel A. Outcome:	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	MA	PHD	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Father Education < 15y											
Comparison Group Mean	5.35	5.37	5.32	5.25	0.52	0.11	0.55	0.54	0.27	0.29	0.18
GCP Effect	-0.031 (0.022)	-0.092* (0.046)	-0.061* (0.029)	-0.068* (0.038)	0.024 (0.025)	0.048* (0.017)	0.102* (0.025)	0.110* (0.025)	0.025 (0.022)	-0.042* (0.022)	0.042* (0.020)
Number of Observations	806 (815)										
Father Education ≥ 15y											
Comparison Group Mean	5.50	5.62	5.45	5.39	0.57	0.14	0.55	0.56	0.31	0.31	0.20
GCP Effect	-0.045* (0.019)	-0.060* (0.039)	-0.094* (0.027)	-0.118* (0.035)	0.016 (0.023)	0.009 (0.017)	0.091* (0.023)	0.084* (0.023)	0.044* (0.021)	-0.052* (0.021)	0.067* (0.020)
Number of Observations	929 (920)										
Panel B. Outcome:	Employed 2018	Self Employed 2018	Salaried Worker Income 2018	Salaried Worker Income (employed=1) 2018	Income (employed=1) 2018	High-tech services 2018	High-tech manufacturing 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Father Education < 15y											
Comparison Group Mean	0.750	0.090	2.140	2.840	3.000	0.300	0.050	0.400	0.050	0.410	0.280
GCP Effect	0.025 (0.022)	0.012 (0.015)	0.052 (0.112)	-0.035 (0.117)	-0.010 (0.115)	-0.008 (0.023)	-0.001 (0.011)	-0.007 (0.025)	0.001 (0.011)	-0.009 (0.025)	0.036 (0.023)
Number of Observations	806 (815)										
Father Education ≥ 15y											
Comparison Group Mean	0.760	0.090	2.300	3.050	3.190	0.320	0.070	0.450	0.060	0.400	0.260
GCP Effect	-0.036* (0.021)	-0.005 (0.013)	-0.122 (0.110)	-0.083 (0.116)	-0.080 (0.114)	-0.033 (0.022)	-0.034* (0.011)	-0.067* (0.023)	0.000 (0.012)	-0.014 (0.023)	-0.019 (0.021)
Number of Observations	929 (920)										

Notes: This table presents the treatment effect heterogeneity, by socioeconomic status, proxies by father education. Each column represents a different outcome variable. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. The effects are calculated based on the estimates of the main estimation equation described in the text on each group separately. * represents statistical significance at the 90% level.

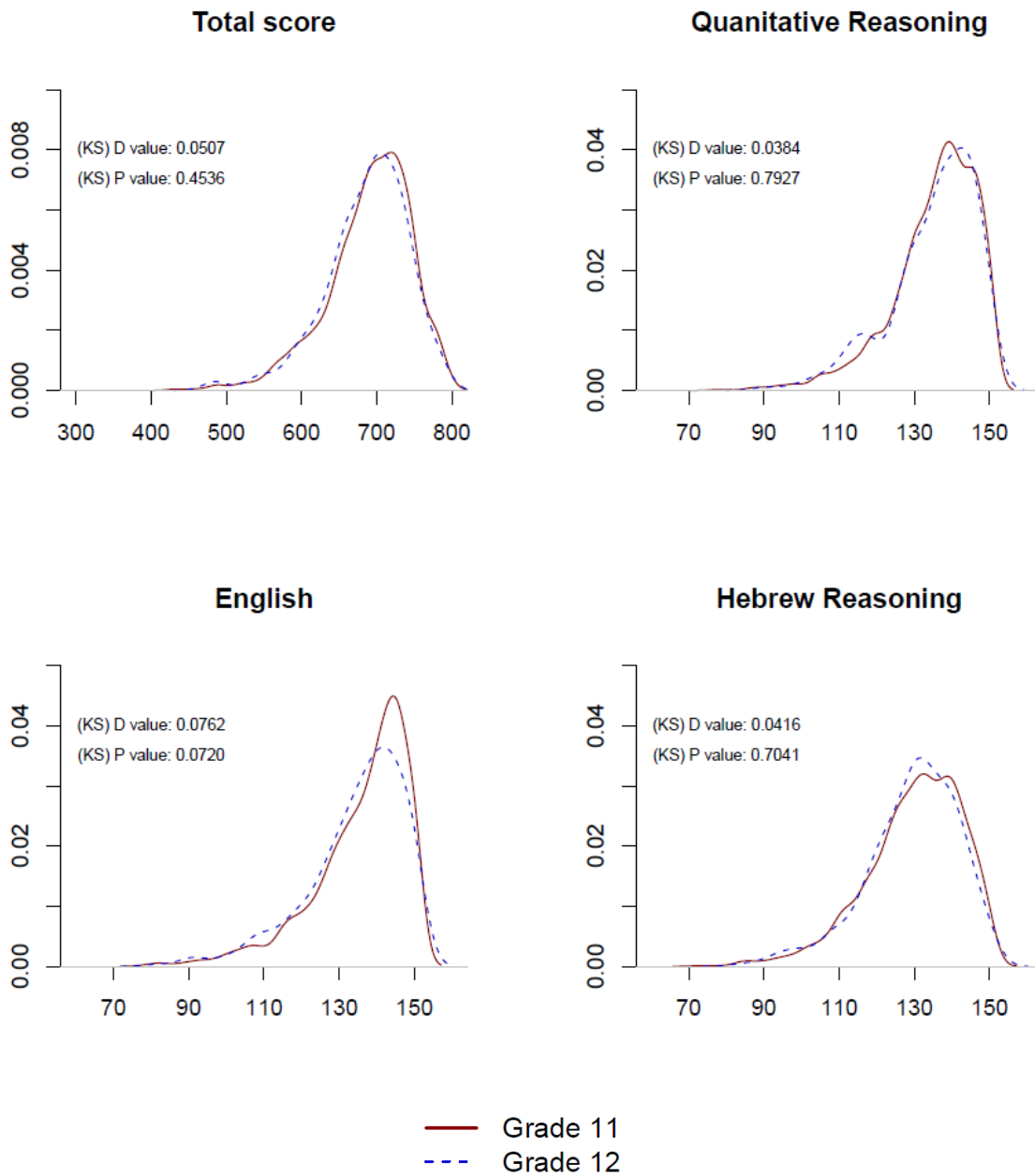
Figure 1: UPET Scores Distributions, by Grade of Testing

(A) 2006-2010 Sample



Notes: This figure plots the distribution of UPET scores by grade of taking the test- the solid red line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010.

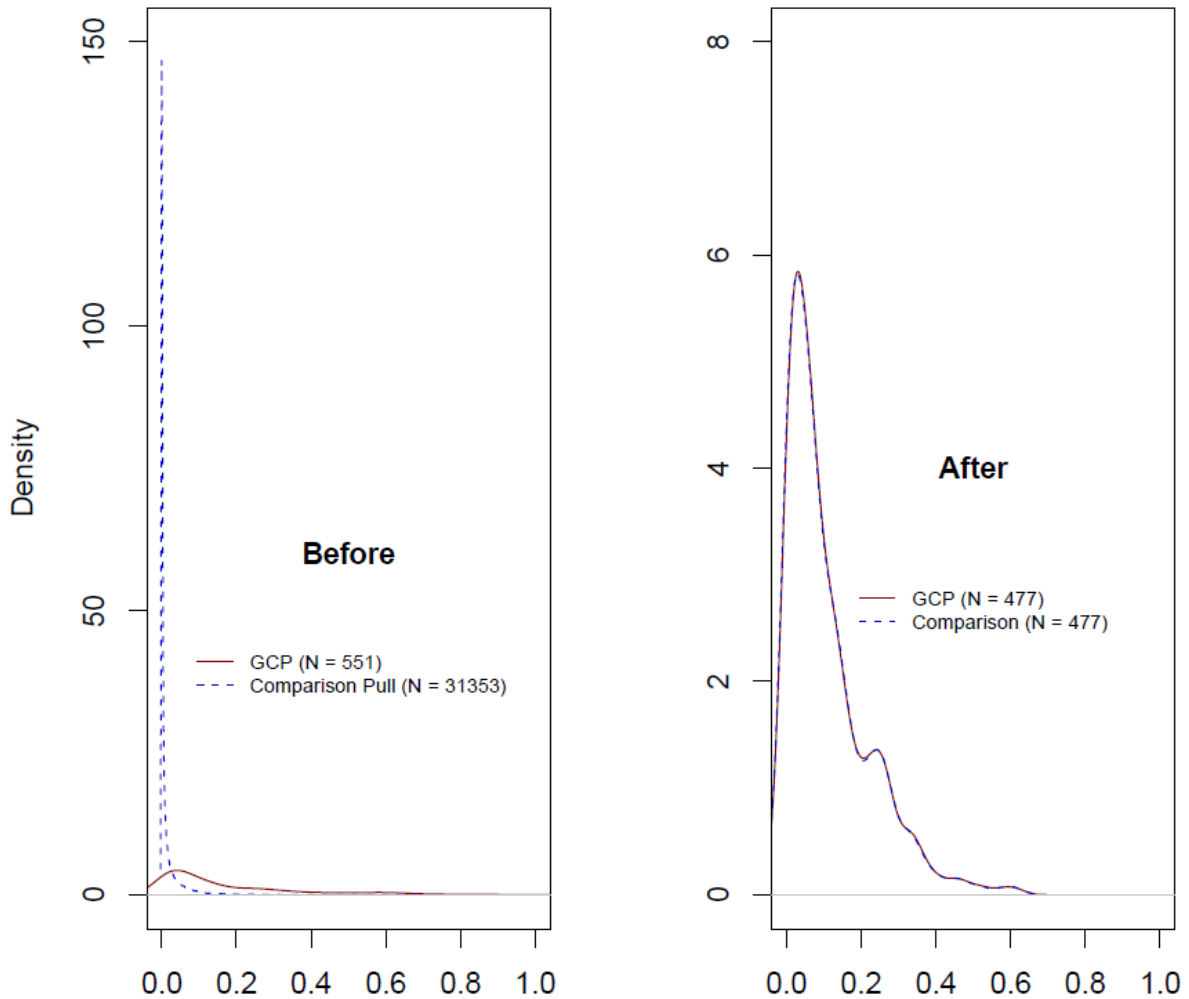
(B) 1992-2005 Sample



Notes: This figure plots the distribution of UPET scores by grade of taking the test- the solid red line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only GCP participants from the cohorts of high-school graduates in 1992-2005.

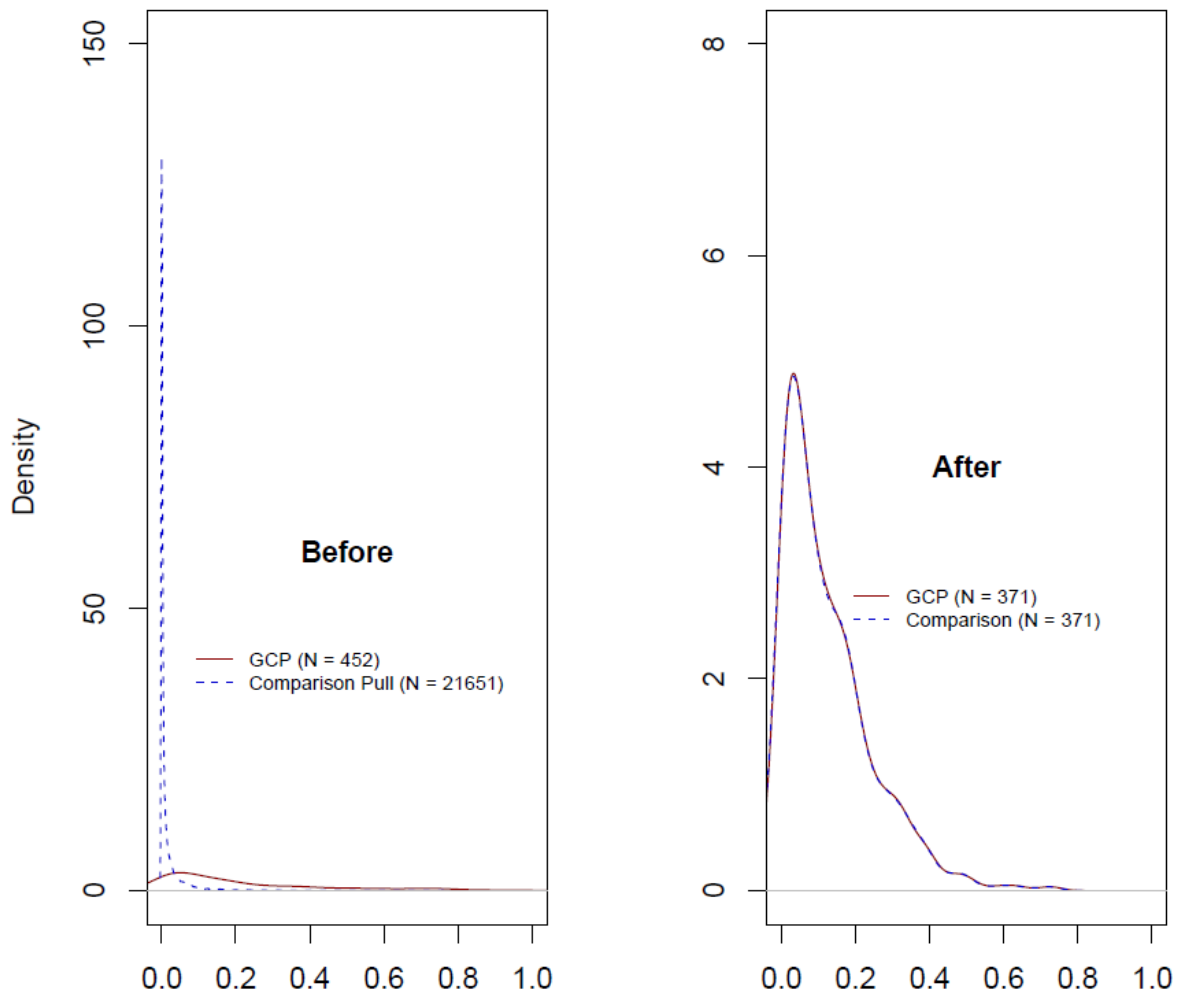
Figure 2: Propensity Score Distributions, Before and After, 2006-2010 Sample

(A) Matching I, Pre-Treatment Middle-school Test Scores as Ability Measure



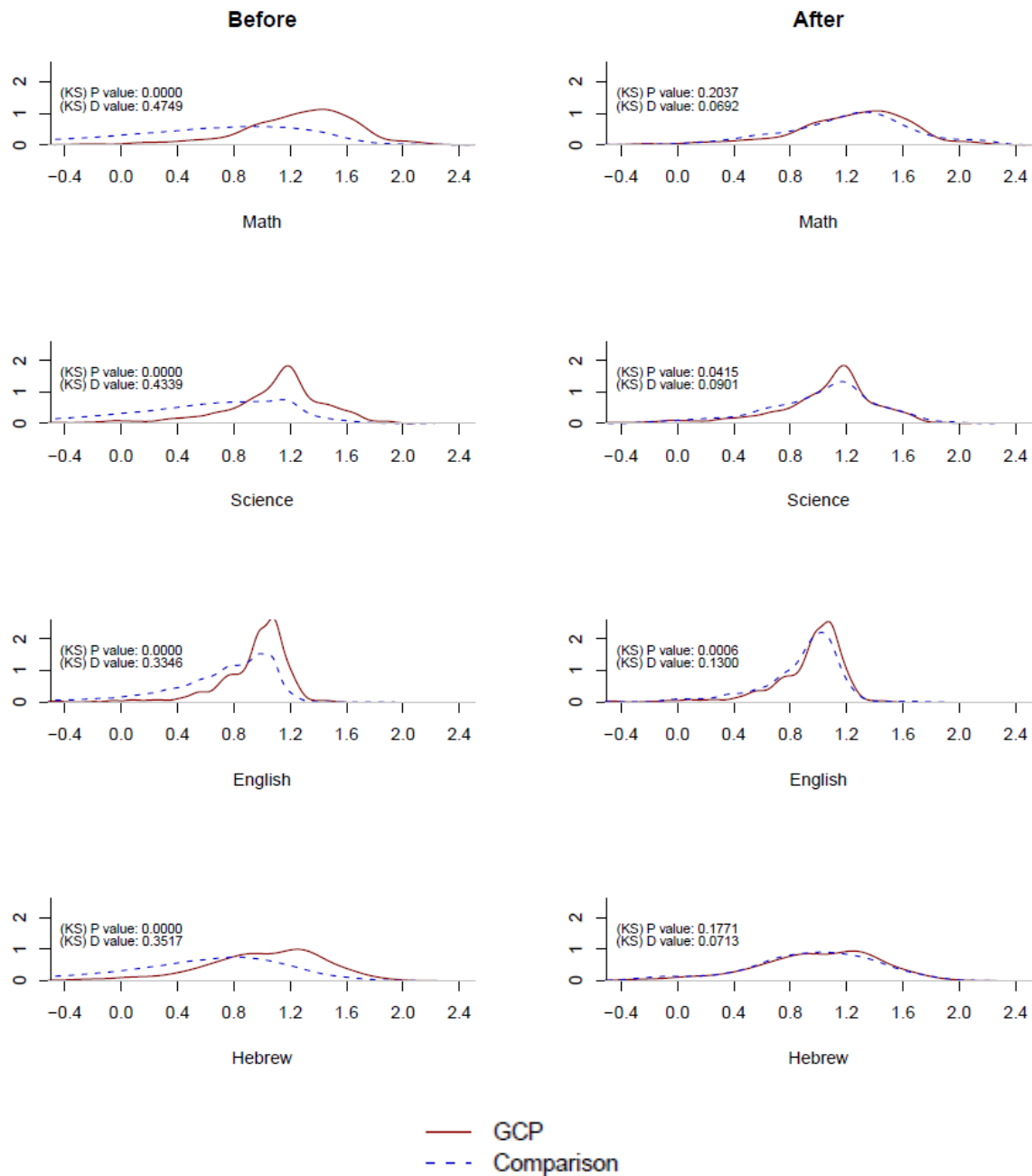
Notes: This figure plots the distribution of the propensity score by groups- the solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching, and the graph on the right shows the distributions after the matching. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010.

(B) Matching II, UPET Scores as Ability Measure



Notes: This figure plots the distribution of the propensity score by groups- the solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching, and the graph on the right shows the distributions after the matching. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010, who took the UPET at any age.

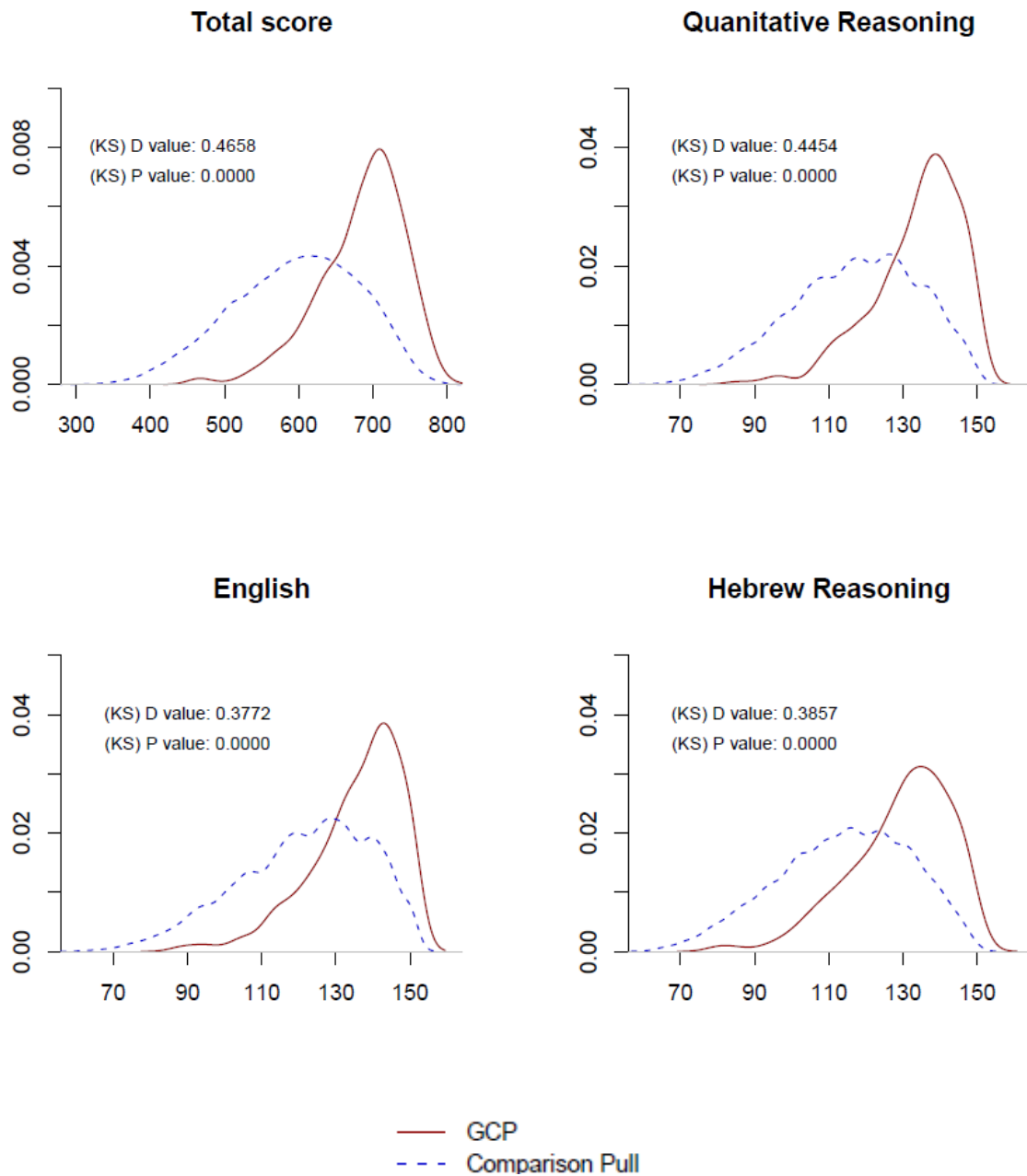
Figure 3: Pre-treatment Middle-school Test Scores, Before and After Matching I, 2006-2010 Sample



Notes: This figure plots the distribution of the Pre-treatment Middle-school Test (Metzav) test scores by groups. The solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graphs on the left show the distributions before the matching, and the graphs on the right show the distributions after the matching (version II). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

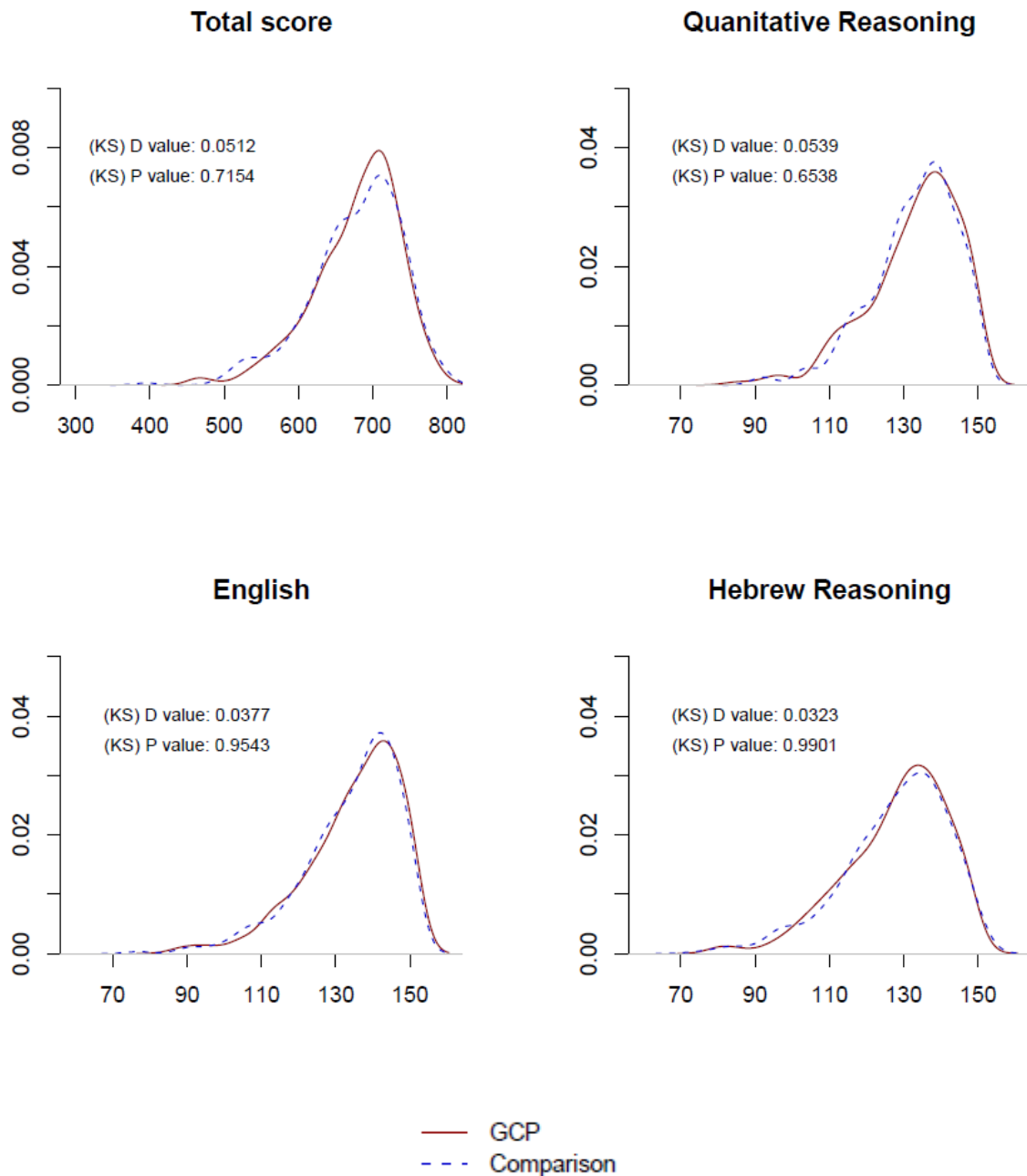
Figure 4: UPET Scores Distributions, GCP Participants and Comparison, 2006-2010 Sample

(A) Before Matching II



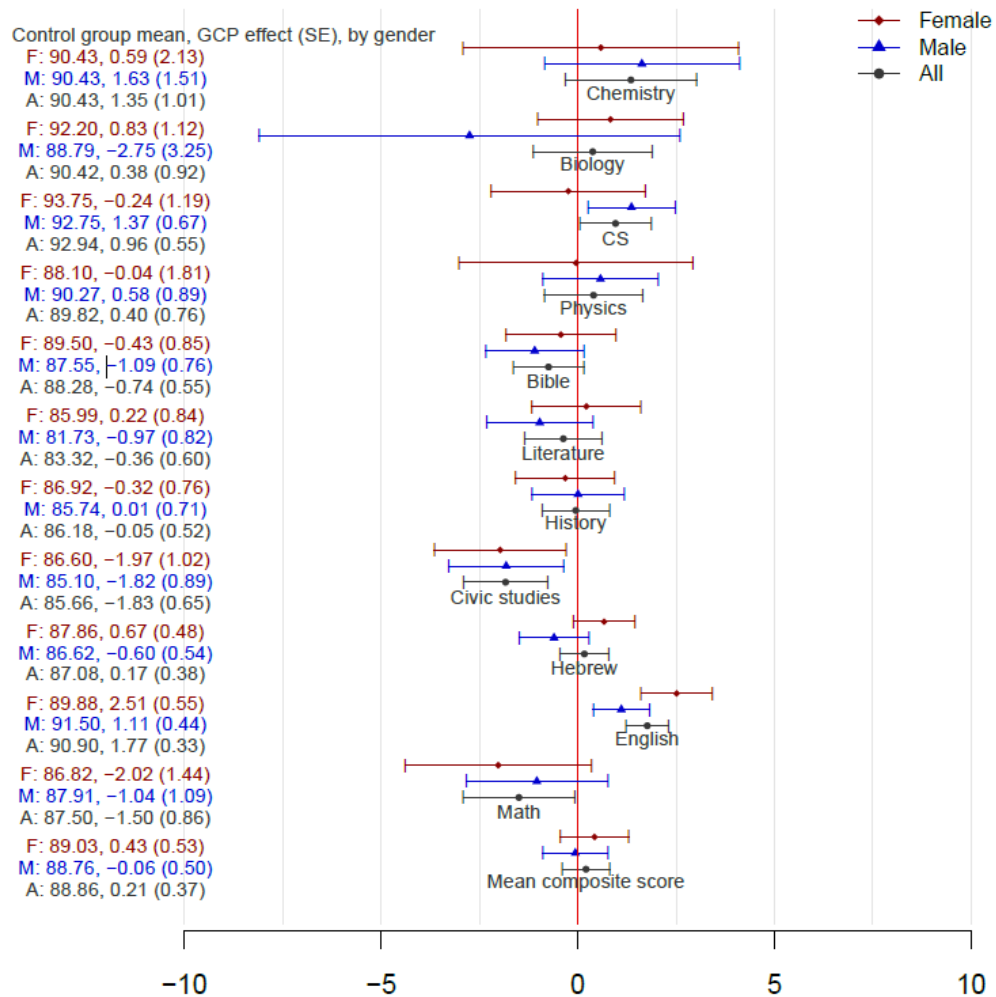
Notes: This figure plots the distribution of UPET scores by group. The solid red line represents the sample of GCP students, and the blue dashed line represents the pull of the comparison group (before matching). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes only students who participated in the Metzav middle school test during their 8th grade, that is about half of the students in cohorts of high-school graduates in 2006-2010, who took the UPET at any age.

(B) After Matching II



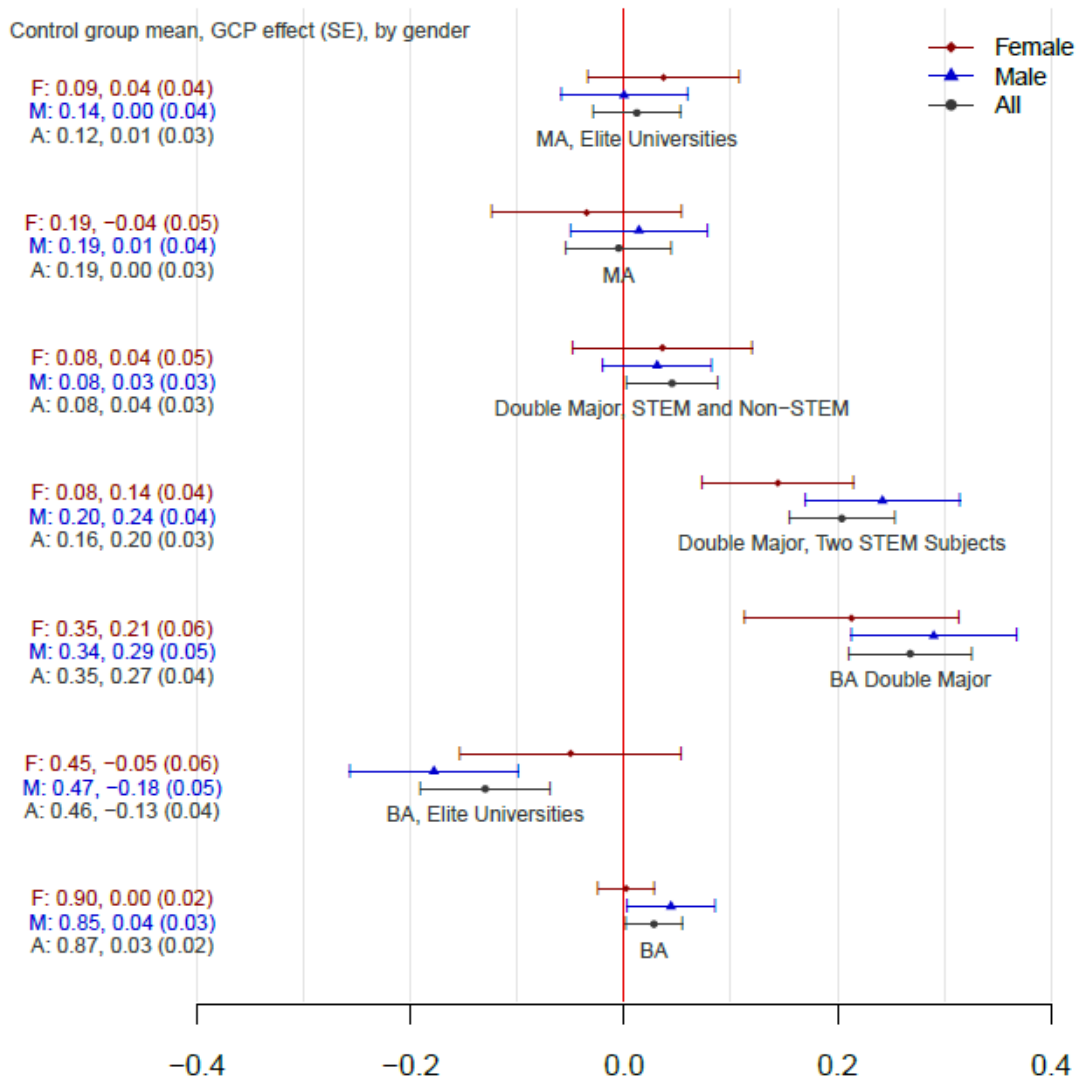
Notes: This figure plots the distribution of UPET scores by group. The solid red line represents the sample of GCP students, and the blue dashed line represents the matched comparison group. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010, who took the UPET at any age.

Figure 5: GCP Effects on Bagrut Test Scores, 2006-2010 Sample



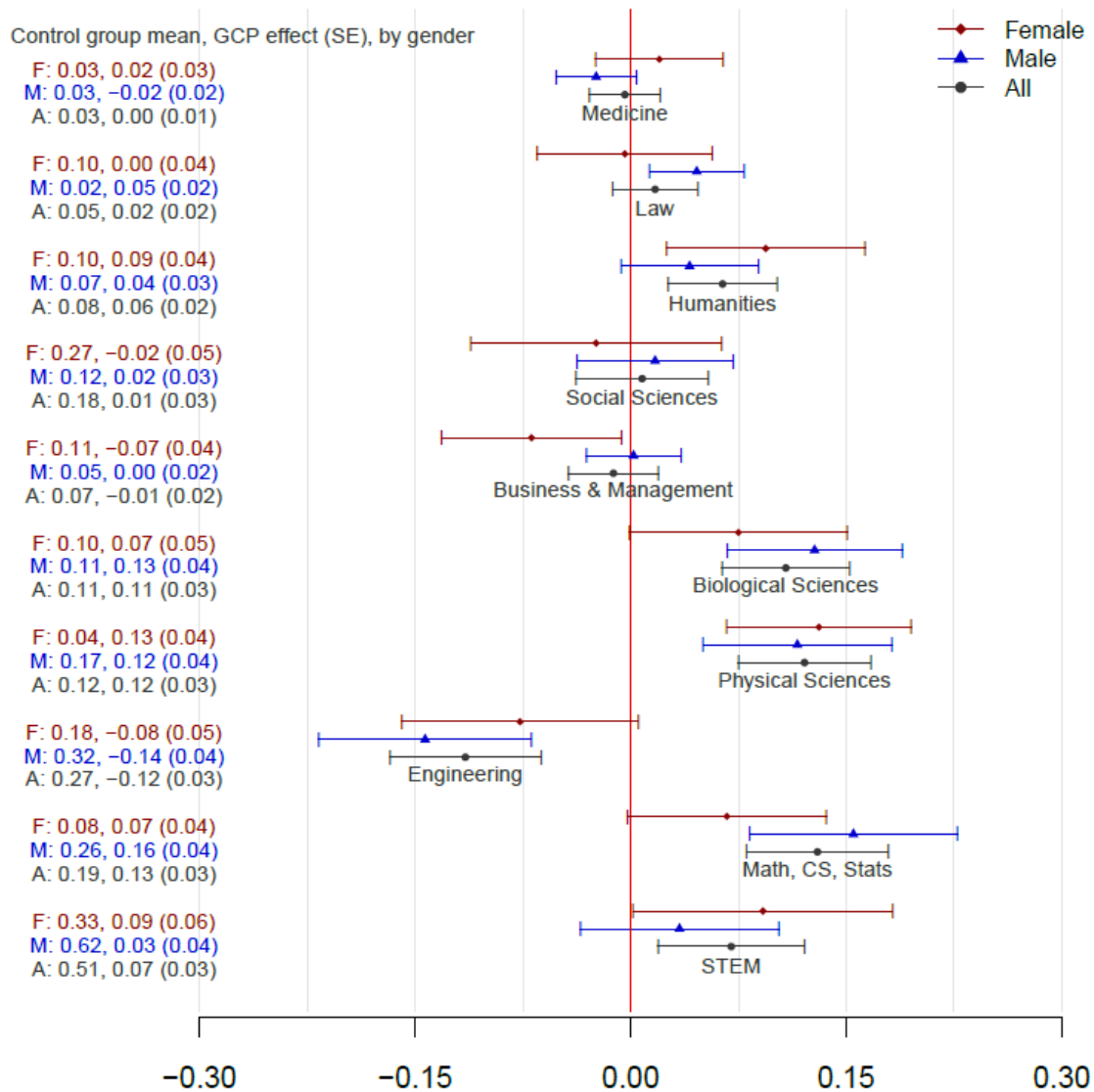
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory subjects). These scores are on a scale of 0-100. The red dots lines represent the sample of females, blue dots and lines represent the sample of males, and dark grey dots and lines represent the entire sample (males and females). The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

Figure 6: GCP Effects on University Degree Attainment, 2006-2010 Sample



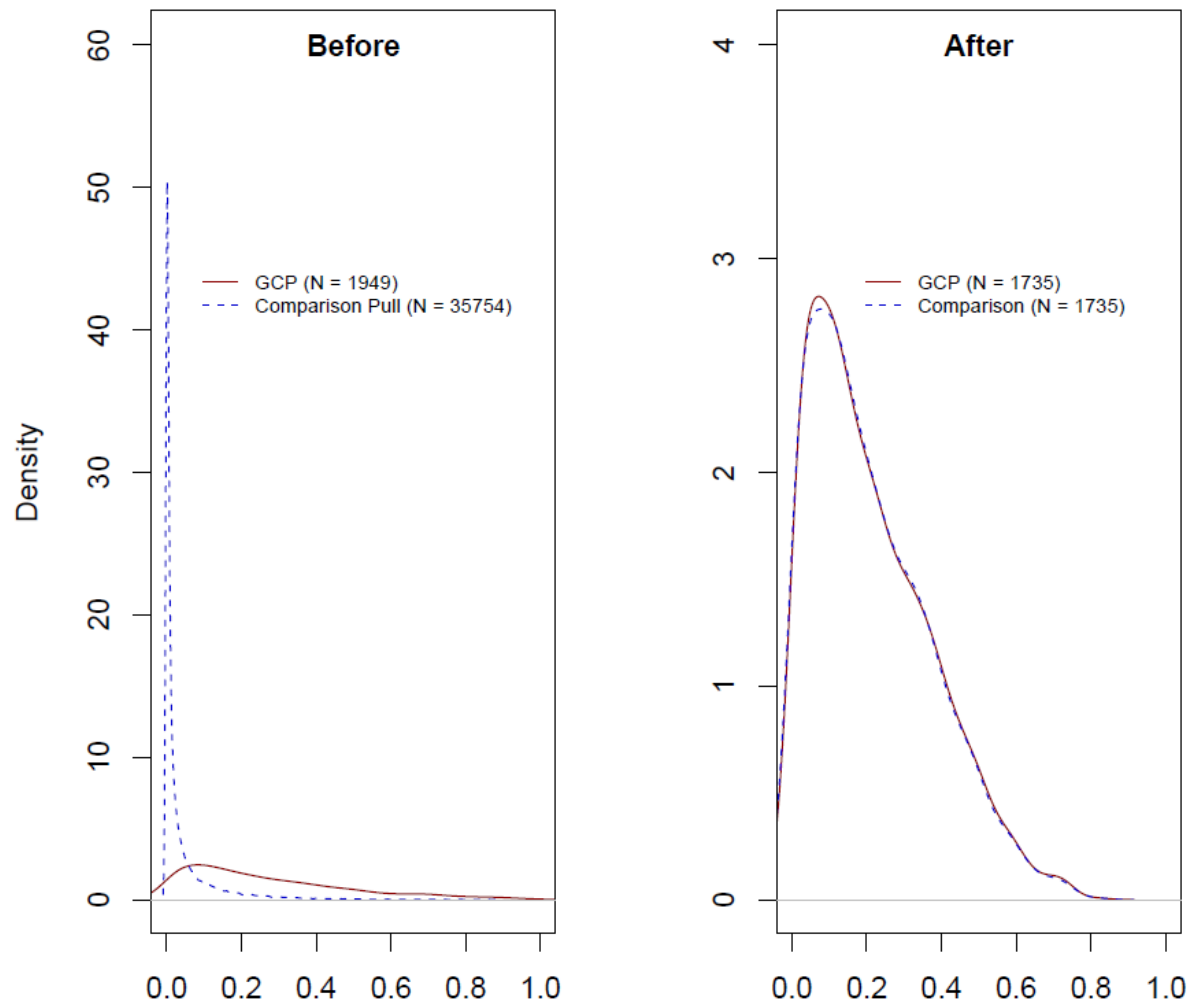
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on university degree attainment. The red dots lines represent the sample of females, blue dots and lines represent the sample of males, and dark grey dots lines represent the entire sample (males and females). The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

Figure 7: GCP Effects on University Field of Studies, 2006-2010 Sample



Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on the university field of studies. Red dots lines represent the sample of females, blue dots lines represent the sample of males, and dark grey dots lines represent the entire sample (males and females). The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

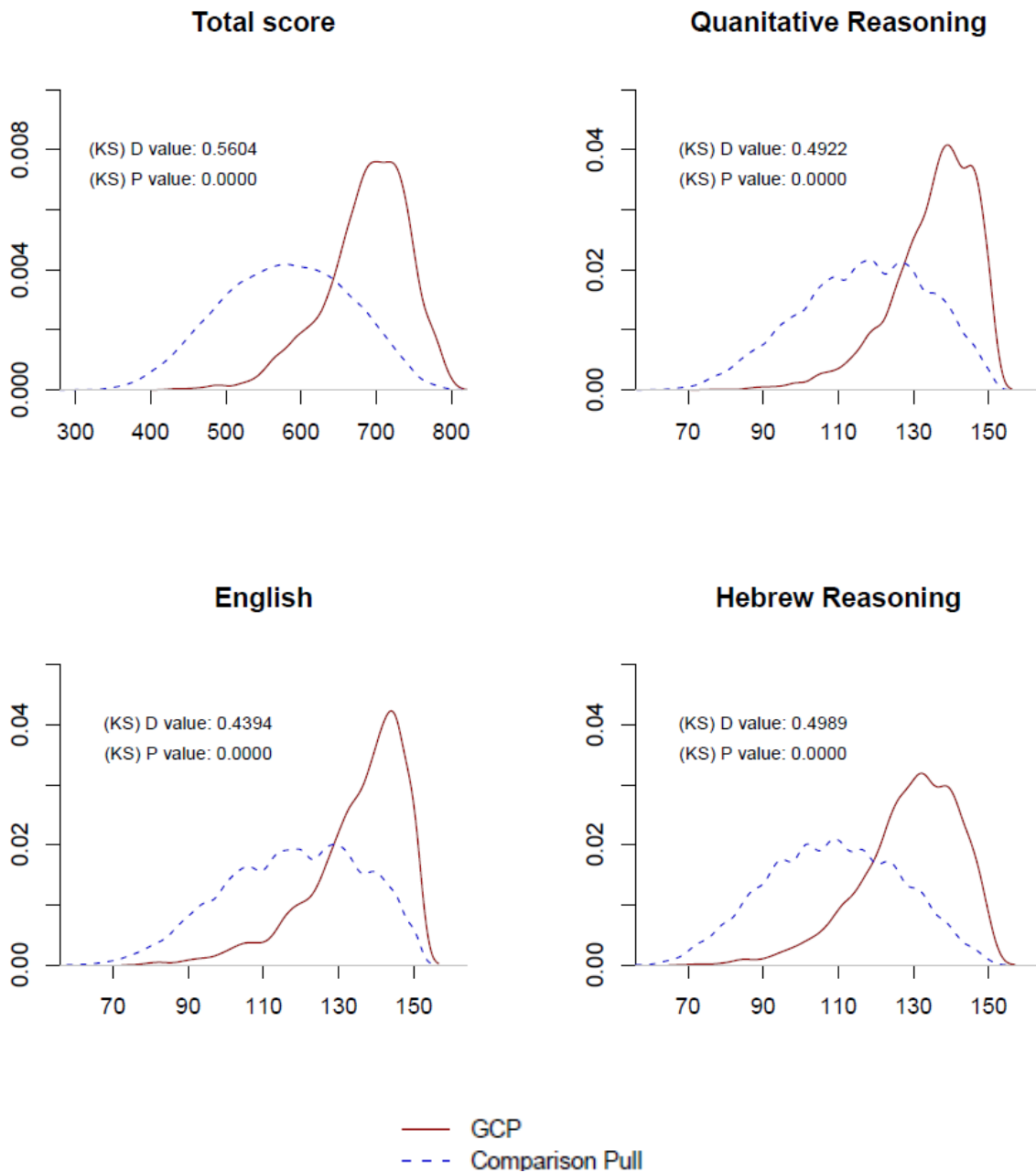
Figure 8: Propensity Score Distributions, Before and After, 1992-2005 Sample



Notes: This figure plots the distribution of the propensity score by groups- the solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching, and the graph on the right shows the distributions after the matching. The sample includes only students from the cohorts of high-school graduates in 1992-2005 who took the UPET during 10th or 11th grade.

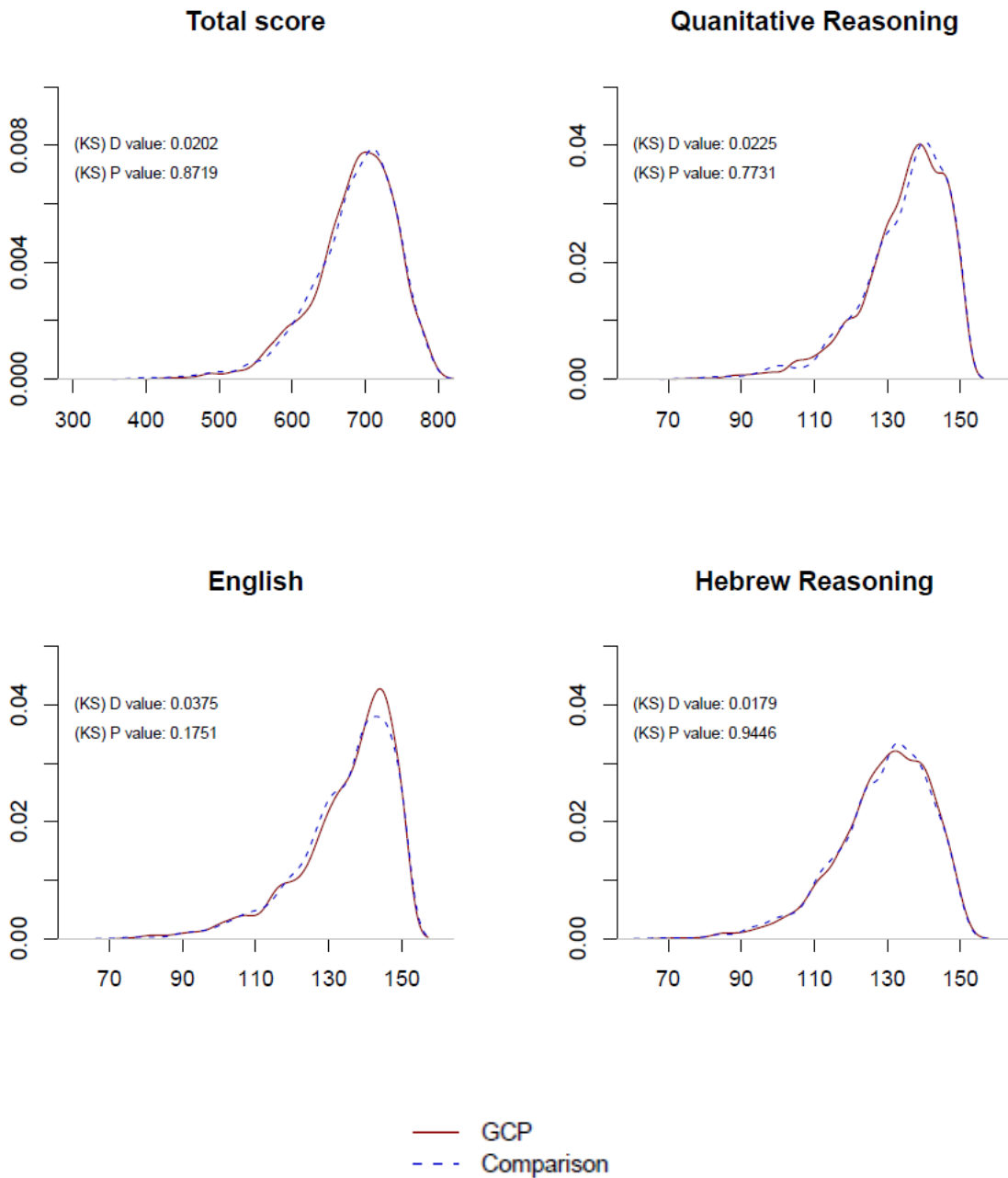
Figure 9: UPET Scores Distributions, GCP Participants and Comparison, 1992-2005 Sample

(A) Before Matching



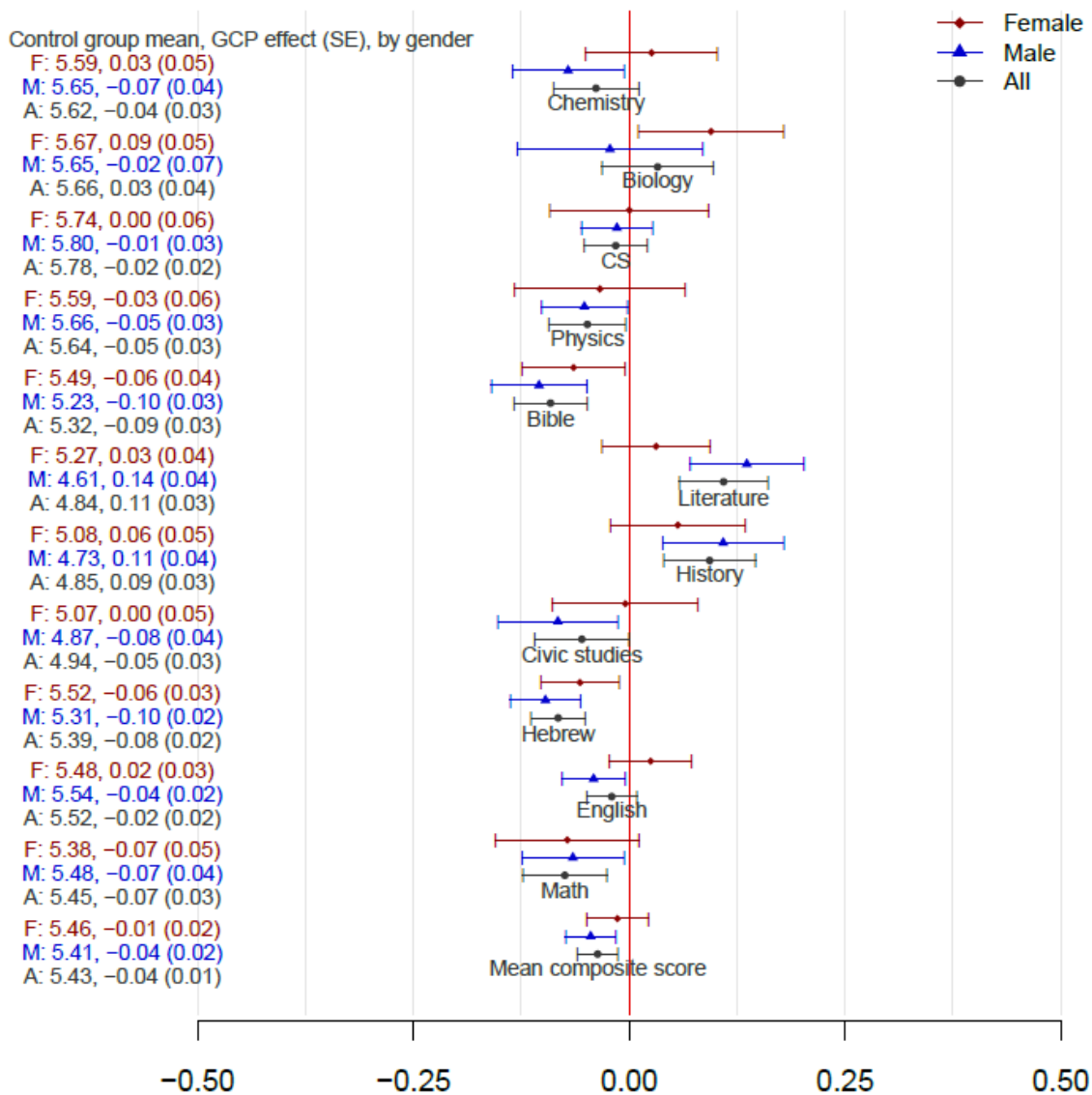
Notes: This figure plots the distribution of UPET scores by group. The solid red line represents the sample of GCP students, and the blue dashed line represents the pull of the comparison group (before matching). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes only students from the cohorts of high-school graduates in 1992-2005 who took the UPET during 10th or 11th grade.

(B) After Matching



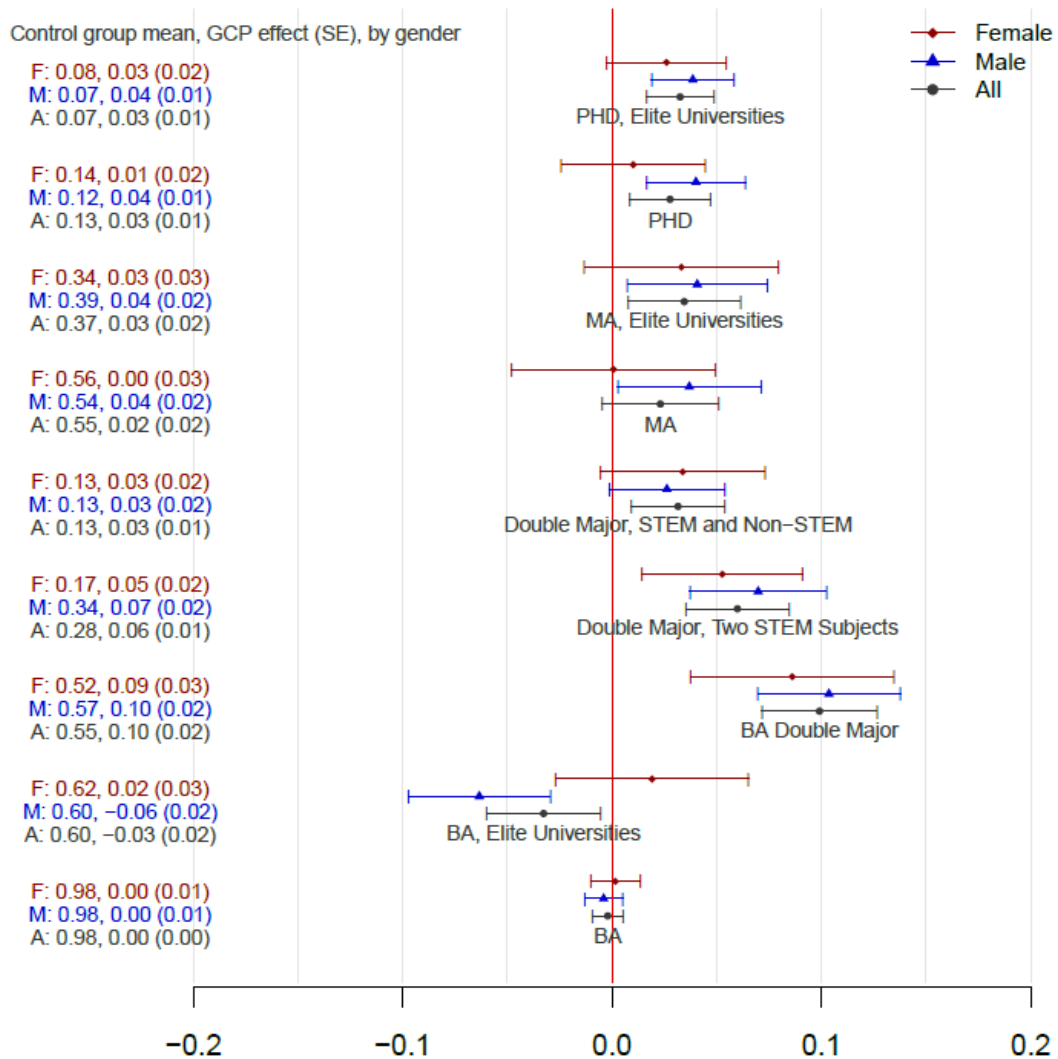
Notes: This figure plots the distribution of UPET scores by group. The solid red line represents the sample of GCP students, and the blue dashed line represents the matched comparison group. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only students from the cohorts of high-school graduates in 1992-2005 who took the UPET during 10th or 11th grade.

Figure 10: GCP Effects on Bagrut Test Scores, 1992-2005 Sample



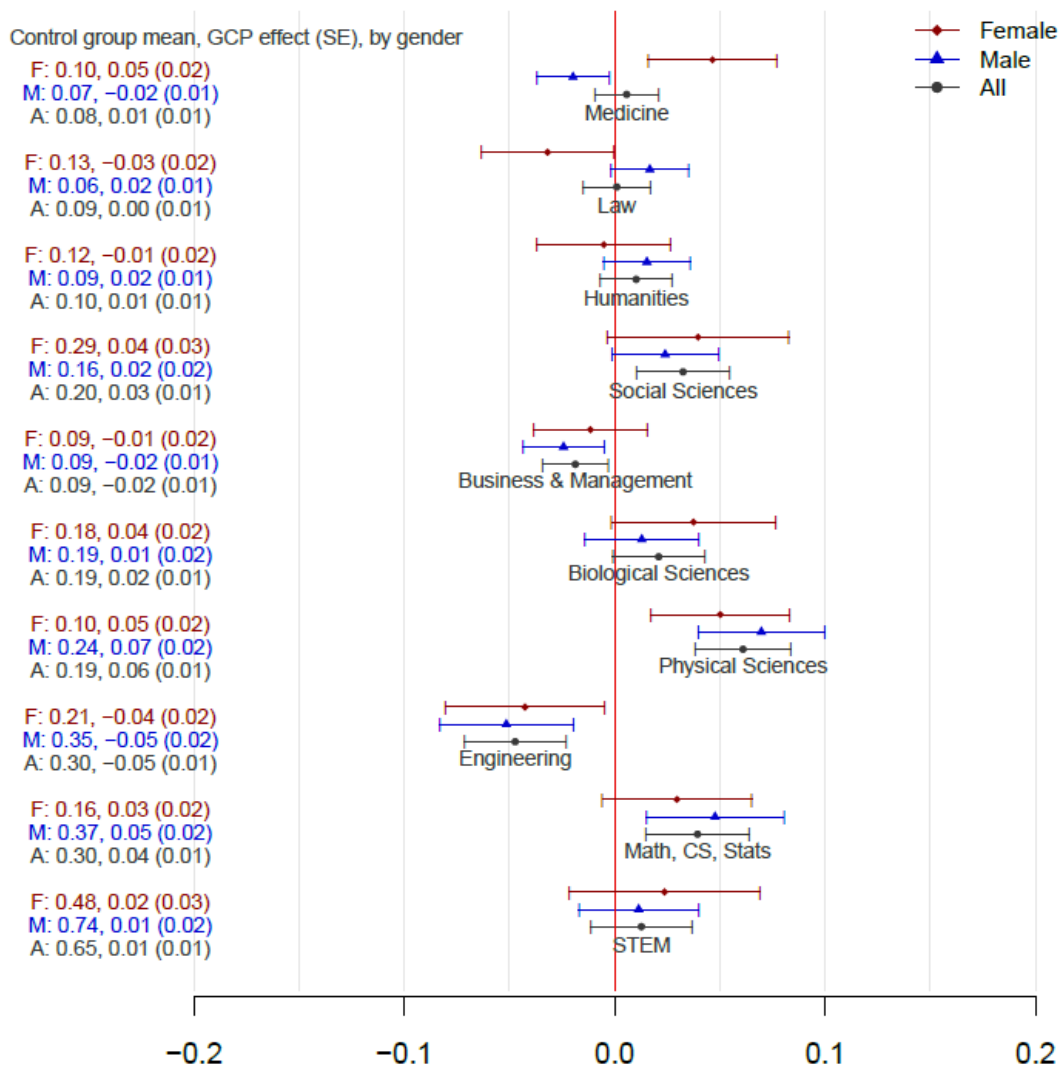
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (all compulsory subjects). These scores are normalized on a scale of 0-6. The Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 11: GCP Effects on University Degree Attainment, 1992-2005 Sample



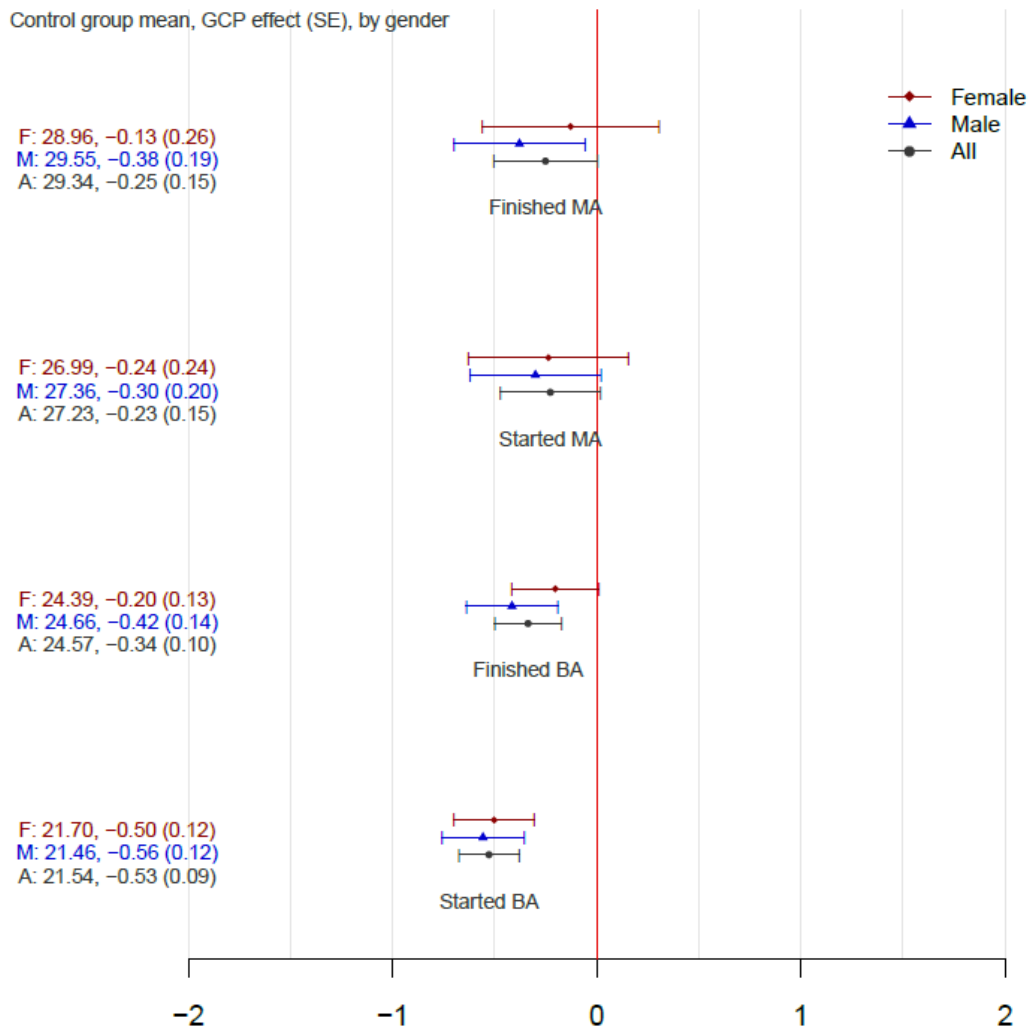
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on university degree attainment. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 12: GCP Effects on University Field of Studies, 1992-2005 Sample



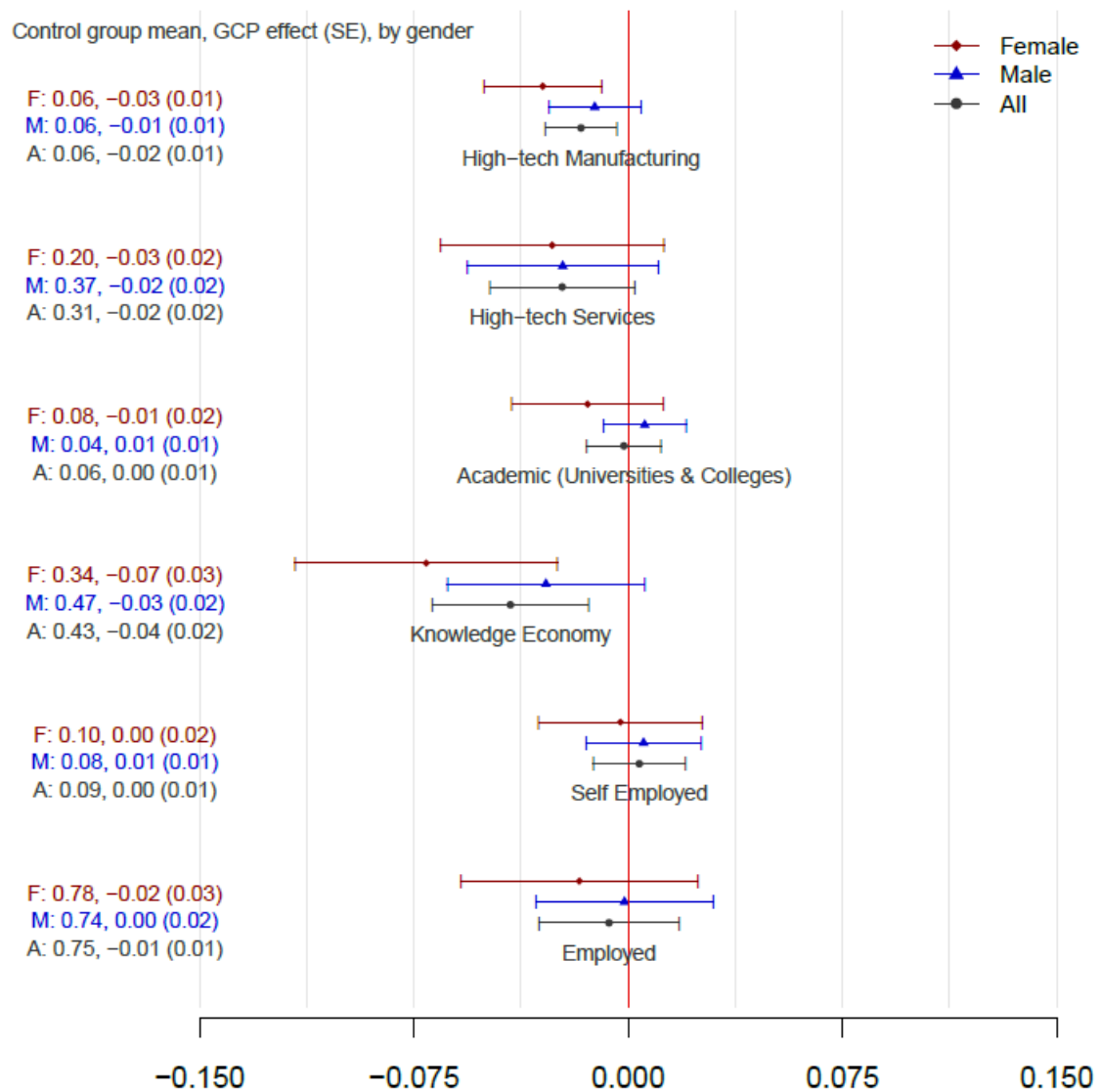
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on the university field of studies. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 13: GCP Effects on Studies Timing, 1992-2005 Sample



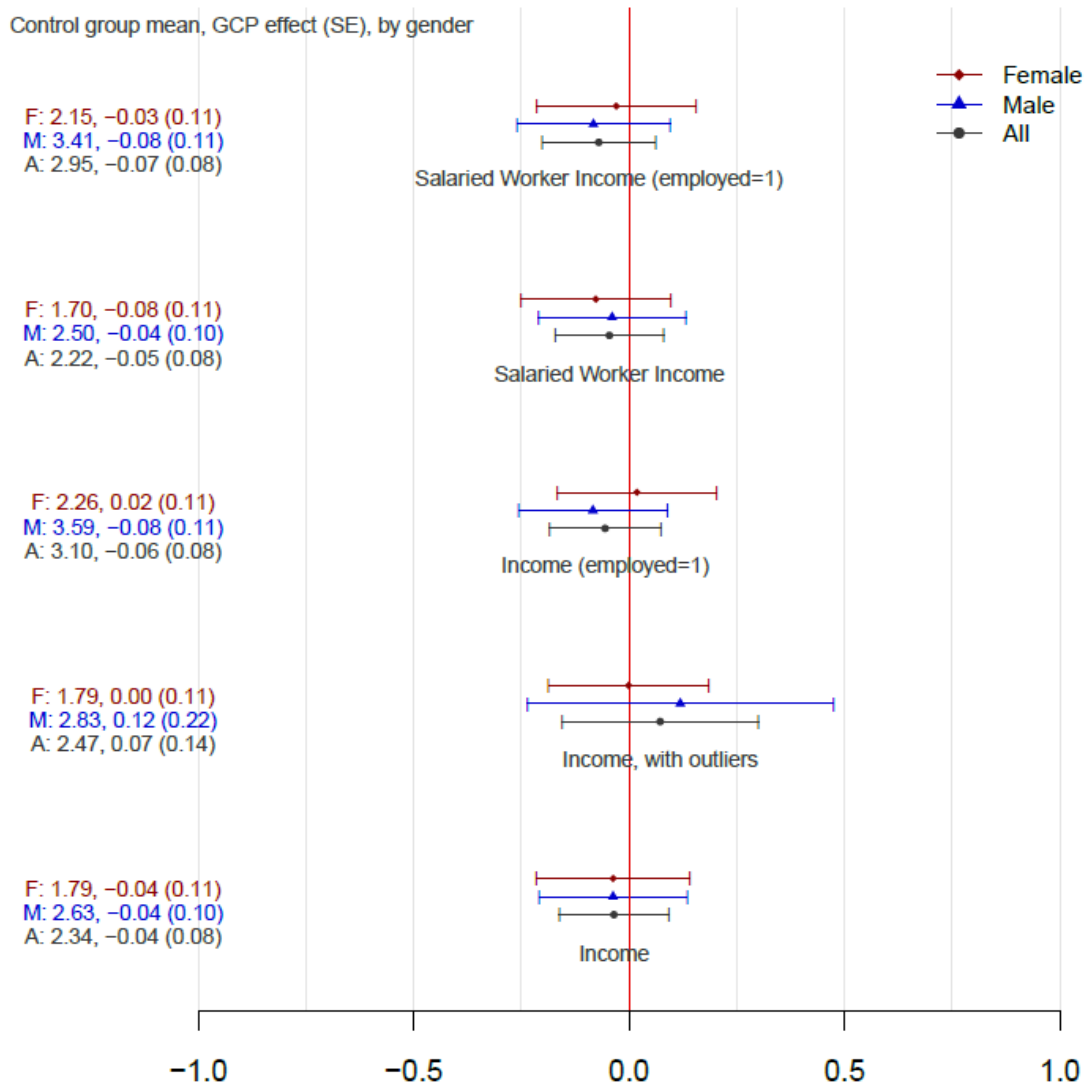
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on studies' timing. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 14: GCP Effect on Employment, by Sector in 2018, 1992-2005 Sample



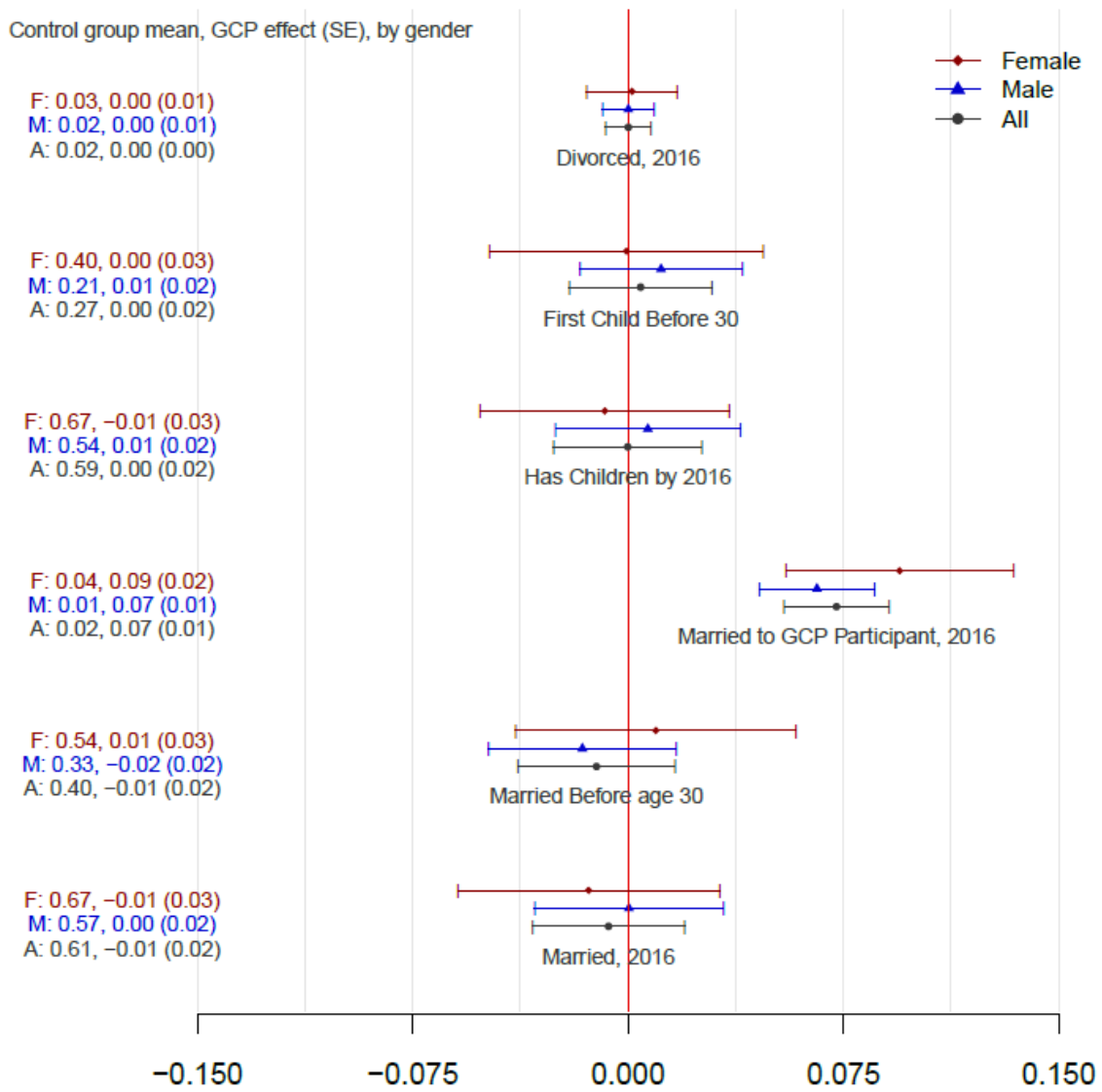
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on employment by Sector in 2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 15: GCP Effect on Annual Income, in 2018, 1992-2005 Sample



Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on annual income (in 100K NIS) in 2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure 16: GCP Effect on Personal Outcomes, 1992-2005 Sample



Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on personal outcomes. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

ONLINE APPENDIX

**Appendix Table A1: Correlation between UPET Scores and 8th Grade Test Scores,
2006-2010 Sample**

	Quantitative	Verbal	English	Total
	(1)	(2)	(3)	(4)
A. GCP Participants				
Math	0.30	0.11	-0.02	0.19
Science	0.16	0.19	0.17	0.22
English	0.15	0.03	0.17	0.13
Hebrew	0.03	0.08	-0.01	0.05
B. Comparison Pull				
Math	0.48	0.27	0.17	0.39
Science	0.31	0.29	0.23	0.33
English	0.20	0.24	0.38	0.29
Hebrew	0.23	0.32	0.19	0.30

Notes: This table presents correlations between the UPET scores and 8th-grade test scores. The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010. Panel A includes only GCP participants, and panel B includes only students from cities with no GCP. Columns represent UPET scores, and rows represent 8th grade test scores.

Appendix Table A2: Localities with and without a GCP, 2006-2010 Sample

	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)
Father education	13.64	13.56	0.08	0.15
Mother education	13.40	13.64	-0.24	0.00*
Number of Students in locality	4208.08	775.67	3432.41	0.00*
Share Eligible for Bagrut	0.71	0.70	0.00	0.55
UPET Score (total)	564.47	565.11	-0.65	0.70
UPET Score (quantitative)	112.68	112.79	-0.11	0.66
UPET Score (verbal)	112.10	112.76	-0.66	0.09*
UPET Score (English)	109.42	109.26	0.16	0.63
Share 5 Credits Bagrut in English	0.41	0.46	-0.05	0.00*
Share 5 Credits Bagrut in Math	0.14	0.16	-0.02	0.00*
Share 5 Credits Bagrut in Physics	0.08	0.10	-0.02	0.00*
Bagrut Credit Units	18.35	21.00	-2.65	0.00*
BA Attainment	0.50	0.55	-0.05	0.00*
Number of Observations	510	510		

Notes: This table presents complimentary descriptive statistics for the full sample (girls and boys). Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. All columns represent students' average level in the same locality concerning different outcome variables. The sample includes only students who participated in the Metzav middle school test during their 8th grade. That is about half of the students in cohorts of high-school graduates in 2006-2010 * represents statistical significance at the 90% level.

Appendix Table A3: Localities with and without a GCP, 1992-2005 Sample

	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)
Father education	12.64	12.38	0.26	0.00*
Mother education	12.34	12.31	0.03	0.57
Number of Students in locality	3414.86	894.24	2520.62	0.00*
Share Eligible for Bagrut	0.69	0.71	-0.02	0.00*
UPET Score (total)	559.42	556.72	2.70	0.00*
UPET Score (quantitative)	111.49	110.96	0.53	0.00*
UPET Score (verbal)	112.43	111.70	0.73	0.00*
UPET Score (English)	108.21	107.91	0.30	0.06*
Share 5 Credits Bagrut in English	0.14	0.14	0.00	0.52
Share 5 Credits Bagrut in Math	0.09	0.10	0.00	0.00*
Share 5 Credits Bagrut in Physics	0.40	0.39	0.01	0.21
Bagrut Credit Points	18.08	19.20	-1.12	0.00*
BA	0.60	0.64	-0.03	0.00*
Number of Observations	1,734	1,734		

Notes: This table presents complimentary descriptive statistics for the full sample (girls and boys). Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. All columns represent students' average level in the same locality concerning different outcome variables. The sample includes only students in cohorts of high-school graduates in 1992-2005 who participated in the UPET during the 10th or 11th grade. * represents statistical significance at the 90% level.

Appendix Table A4: Geographical Mobility, 2006-2010 Sample

	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)
Moved between 6th and 10th grade	0.16	0.21	-0.05	0.14
Moved between 9th and 10th grade	0.04	0.07	-0.03	0.13
Moved between 6th and 9th grade	0.14	0.17	-0.03	0.38
Number of Observations	477	477		

Notes: This table compares the share of students who moved between cities during the years before high school. Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. * represents statistical significance at the 90% level.

Appendix Table A5: Alternative Matching Methods, 1992-2005 Sample

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Replacement N: 1,825 (1,825)	-0.054* (0.014)	-0.112* (0.029)	0.113* (0.018)	0.031* (0.013)	0.047* (0.016)
Caliper = 0.05 N: 1,691 (1,691)	-0.027* (0.015)	-0.057* (0.030)	0.112* (0.017)	0.028* (0.012)	0.043* (0.015)
Caliper = 0.2 N: 1,777 (1,777)	-0.050* (0.014)	-0.080* (0.030)	0.126* (0.017)	0.037* (0.011)	0.051* (0.015)

Notes: This table presents the main results of this paper derived based on the alternative matching methods. The sample is our main (1992-2005 graduates) sample described in the text. Replacement refers to PSM identical to the one described in the text, with the only change we allow matching with replacement. Caliper = 0.05 (Caliper = 0.2) refers to PSM identical to the one described in the text, with the only change of allowing smaller (larger) caliper. * represents statistical significance at the 90% level.

Appendix Table A6: Standard Errors Calculations, 1992-2005 Sample

Panel A.					
Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	-0.0267	-0.0649	0.1015	0.0190	0.0484
SE	0.0172	0.0326	0.0165	0.0119	0.0157
SE Abadie Imbens	0.0146	0.0311	0.0164	0.0119	0.0151

Panel B.					
Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	-0.0271	-0.0730	0.0980	0.0221	0.0446
SE	0.0141	0.0294	0.0167	0.0118	0.0147
SE Bootstrap	0.0147	0.0287	0.0167	0.0120	0.0147

Notes: This table presents standard errors calculations. The upper panel compares the standard errors that we use in our descriptive tables (7, A12, A13) with the correction offered by Abadie and Imbens (2008). The lower panel compares the standard errors used in our main figures (5-11) with bootstrapped standard errors. We use our main sample (1992-2005 graduates) described in the text.

Appendix Table A7: Peers Characteristics and Educational Outcomes, 2006-2010

	Sample			
	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)
Father education	16.06	14.42	1.64	0.00*
Mother education	15.81	14.53	1.28	0.00*
Number of Students in class	26.04	34.71	-8.68	0.00*
Share Eligible for Bagrut	0.96	0.87	0.09	0.00*
UPET Score (total)	685.52	593.70	91.82	0.00*
UPET Score (quantitative)	134.40	118.43	15.96	0.00*
UPET Score (verbal)	135.27	118.92	16.36	0.00*
UPET Score (English)	129.37	113.15	16.22	0.00*
Share 5 Credits Bagrut in English	0.98	0.75	0.23	0.00*
Share 5 Credits Bagrut in Math	0.82	0.37	0.45	0.00*
Share 5 Credits Bagrut in Physics	0.53	0.25	0.29	0.00*
Bagrut Credit Points	30.77	26.22	4.55	0.00*
BA	0.92	0.75	0.17	0.00*
Number of Observations	510	510		

Notes: This table presents complimentary descriptive statistics for the full sample (girls and boys). Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. All columns represent the average level of peers (students in the same class in high school) regarding different outcome variables. The sample includes only students who participated in the Metzav middle school tests during their 8th grade. That is about half of the students in cohorts of high-school graduates in 2006-2010

* represents statistical significance at the 90% level.

Appendix Table A8: Peers Characteristics and Educational Outcomes, 1992-2005 Sample

	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)
Father education	14.64	13.63	1.01	0.00*
Mother education	14.44	13.57	0.87	0.00*
Number of Students in class	26.82	45.54	-18.71	0.00*
Share Eligible for Bagrut	0.97	0.87	0.10	0.00*
UPET Score (total)	694.59	594.52	100.06	0.00*
UPET Score (quantitative)	135.57	118.53	17.05	0.00*
UPET Score (verbal)	136.67	119.11	17.56	0.00*
UPET Score (English)	131.41	113.32	18.09	0.00*
Share 5 Credits Bagrut in English	0.98	0.71	0.26	0.00*
Share 5 Credits Bagrut in Math	0.84	0.40	0.44	0.00*
Share 5 Credits Bagrut in Physics	0.51	0.28	0.22	0.00*
Bagrut Credit Points	29.31	25.56	3.75	0.00*
BA	0.97	0.87	0.09	0.00*
Number of Observations	1,734	1,734		

Notes: This table presents complimentary descriptive statistics for the full sample (girls and boys). Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. All columns represent the average level of peers (students in the same class in high school) concerning different outcome variables. The sample includes only students in cohorts of high-school graduates in 1992-2005 who participated in the UPET during the 10th or 11th grade. * represents statistical significance at the 90% level.

**Appendix Table A9: Comparison of GCP Participants, by Length of Participation, 2006-2010
Sample**

	Only in High- School	Since Middle School	Difference	p-value
	(1)	(2)	(3)	(4)
Father education	16.19	15.12	1.07	0.00*
Mother education	15.47	15.17	0.30	0.00*
Born in Israel	0.84	0.89	-0.05	0.08*
Father Born in Israel	0.62	0.65	-0.03	0.48
Mother born Israel	0.68	0.65	0.04	0.43
Father income 2003	1.77	1.62	0.15	0.42
Mother income 2003	0.83	1.02	-0.20	0.04*
5 Credits Math Bagrut	0.79	0.83	-0.03	0.36
5 Credits Physics Bagrut	0.51	0.53	-0.02	0.70
8th Grade Test Score (math)	1.15	1.27	-0.12	0.01*
8th Grade Test Score (English)	0.89	0.94	-0.05	0.10
8th Grade Test Score (Hebrew)	0.94	0.99	-0.05	0.27
8th Grade Test Score (science)	1.06	1.02	0.04	0.28
UPET Score (total)	680.95	685.24	-4.29	0.47
UPET Score (quantitative)	132.79	134.53	-1.74	0.15
UPET Score (verbal)	129.01	129.05	-0.04	0.98
UPET Score (English)	135.27	135.49	-0.22	0.86
Number of Observations	510	510		

Notes: This table presents complimentary descriptive statistics for the full sample (girls and boys). Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Columns (3) and (4) show the difference between the means and the corresponding p-value. The sample includes only GCP participants who participated in the Meitzav middle school tests during their 8th grade. This includes about half of the GCP participants in cohorts of high-school graduates in 2006-2010. * represents statistical significance at the 90% level.

Appendix Table A10: Descriptive Statistics, Demographics and Psychometric Scores, 1992-2005 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Family Background:												
Father education	14.61	14.65	-0.04	0.76	14.68	14.55	0.13	0.55	14.57	14.70	-0.13	0.43
Mother education	14.35	14.43	-0.08	0.54	14.50	14.37	0.13	0.52	14.27	14.46	-0.19	0.23
Born in Israel	0.87	0.87	0.01	0.55	0.87	0.84	0.03	0.14	0.88	0.88	0.00	1.00
Father Born in Israel	0.53	0.52	0.00	0.81	0.53	0.50	0.03	0.30	0.53	0.53	0.00	1.00
Mother born Israel	0.58	0.59	-0.01	0.68	0.56	0.56	0.00	1.00	0.59	0.60	-0.01	0.63
Father income 1990	0.51	0.49	0.02	0.21	0.51	0.49	0.02	0.47	0.50	0.48	0.02	0.31
Mother income 1990	0.21	0.19	0.02	0.02*	0.21	0.19	0.02	0.12	0.21	0.19	0.02	0.04*
Siblings	1.92	1.90	0.03	0.57	1.86	1.91	-0.05	0.47	1.95	1.89	0.06	0.29
B. UPET scores:												
Total	686.14	684.99	1.15	0.56	671.38	666.99	4.39	0.23	693.99	694.55	-0.56	0.80
Verbal Reasoning	129.24	128.93	0.31	0.48	129.33	127.88	1.45	0.05*	129.19	129.49	-0.30	0.57
Quantitative Reasoning	134.84	134.81	0.03	0.94	130.10	130.02	0.08	0.92	137.36	137.36	0.00	1.00
English	135.19	134.88	0.31	0.47	131.62	130.91	0.71	0.38	137.09	136.99	0.10	0.84
Number of Observations	1,735	1,735			602	602			1,133	1,133		

Notes: This table presents descriptive statistics for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. Income is measured in 100K NIS. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. * represents statistical significance at the 90% level.

Appendix Table A11: Distribution of Bagrut Subjects at Advanced Level (5 Credits), 1992-2005 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Math	0.89	0.89	0.00	0.67	0.81	0.83	-0.02	0.37	0.93	0.91	0.02	0.07*
Physics	0.57	0.59	-0.01	0.49	0.33	0.34	-0.01	0.71	0.70	0.72	-0.02	0.30
Computer Sciences	0.48	0.47	0.01	0.50	0.30	0.24	0.06	0.02*	0.57	0.59	-0.02	0.34
Chemistry	0.44	0.45	-0.01	0.47	0.49	0.50	-0.01	0.73	0.42	0.43	-0.01	0.63
Biology	0.18	0.18	0.00	0.86	0.29	0.29	0.00	1.00	0.12	0.12	0.00	1.00
Arabic	0.09	0.09	-0.01	0.55	0.10	0.14	-0.04	0.03*	0.08	0.07	0.01	0.36
Literature	0.08	0.08	0.00	0.80	0.15	0.16	-0.01	0.63	0.04	0.04	0.00	1.00
Social Sciences	0.07	0.07	0.00	1.00	0.11	0.11	0.00	1.00	0.04	0.04	0.00	1.00
History	0.04	0.04	0.01	0.44	0.04	0.04	0.00	1.00	0.04	0.04	0.00	1.00
Art	0.03	0.02	0.01	0.25	0.04	0.04	0.00	1.00	0.02	0.01	0.01	0.03*
Total credits	30.00	29.83	0.17	0.17	29.27	29.00	0.27	0.13	30.38	30.27	0.11	0.48
Number of Observations	1,735	1,735			602	602			1,133	1,133		

Notes: This table presents descriptive statistics for the full sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. Total credits refer to the total credits accumulated in the bagrut. All other variables are indicator variables for five credits bagrut in the subject. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. * represents statistical significance at the 90% level.

Appendix Table A12: GCP Treatment Effect Estimates, 1992-2005 Sample, by Matching Models

Panel A.		Bagrut Exam Scores					Higher Education				
Outcome:	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	MA	PHD	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Main, no bagrut	0.01	-0.10*	-0.03	-0.04	0.04*	0.04*	0.12*	0.06*	0.06*	-0.09*	0.07*
N: 1,892 (1,892)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Main, only quant.	0.05*	-0.07*	0.01	0.04	0.03*	0.03*	0.13*	0.06*	0.07*	-0.10*	0.07*
N: 1,760 (1,760)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)

Panel B.		Labor market outcomes							Personal		
Outcome:	Employed 2018	Self Employed 2018	Salaried Income 2018	Salaried Income (employed=1) 2018	Income (employed=1) 2018	HT services 2018	HT manufacturing 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Main, no bagrut	-0.01	0.01	-0.06	-0.09	-0.07	-0.02	-0.02*	-0.04*	0.01	-0.02	0.00
N: 1,892 (1,892)	(0.01)	(0.01)	(0.07)	(0.08)	(0.08)	(0.01)	(0.01)	0.02	(0.01)	(0.02)	(0.01)
Main, only quant.	-0.03*	0.01	-0.15*	-0.13	-0.09	-0.01	-0.03*	-0.03*	0.01	-0.02	0.00
N: 1,760 (1,760)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	0.02	(0.01)	(0.02)	(0.02)

Notes: This table presents the treatment effect estimated on girls by sample. Each row represents different matching specifications, and each column represents different outcome variables. The sample ‘Main no bagrut’ is identical to the main sample, except we omit all Bagrut dummies from the matching. The sample ‘Main only quant.’ is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the quantitative score). * represents statistical significance at the 90% level.

Appendix Table A13: Bagrut Test Scores in Compulsory Subjects, 1992-2005 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean composite score	5.39	5.43	-0.03	0.05*	5.45	5.46	-0.01	0.73	5.36	5.41	-0.05	0.02*
English	5.51	5.52	-0.01	0.70	5.52	5.48	0.04	0.31	5.51	5.54	-0.03	0.29
Math	5.37	5.45	-0.07	0.02*	5.31	5.38	-0.07	0.20	5.40	5.48	-0.08	0.05*
Bible	5.21	5.32	-0.11	0.00*	5.43	5.49	-0.06	0.15	5.10	5.23	-0.13	0.00*
Literature	4.93	4.84	0.09	0.01*	5.30	5.27	0.03	0.46	4.73	4.61	0.12	0.01*
History	4.96	4.85	0.10	0.00*	5.12	5.08	0.04	0.43	4.87	4.73	0.14	0.00*
Civic studies	4.90	4.94	-0.04	0.29	5.06	5.07	-0.01	0.87	4.82	4.87	-0.05	0.25
Hebrew	5.30	5.39	-0.09	0.00*	5.46	5.52	-0.06	0.07*	5.22	5.31	-0.09	0.00*
Number of Observations	1,735	1,735			602	602			1,133	1,133		

Notes: This table presents descriptive statistics for the entire sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. All scores are on a scale of 0-100. We included all compulsory subjects. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. * represents statistical significance at the 90% level.

Appendix Table A14: Descriptive Statistics, University Degrees and Field of Study, 1992-2005 Sample

	Both Genders				Girls				Boys			
	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value	GCP	Comparison	difference	p-value
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Degrees:												
BA	0.98	0.98	0.00	0.89	0.99	0.98	0.01	0.16	0.98	0.98	0.00	1.00
BA, Double Major	0.65	0.55	0.10	0.00*	0.61	0.52	0.09	0.00*	0.67	0.57	0.10	0.00*
MA	0.57	0.55	0.02	0.16	0.56	0.56	0.00	1.00	0.57	0.54	0.03	0.15
PHD	0.15	0.13	0.03	0.02*	0.15	0.14	0.01	0.62	0.15	0.12	0.03	0.04*
B. Field of Study:												
STEM	0.66	0.65	0.01	0.62	0.50	0.48	0.02	0.49	0.74	0.74	0.00	1.00
Math, Computer Sciences, Statistics	0.33	0.30	0.04	0.01*	0.19	0.16	0.03	0.17	0.41	0.37	0.04	0.05*
Engineering	0.25	0.30	-0.05	0.00*	0.16	0.21	-0.05	0.02*	0.29	0.35	-0.06	0.00*
Physical Sciences	0.25	0.19	0.06	0.00*	0.15	0.10	0.05	0.01*	0.30	0.24	0.06	0.00*
Biological Sciences	0.20	0.19	0.02	0.27	0.22	0.18	0.04	0.08*	0.20	0.19	0.01	0.55
Business and Management	0.08	0.09	-0.02	0.10*	0.08	0.09	-0.01	0.53	0.08	0.09	-0.01	0.39
Social Sciences	0.24	0.20	0.04	0.01*	0.33	0.29	0.04	0.13	0.19	0.16	0.03	0.06*
Humanities	0.12	0.10	0.01	0.21	0.13	0.12	0.01	0.60	0.11	0.09	0.02	0.11
Law	0.09	0.09	0.00	0.64	0.10	0.13	-0.03	0.10	0.08	0.06	0.02	0.07*
Medicine	0.08	0.08	0.00	0.66	0.14	0.10	0.04	0.03*	0.05	0.07	-0.02	0.04*
Number of Observations	1,735	1,735			602	602			1,133	1,133		

Notes: This table presents descriptive statistics for the entire sample (columns 1-4), the Girls' sample (5-8), and the Boys' sample (9-12). Columns (1), (5), and (9) show the means among the treatment group, Columns (2), (6), and (10) show the means among the comparison group, Columns (3), (7), (11) and (4), (8), (12) show the difference between the means and the corresponded p-value. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade. * represents statistical significance at the 90% level.

Table A15: GCP Treatment Estimated Effect, 1992-2005 Sample, By Sample and Matching Models

Panel B. Outcome:	Employed 2018	Self Employed 2018	Salaried Worker Income 2018	Salaried Worker Income (employed=1) 2018	Income (employed=1) 2018	High-tech services 2018	High-tech manufactu ring 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Main, exact cohort	0.01	0.01	-0.11	-0.20*	-0.16*	0.00	-0.02*	-0.01	0.01	-0.02	-0.01
N: 1,549 (1,549)	(0.02)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Main 1992-2000	-0.01	0.00	-0.06	-0.07	-0.07	-0.03	-0.01	-0.04*	0.00	-0.02	0.00
N: 1,287 (1,287)	(0.02)	(0.01)	(0.10)	(0.10)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Main, In Israel at 2018	0.00	0.01	-0.05	-0.06	-0.05	-0.03	-0.02*	-0.05*	0.00	-0.01	0.00
N: 1,526 (1,532)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Extend all, exact cohort	-0.03*	-0.01	-0.09	-0.04	-0.06	-0.01	-0.02*	-0.02*	0.01	-0.02	-0.01
N: 2,644 (2,644)	(0.01)	(0.01)	(0.06)	(0.06)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Extend all, 1992-2000	-0.04*	0.00	-0.17*	-0.11	-0.11	-0.02	-0.02*	-0.02	0.01	0.00	-0.01
N: 1,974 (1,974)	(0.01)	(0.01)	(0.07)	(0.08)	(0.08)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)

Notes: This table presents the treatment effect estimated on girls by sample. Each row represents a different sample, and each column represents a different outcome variable. Main\Extend all exact cohort is identical to the main\extend all sample, except that we match exactly the cohort. The sample 'Main\Extend all 1992-2000' is a subset of the 'main\extend all' sample, including only the older cohorts (1992-2000). * represents statistical significance at the 90% level.

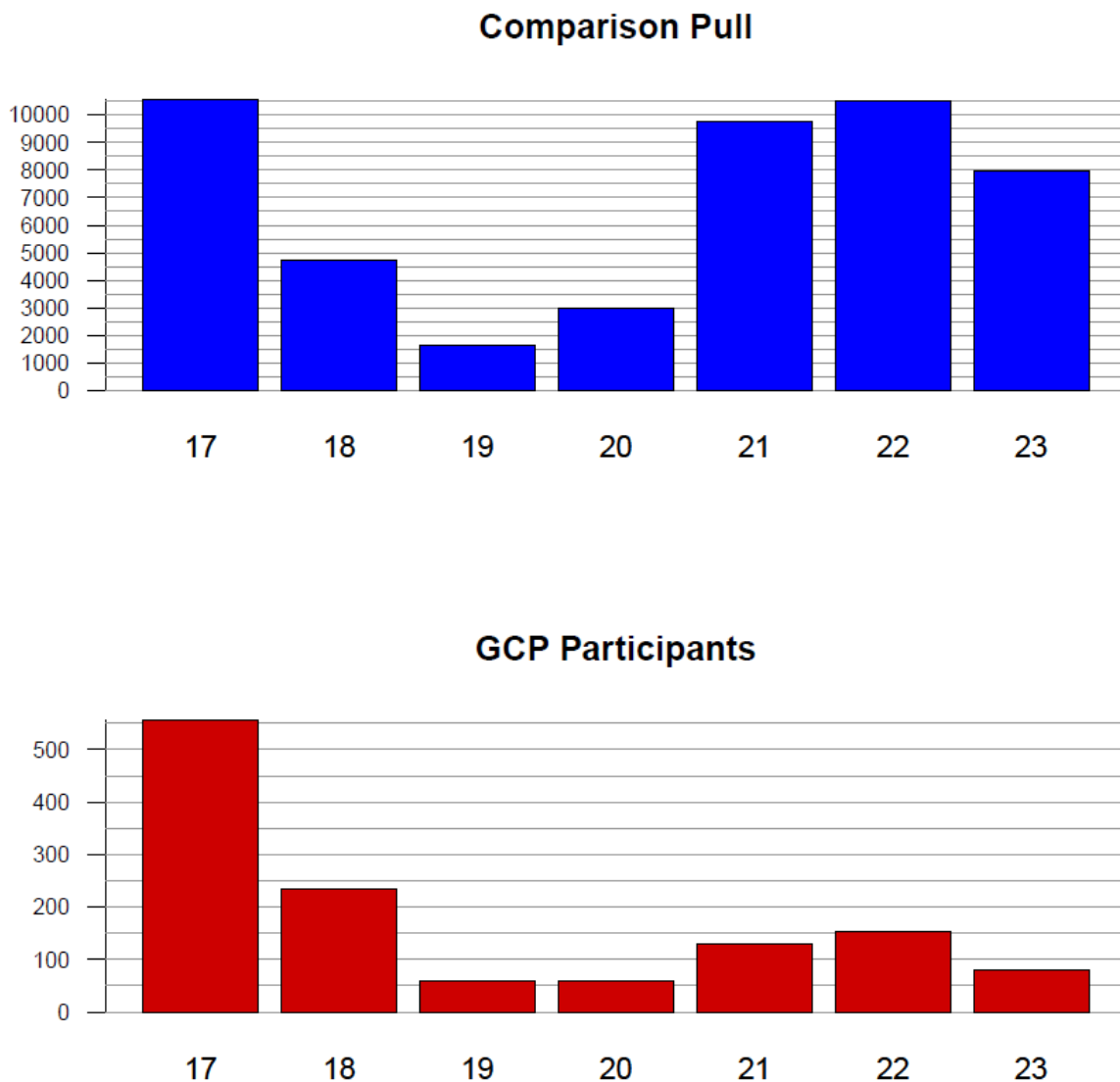
Appendix Table A16: STEM Studies and the Knowledge Economy

University Degree Field of Study	N	Share Working in the Knowledge Economy
	(1)	(2)
No STEM	27,128	20.02
Biological or Physical sciences	5,111	41.67
Math, CS, Stats	7,428	63.46
Engineering	13,607	57.09
Math, CS, Stats and Engineering	1,538	69.70
Number of Observations	54,812	

Notes: This table shows descriptive statistics on the interaction between STEM studies and working in the Knowledge Economy. The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Appendix Figure A1: Age of Taking UPET, GCP Participants and Comparison

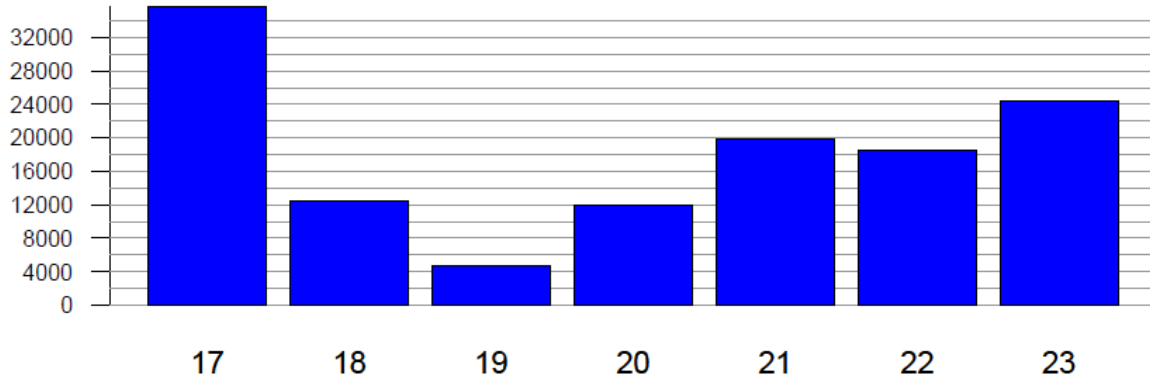
(A) 2006-2010 Graduates Sample



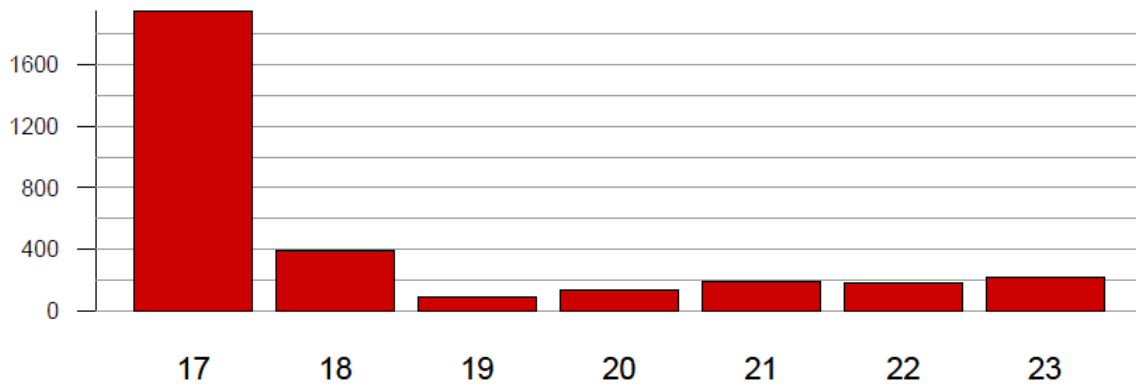
Notes: This figure plots the number of students who participated in the UPET, by grade of taking the test, for the comparison group pull, i.e., students in cities without a GCP (upper panel, in blue) and GCP participants (lower panel, in red). The sample includes only students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010. Lower and upper bars are binned, i.e., 17 represents 17 and below, and 23 represents 23 and above.

(B) 1992-2005 Graduates Sample

Comparison Pull



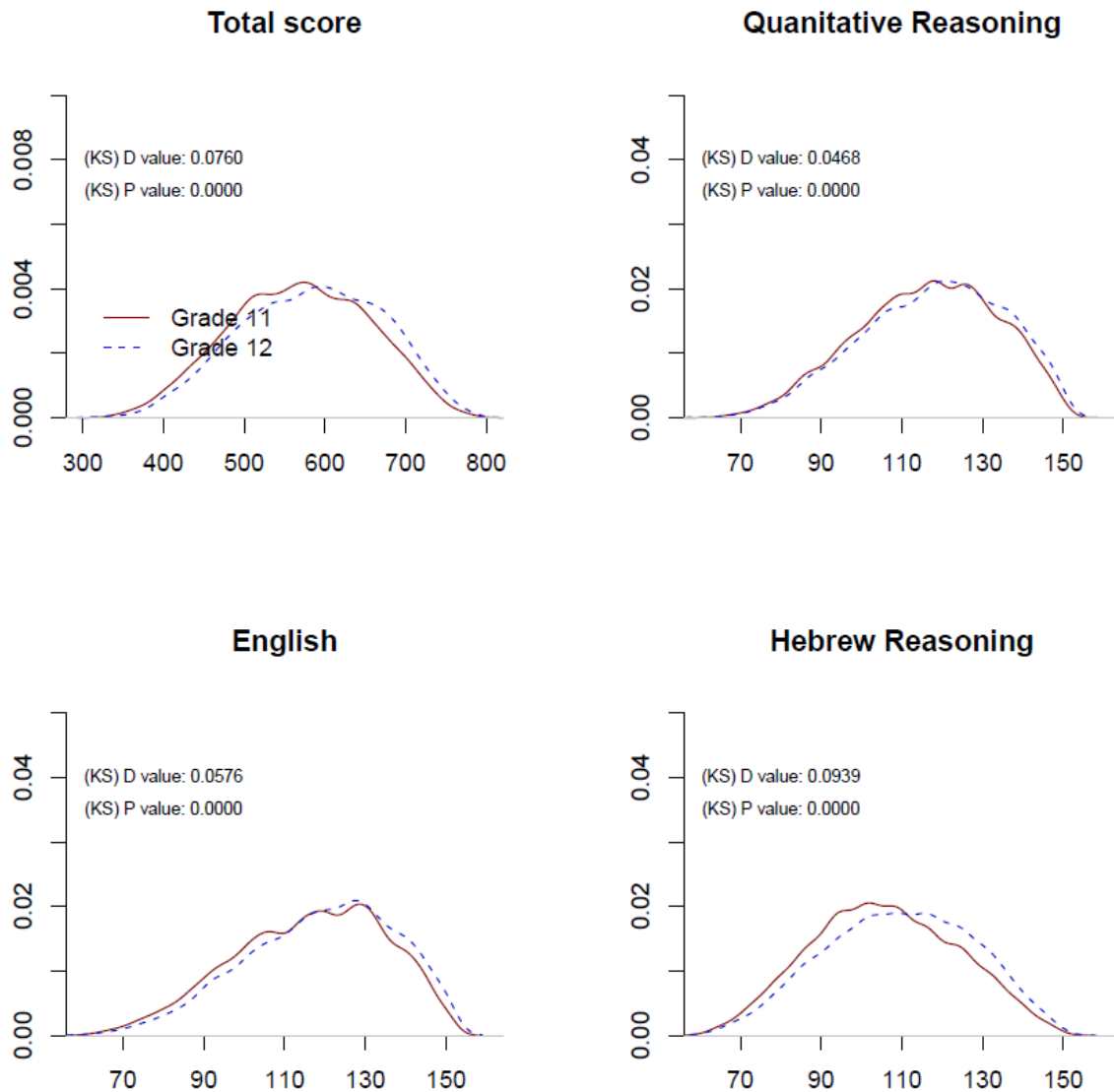
GCP Participants



Notes: This figure plots the number of students who participated in the UPET, by grade of taking the test, for the comparison group pull, i.e., students in cities without a GCP (upper panel, in blue) and GCP participants (lower panel, in red). The sample includes all students who took the UPET at any age, from the cohorts of high-school graduates in 1992-2005. Lower and upper bars are binned, i.e., 17 represents 17 and below, and 23 represents 23 and above.

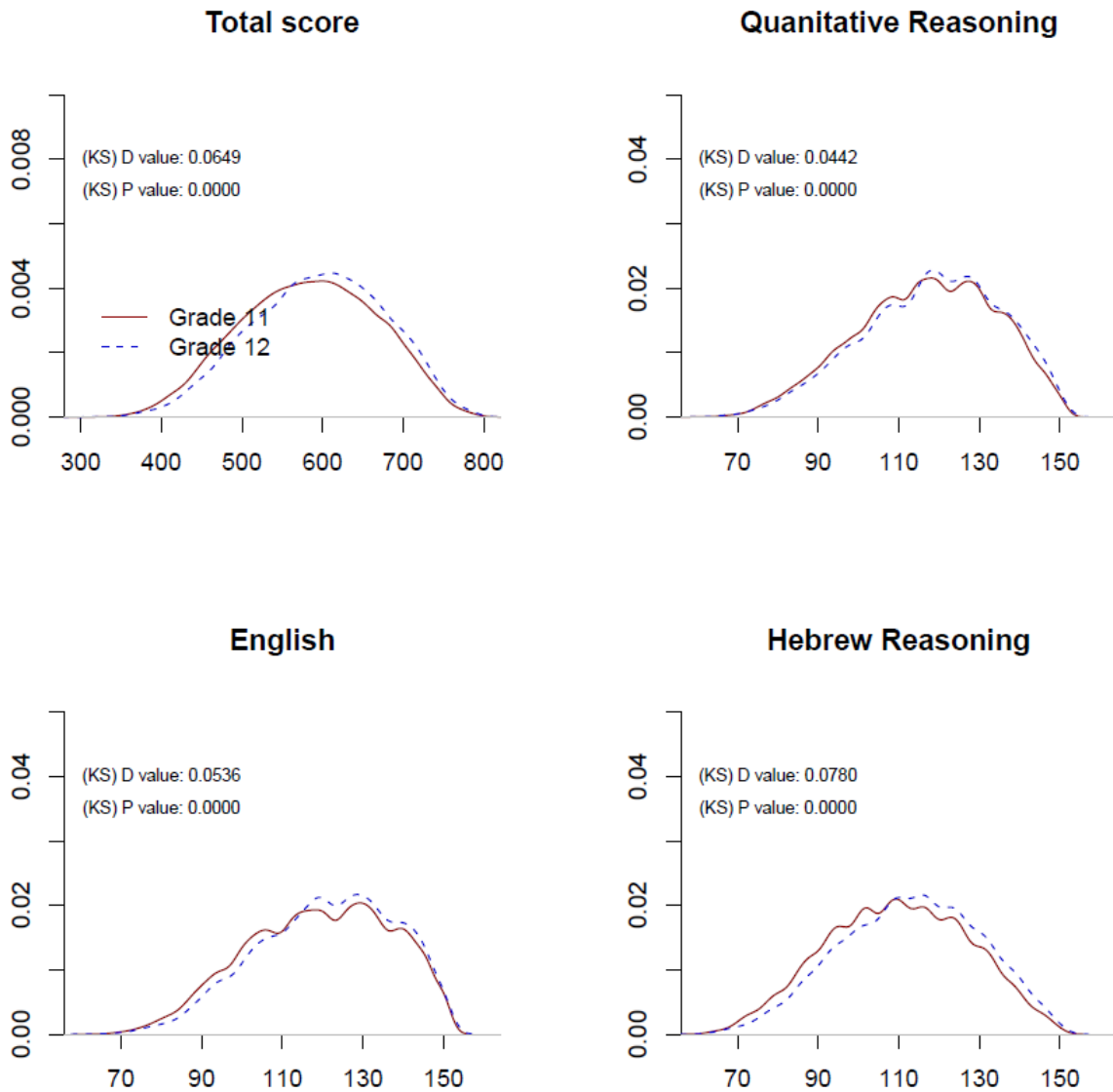
Appendix Figure A2: UPET Scores Distributions, by Grade of Testing, Non-Gifted

(A) 2006-2010 Graduates Sample



Notes: This figure plots the distribution of UPET scores by grade of taking the test- the solid red line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only (non-gifted) students who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010.

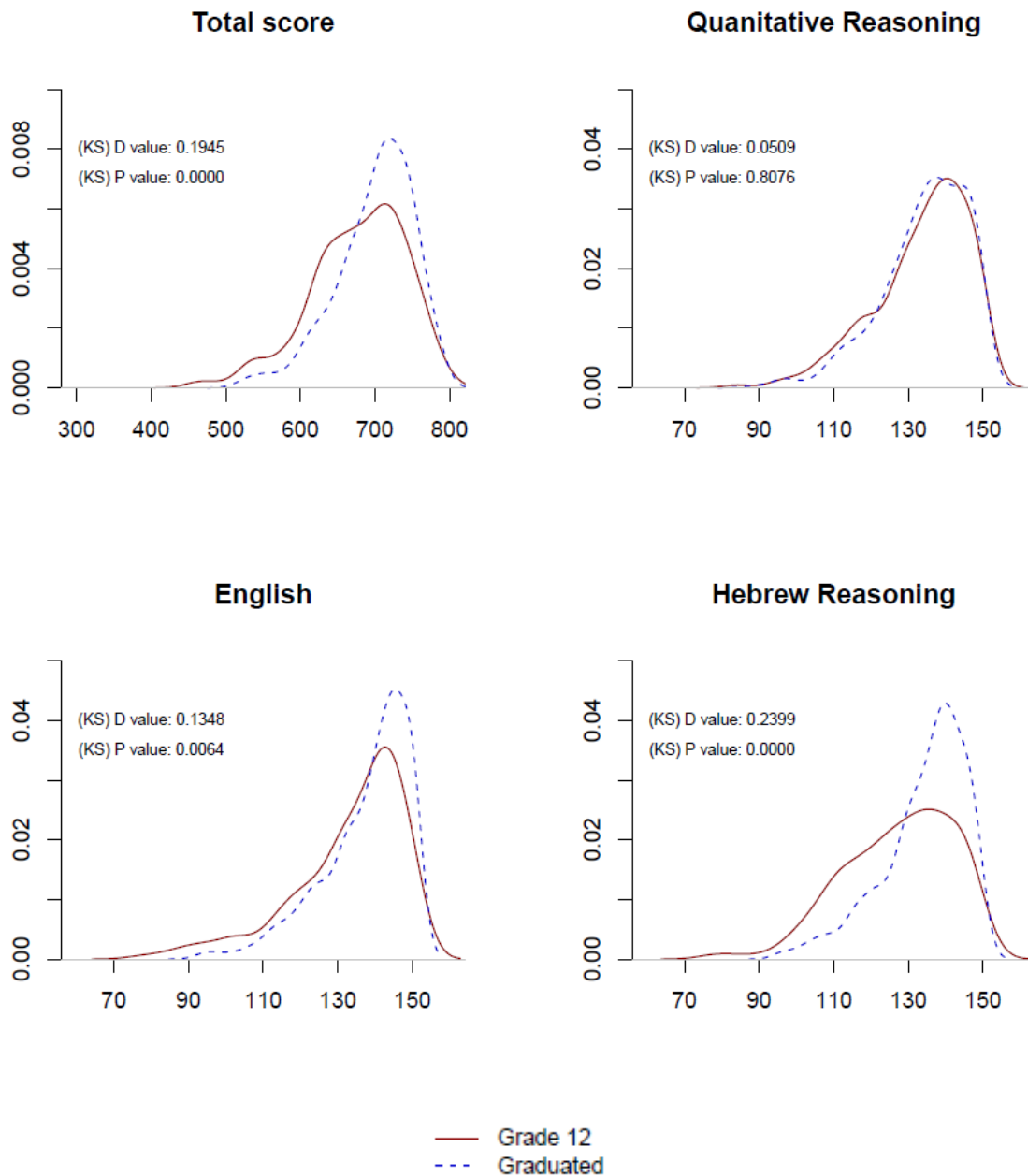
(B) 1992-2005 Graduates Sample



Notes: This figure plots the distribution of UPET scores by grade of taking the test- the solid red line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes (non-gifted) students in cities without a GCP from the cohorts of high-school graduates in 1992-2005.

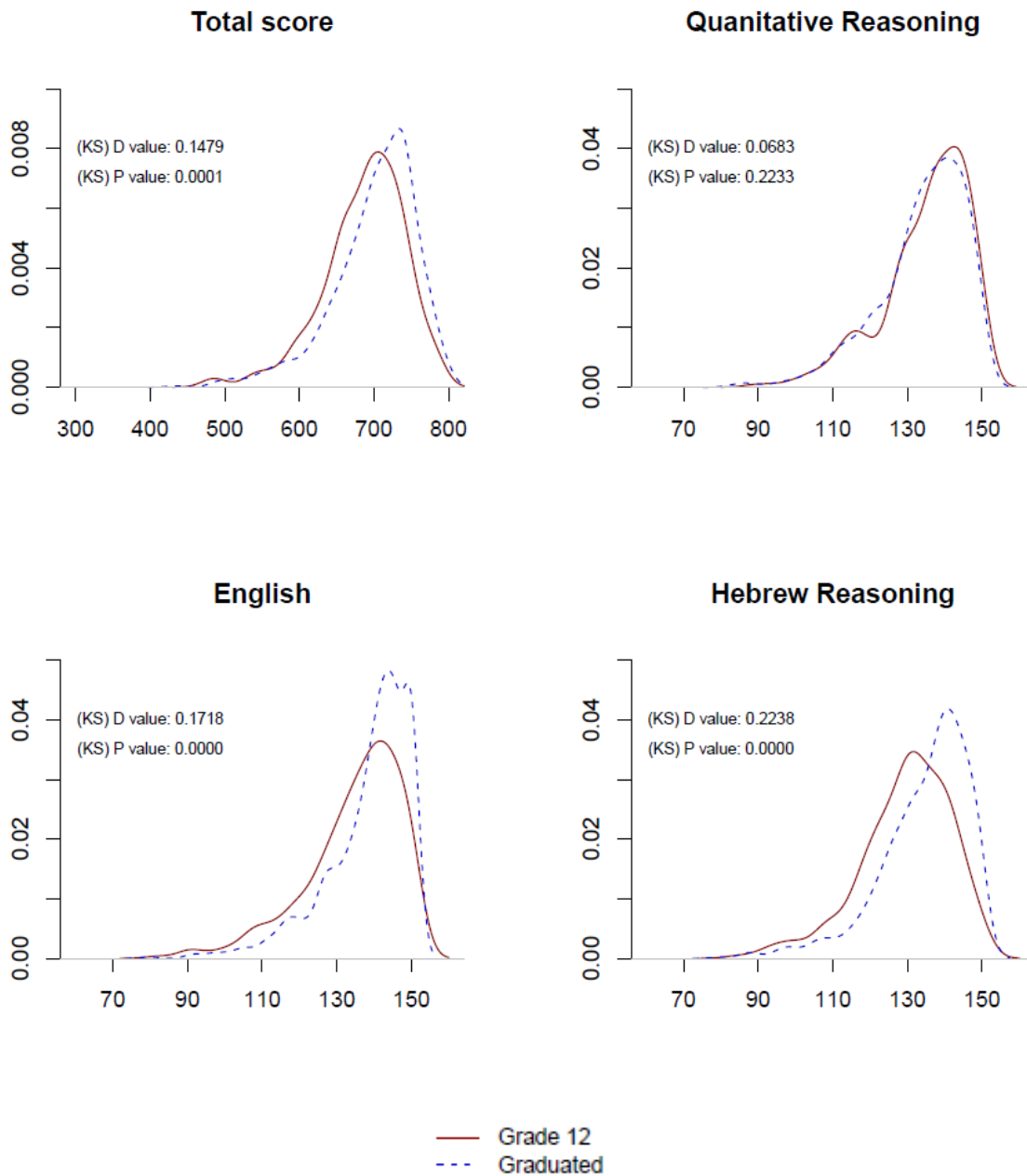
Appendix Figure A3: UPET Scores Distributions, by Grade of Testing, GCP Participants

(A) 2006-2010 Graduates Sample



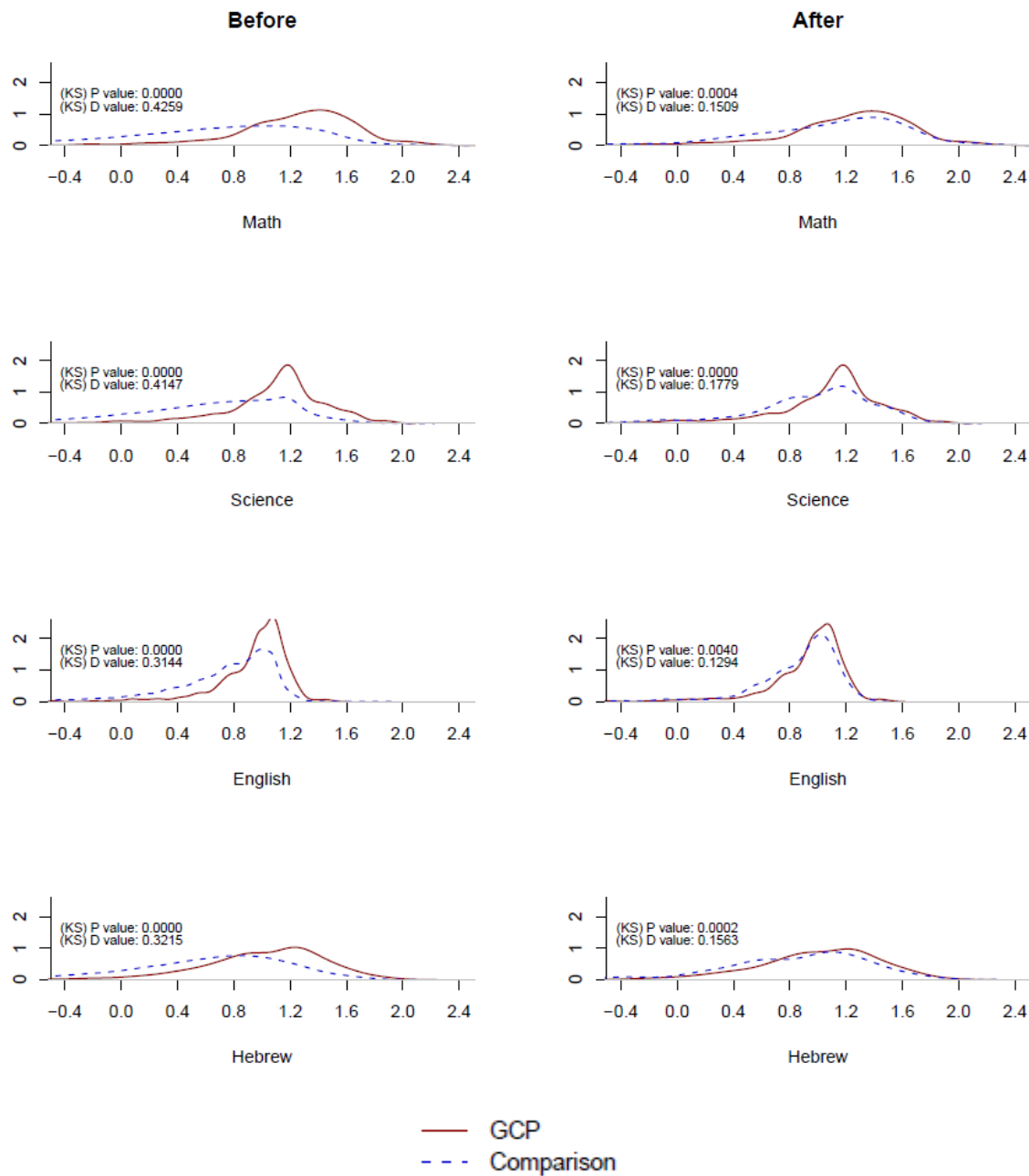
Notes: This figure plots the distribution of UPET scores by grade of taking the test. The solid red line represents the sample of students who took the test during their 12th grade, and the blue dashed line represents the sample of students who took the test during after their 12th grade (during the military service or after). The graph also shows the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes only GCP participants who participated in the Metzav middle school test during their 8th grade, about half of the students in cohorts of high-school graduates in 2006-2010.

(B) 1992-2005 Graduates Sample



Notes: This figure plots the distribution of UPET scores by grade of taking the test. The solid red line represents the sample of students who took the test during their 12th grade, and the blue dashed line represents the sample of students who took the test after 12th grade (during the military service or after). The graph also shows the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes only GCP participants from the cohorts of high-school graduates in 1992-2005.

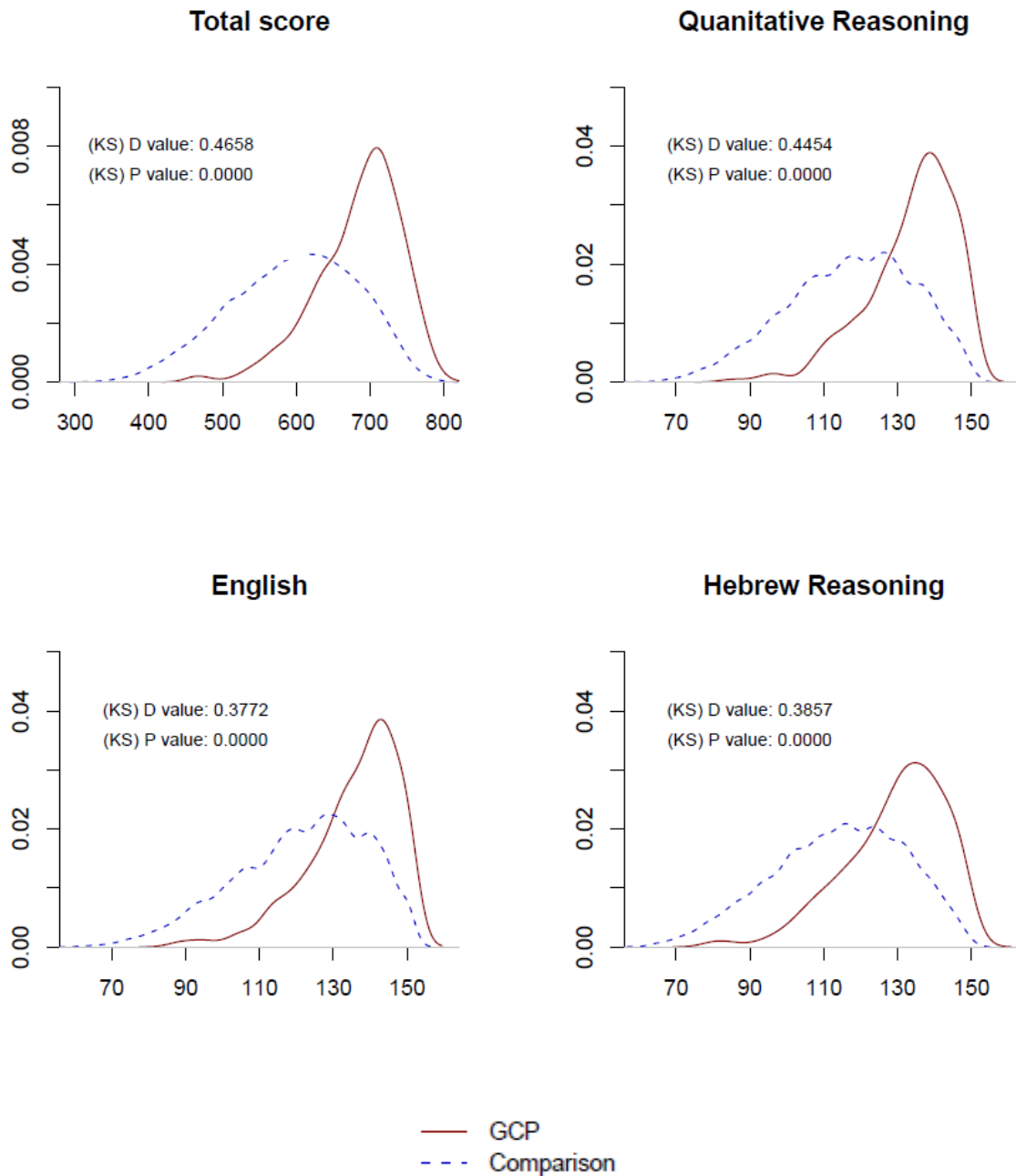
Appendix Figure A4: Pre-treatment Middle-school Test Scores, Before and After Matching II, 2006-2010 Sample



Notes: This figure plots the distribution of the Pre-treatment Middle-school Test (Metzav) test scores by groups. The solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graphs on the left show the distributions before the matching, and the graphs on the right show the distributions after the matching (version II, which does not use the Metzav scores in the logit specification). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010, who took the UPET at any age.

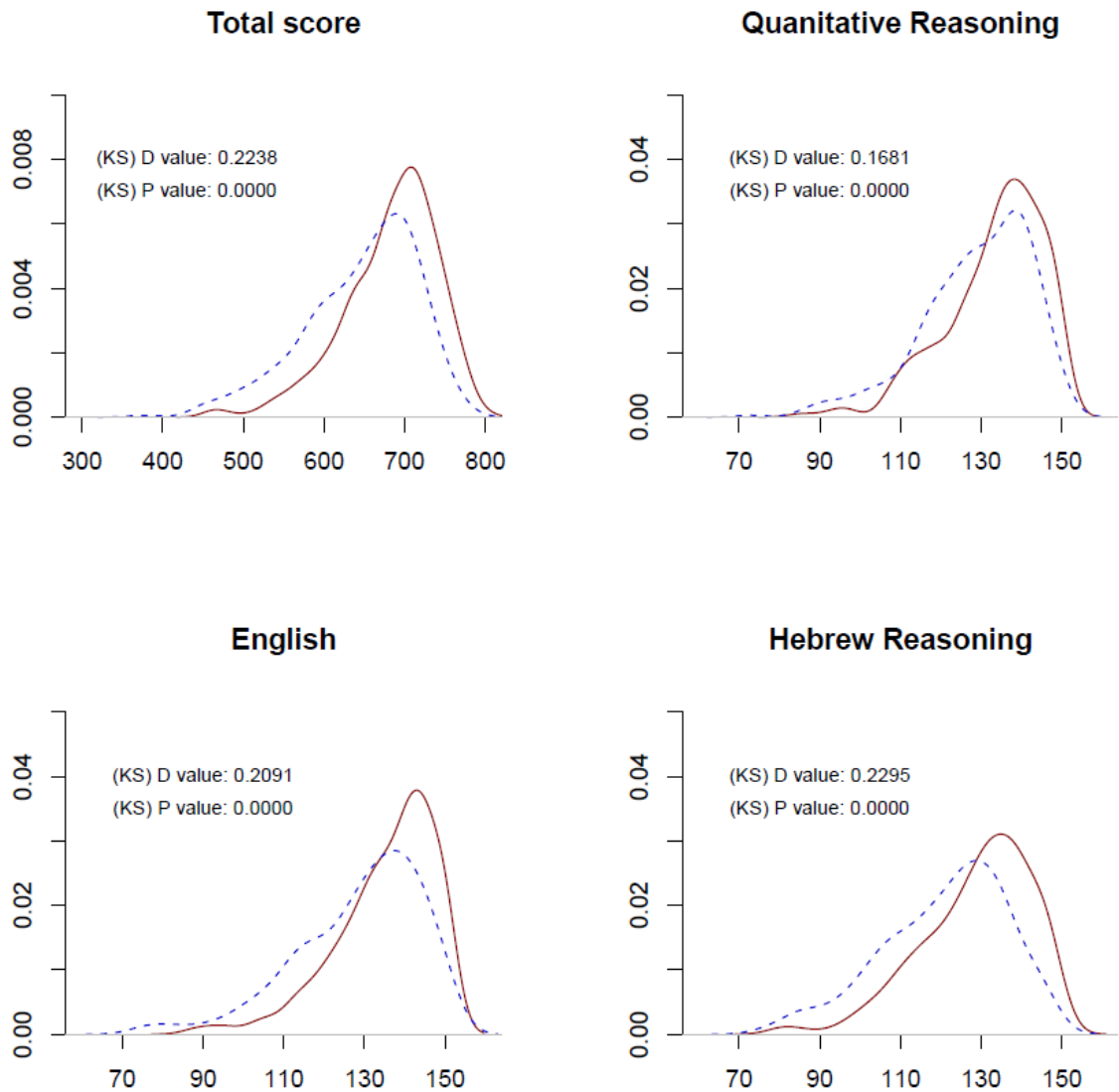
Appendix Figure A5: UPET Scores After Matching I, 2006-2010 Sample

(A) Before Matching



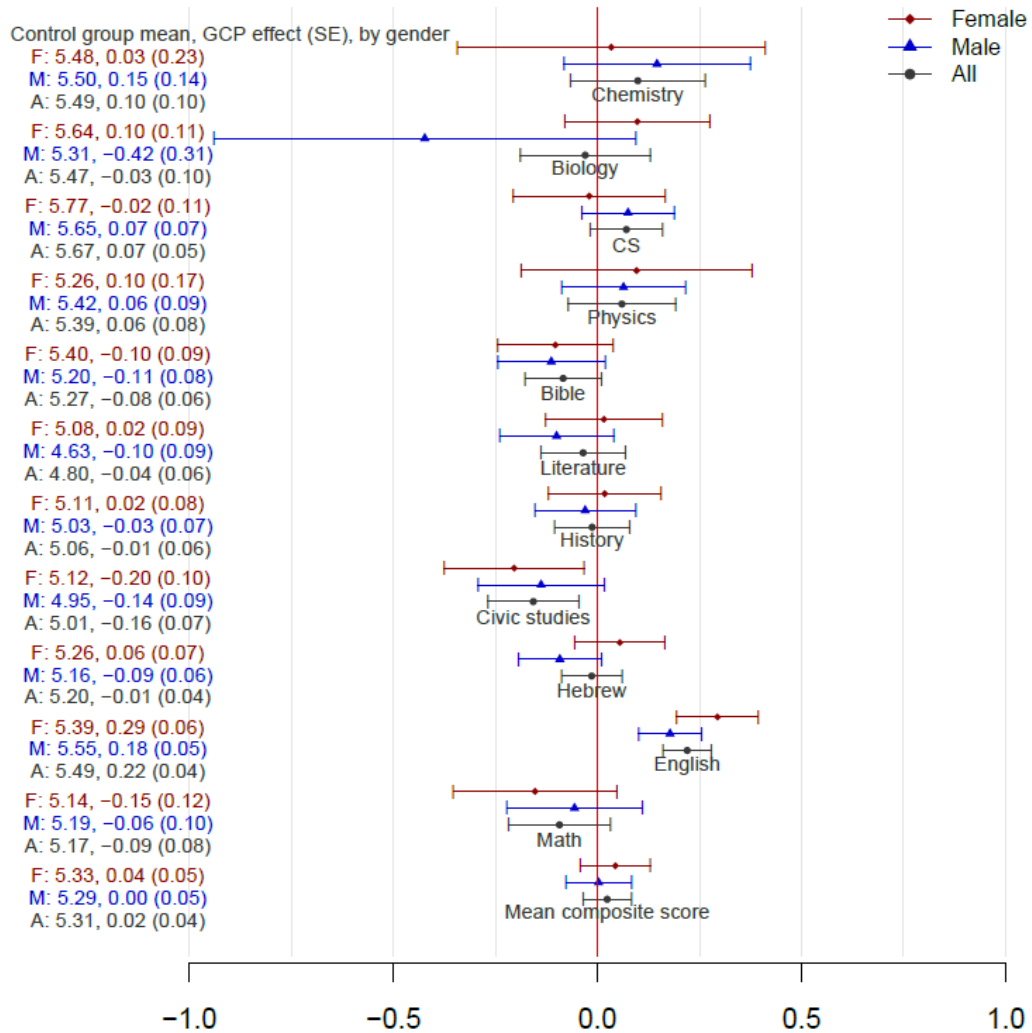
Notes: This figure plots the distribution of UPET scores by group- the solid red line represents the sample of GCP students, and the blue dashed line represents the comparison group pull before the matching. The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only students who participated in the Metzav middle school test during their 8th grade, which is about half of the students in cohorts of high-school graduates in 2006-2010.

(B) After Matching



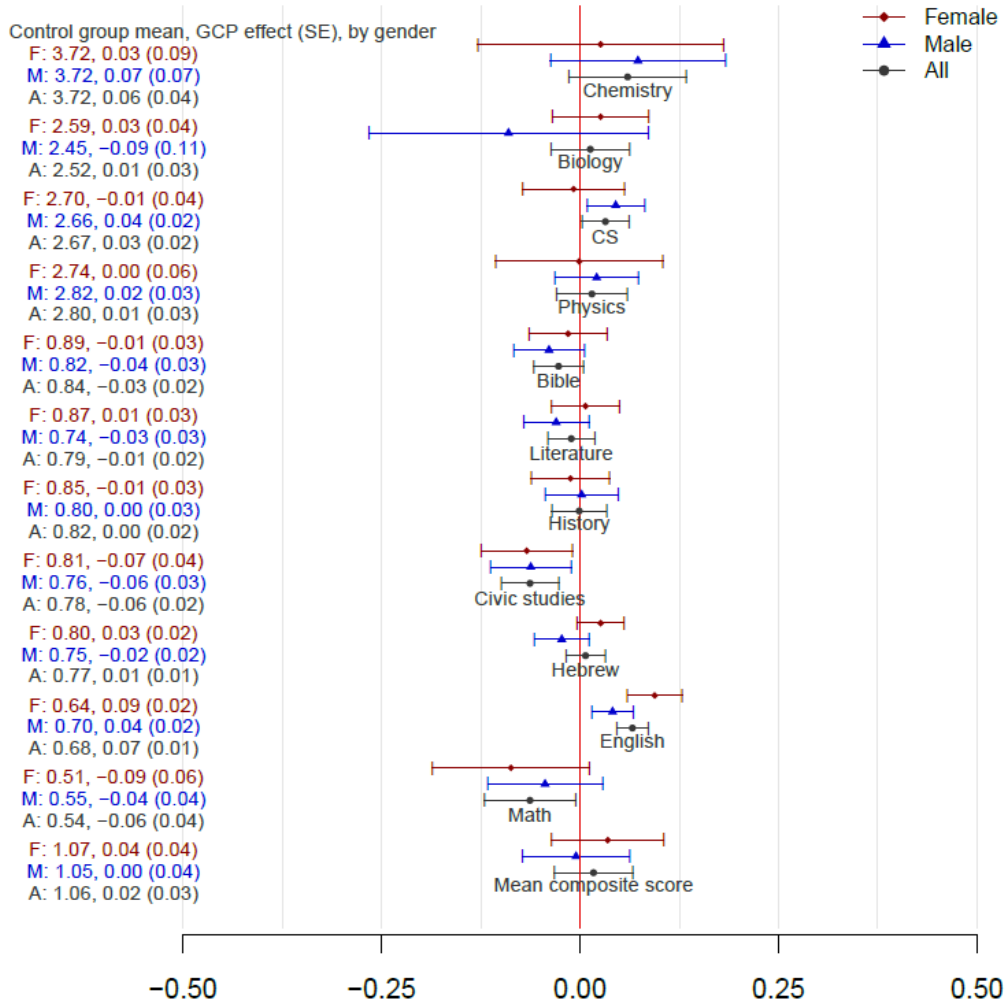
Notes: This figure plots the distribution of UPET scores by group- the solid red line represents the sample of GCP students, and the blue dashed line represents the matched comparison group (version I, which does not use the UPET scores in the logit specification). The graphs also show the Kolmogorov-Smirnov test for the equality of the probability distributions. The sample includes only students who participated in the Metzav middle school test during their 8th grade, which is about half of the students in cohorts of high-school graduates in 2006-2010.

Appendix Figure A6: GCP Effects on Bagrut Test Scores (1-6 Scores), 2006-2010 Sample



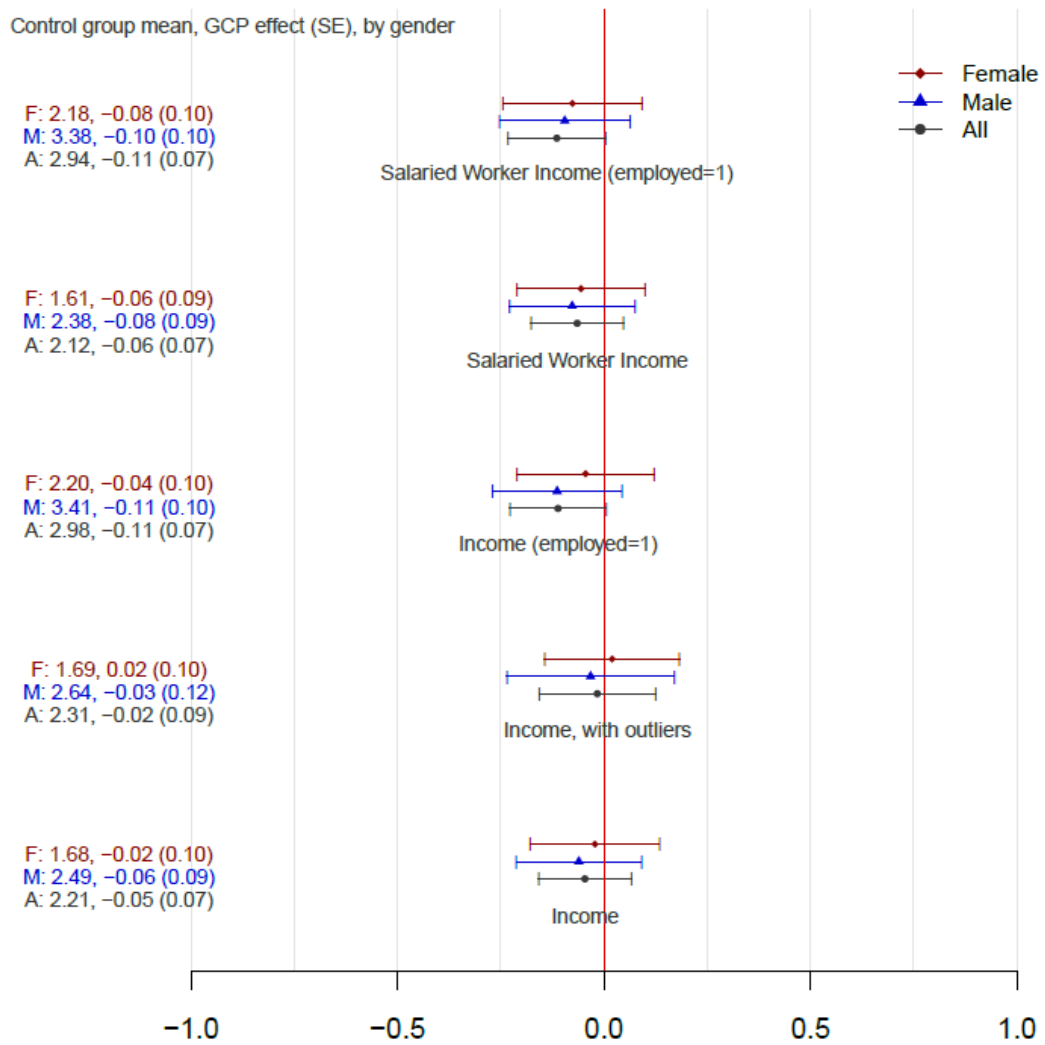
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory and most chosen elective subjects). These scores are normalized on a scale of 0-6. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

Appendix Figure A7: GCP Effects on Bagrut Test Scores (z Scores), 2006-2010 Sample



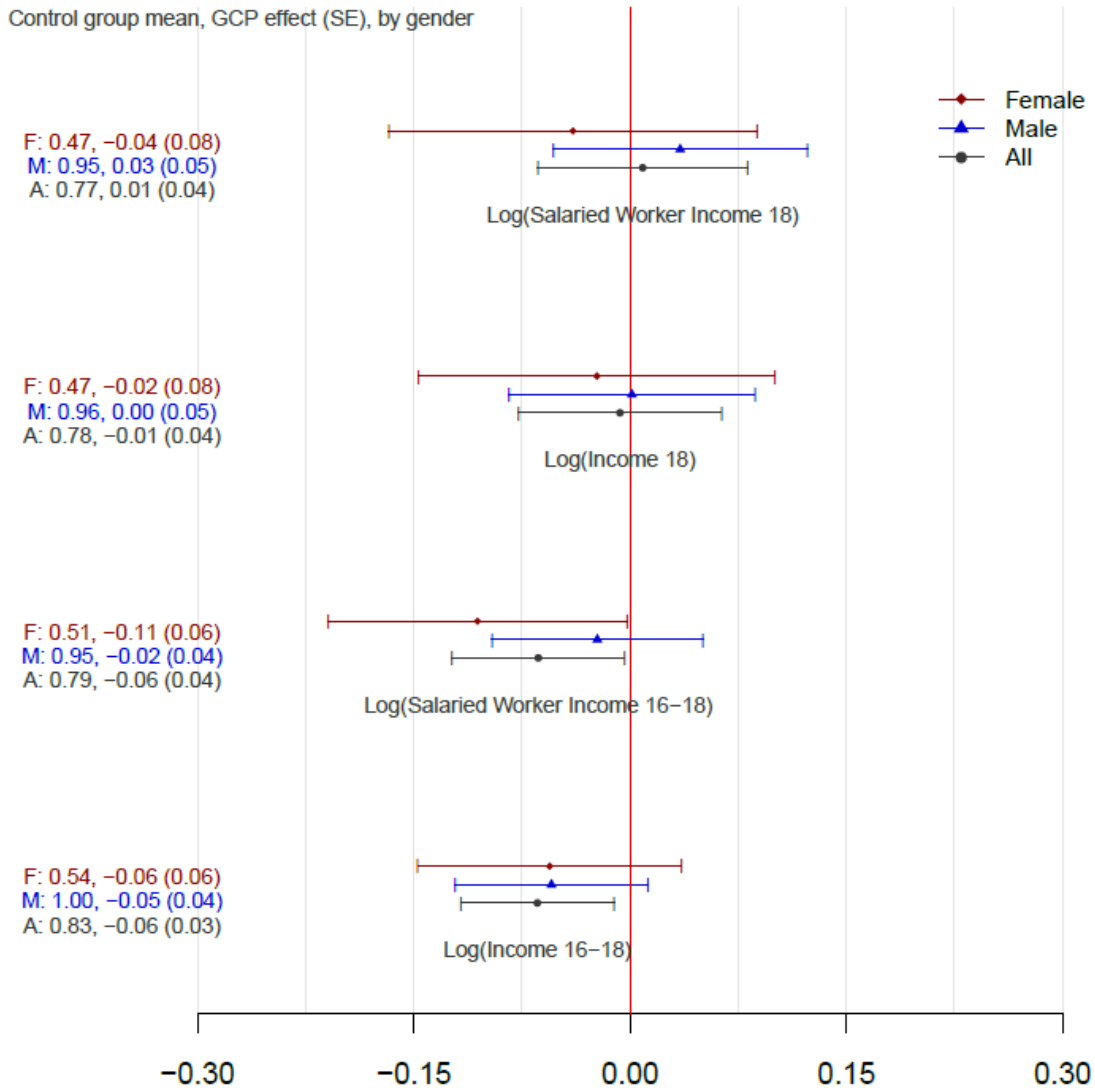
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory and most chosen elective subjects). Scores are normalized based on the normal distribution (yearly). Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students who participated in the 8th-grade Metzav tests, about half of the students in cohorts of high-school graduates in 2006-2010.

Appendix Figure A8: GCP Effects on Average Annual Income, in 2016-2018



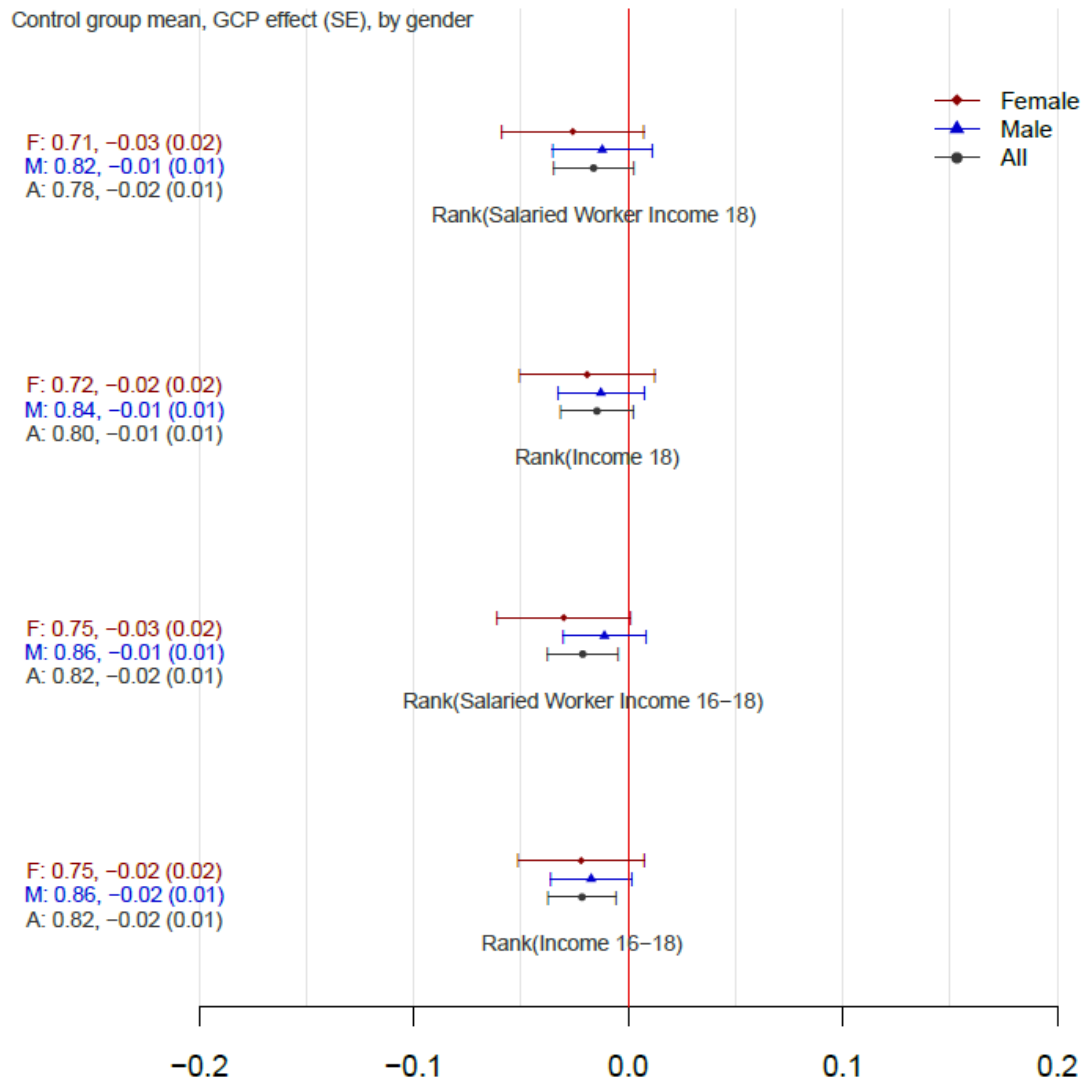
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on average annual income (in 100K NIS), in 2016-2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Appendix Figure A9: GCP Effects on the Log of Average Annual Income



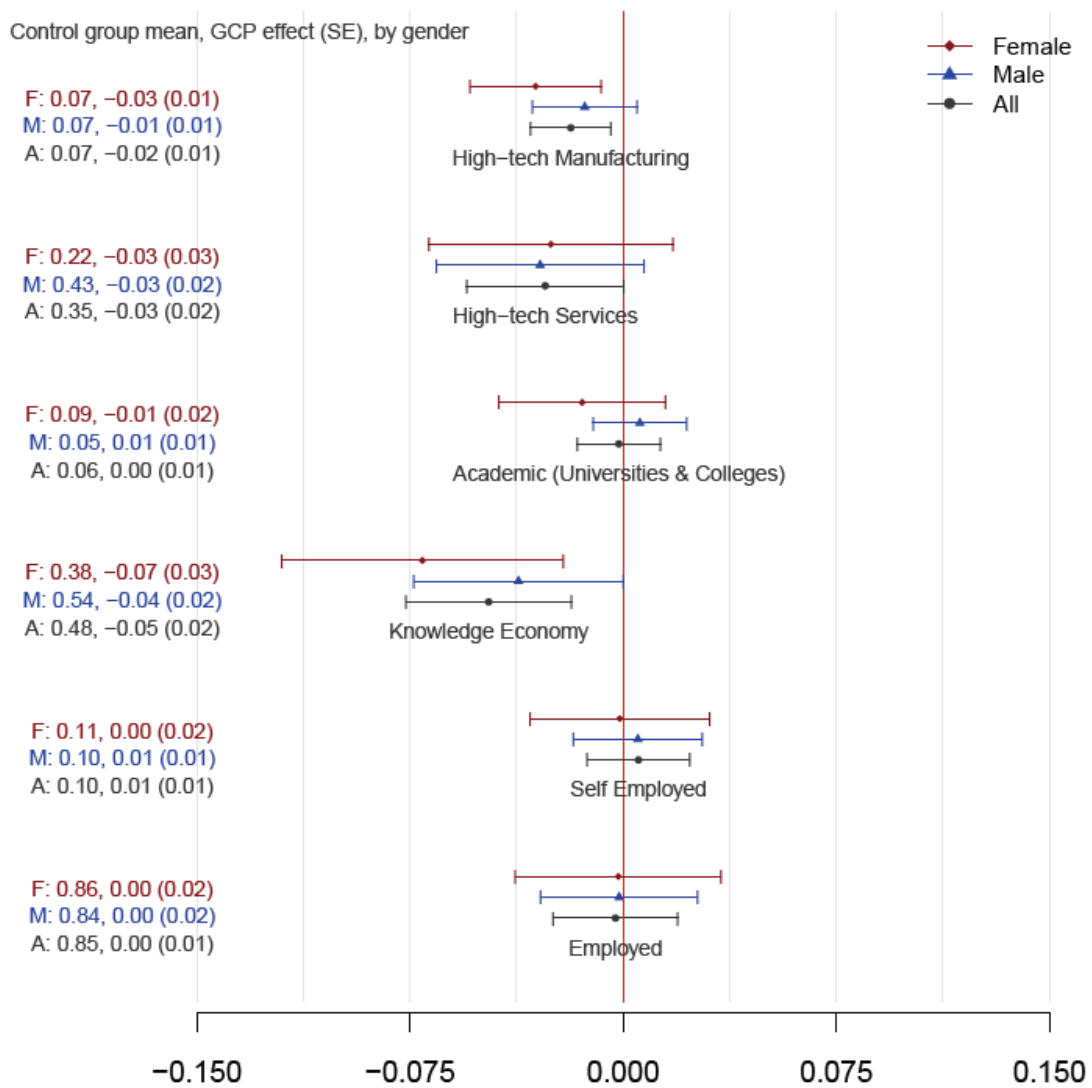
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on the natural log of average annual income (in 100K NIS) in 2016-2018. Red dots and lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Appendix Figure A10: GCP Effects on the Rank of Average Annual Income



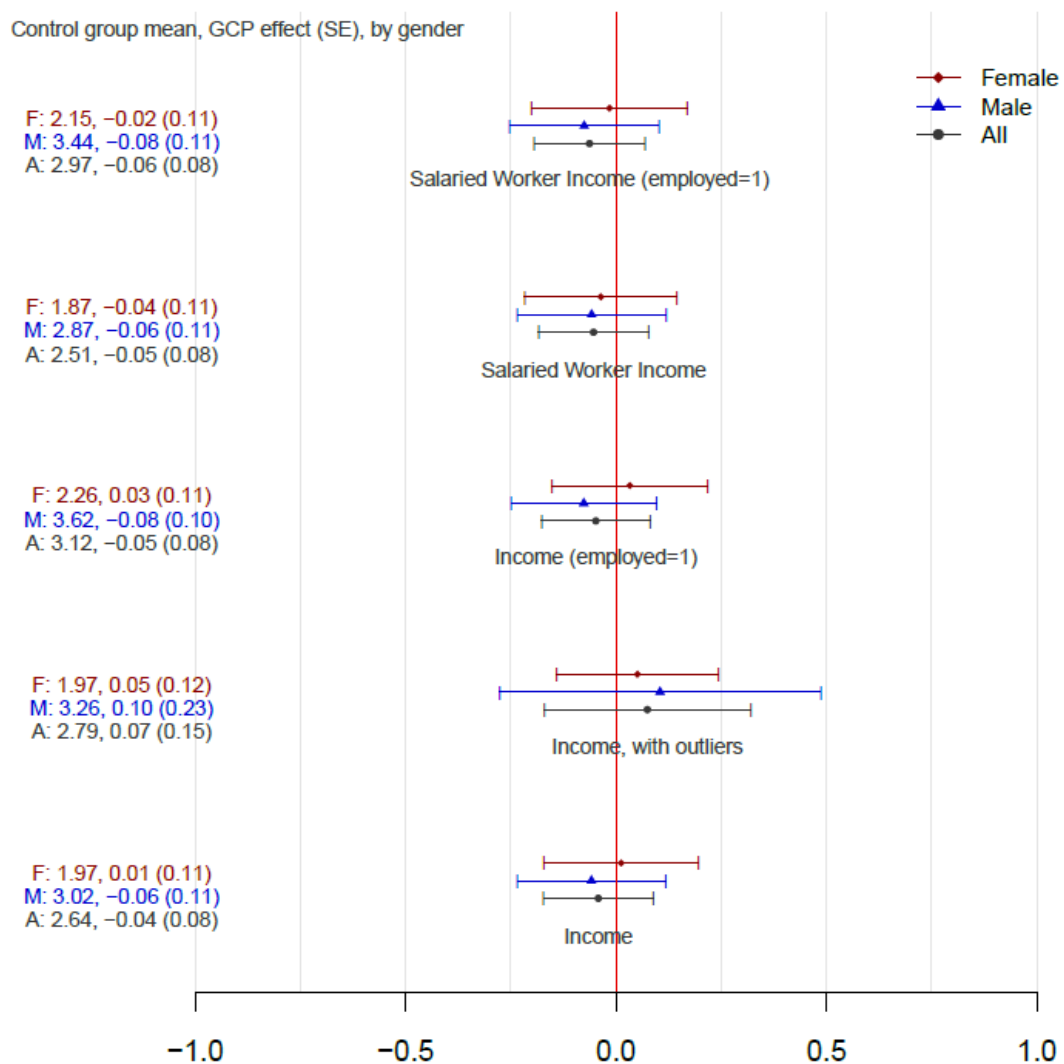
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on the rank average annual income in 2016-2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure A11: GCP Effect on Employment, by Sector in 2018, 1992-2005 Sample (Live in Israel in 2018)



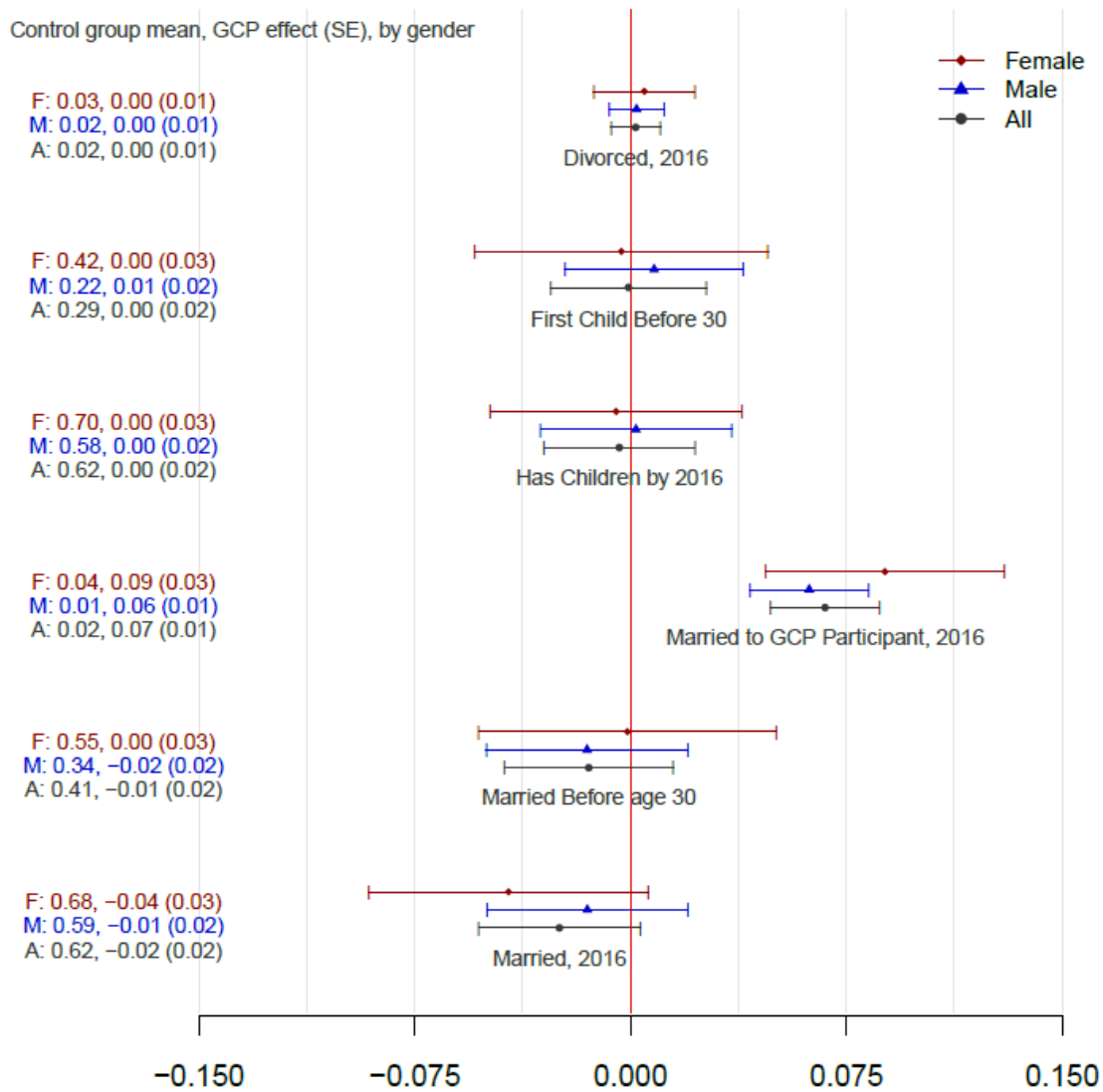
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on employment by Sector in 2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure A12: GCP Effect on Annual Income, in 2018, 1992-2005 Sample (Live in Israel in 2018)



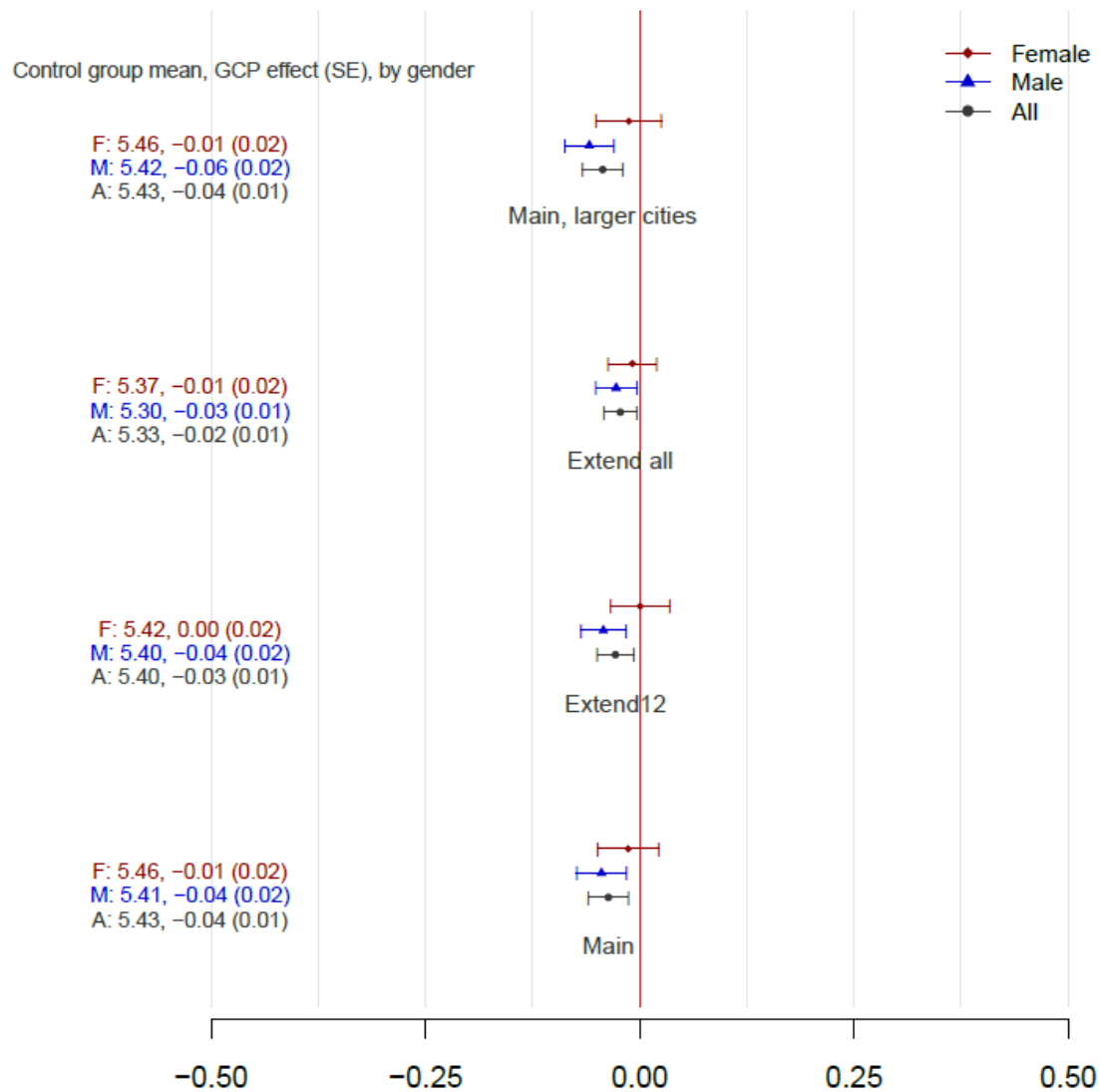
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on annual income in 2018. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Figure A13: GCP Effect on Personal Outcomes, 1992-2005 Sample (Live in Israel in 2018)



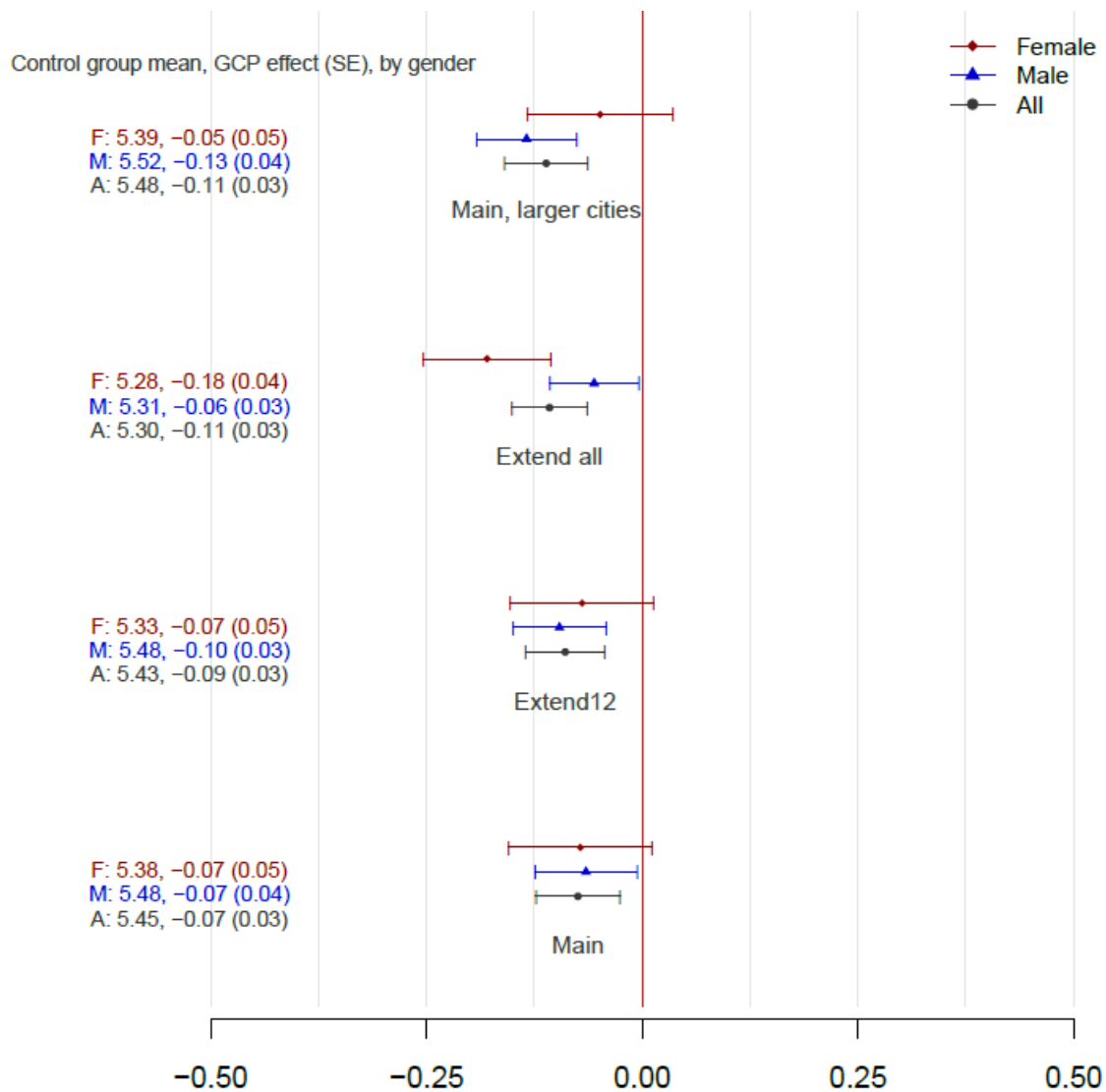
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on personal outcomes. Red lines represent the sample of females, blue lines represent the sample of males, and dark grey lines represent the entire sample (males and females). The sample includes students from the cohorts of high-school graduates in 1992-2005 who took the UPET during their 10-11th grade.

Appendix Figure A14: GCP Effects on Bagrut Mean Composite Score, By Sample



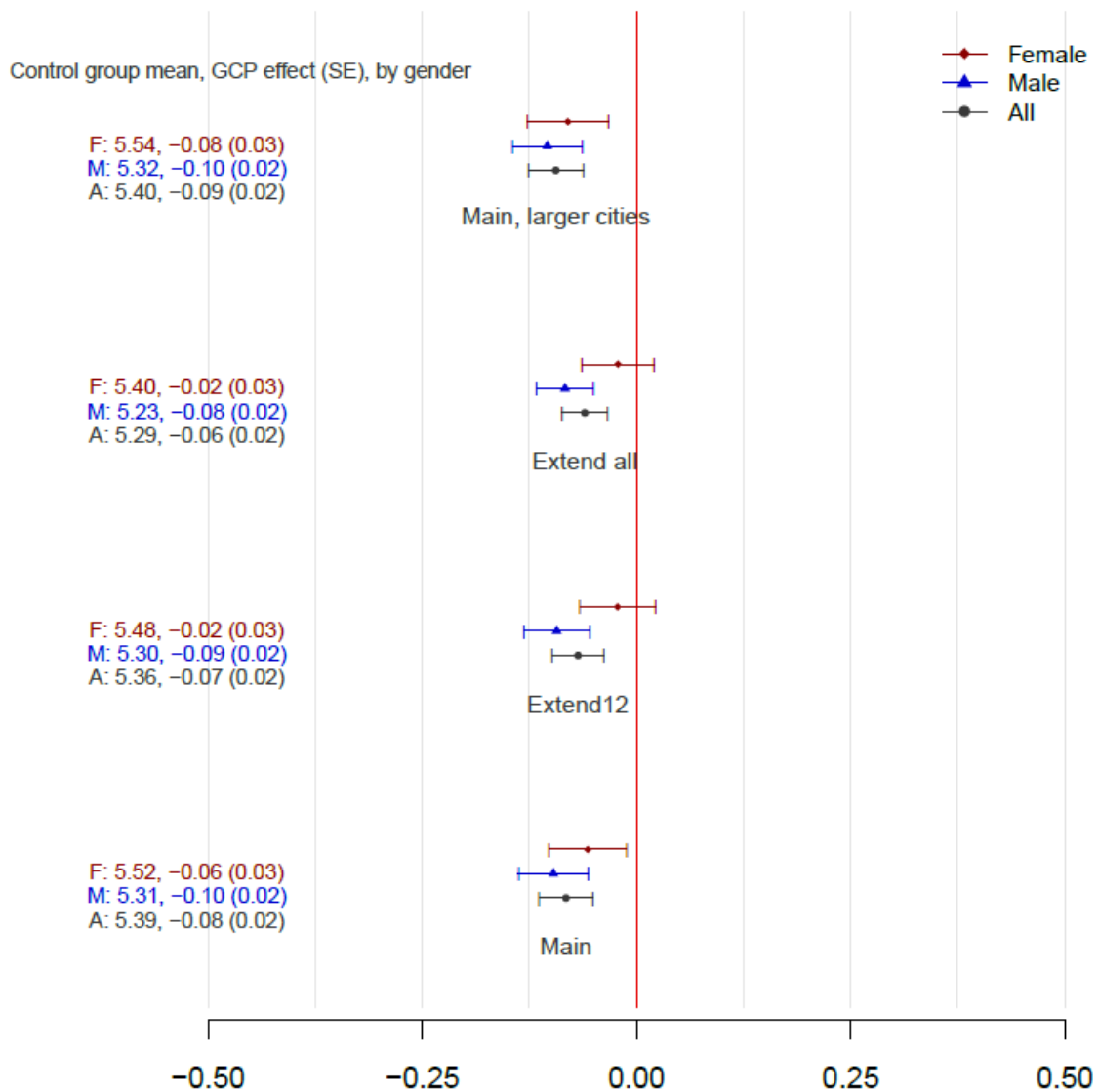
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on *bagrut* mean composite score, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A15: GCP Effects on Bagrut Math Score, By Sample



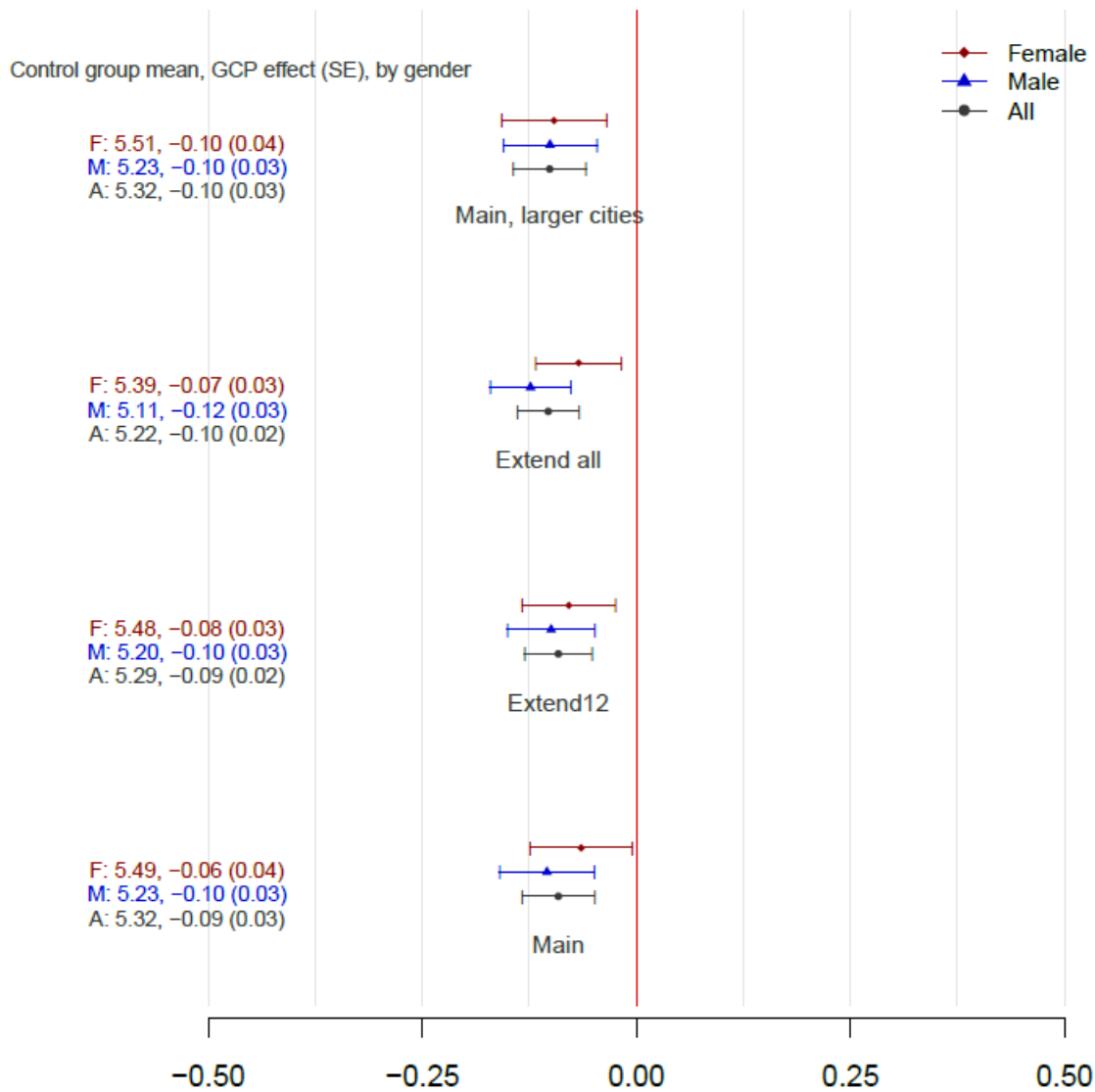
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on *bagrut* math score, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A16: GCP Effects on Bagrut Hebrew Score, By Sample



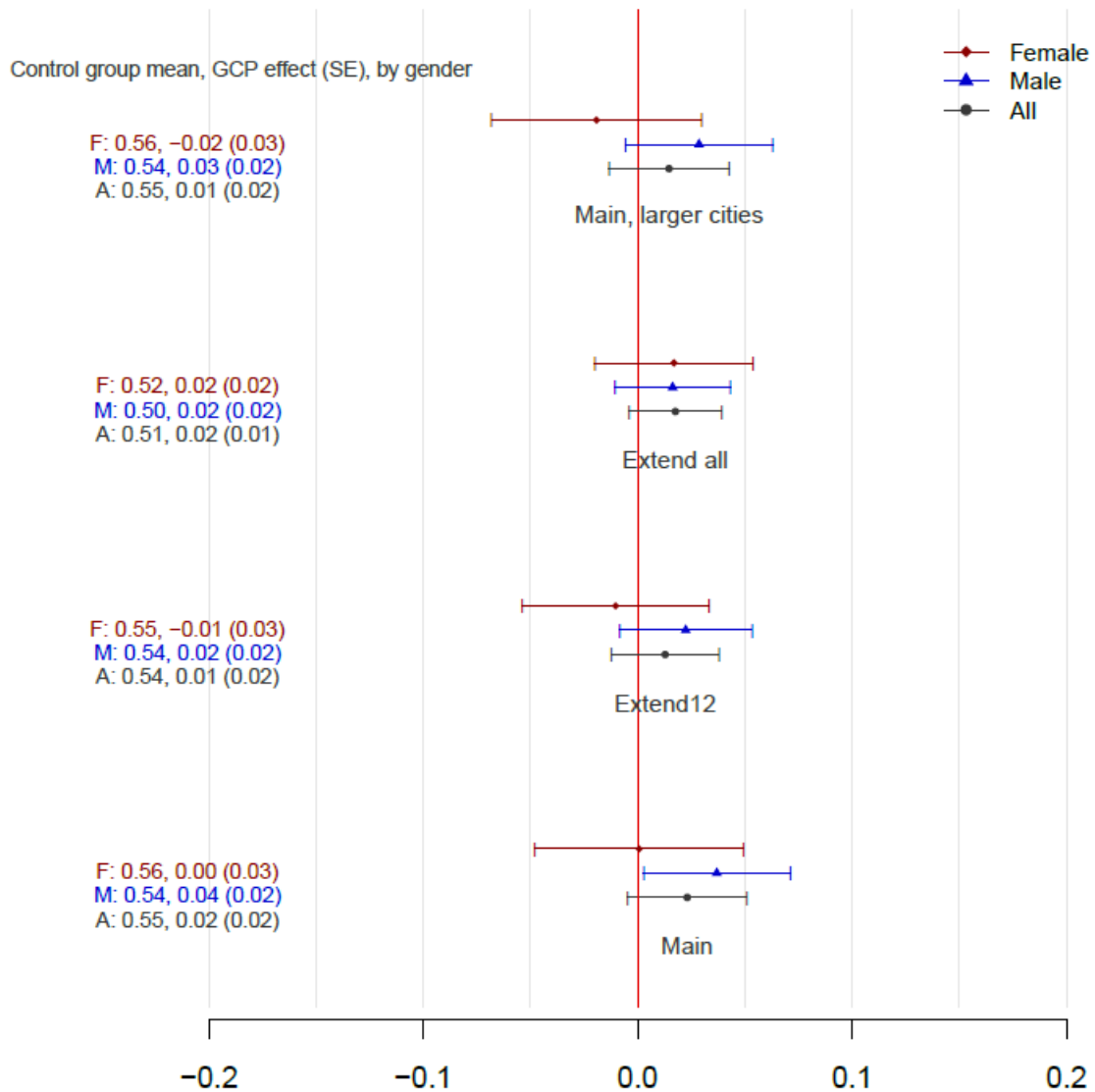
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on *bagrut* Hebrew score, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A17: GCP Effects on Bagrut Bible Score, By Sample



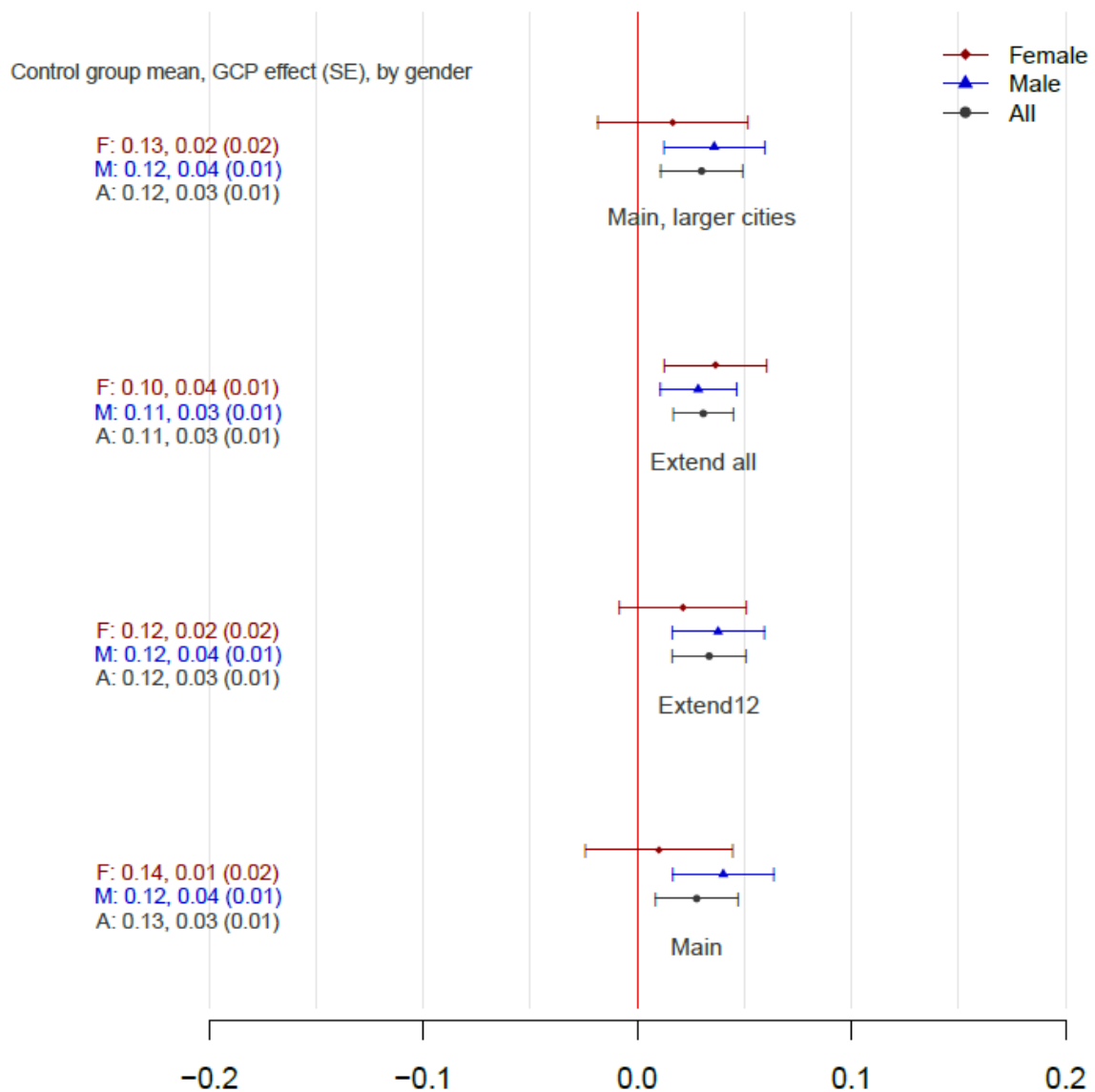
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on *bagrut* Bible score, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A18: GCP Effects on MA Degree Attainment, By Sample



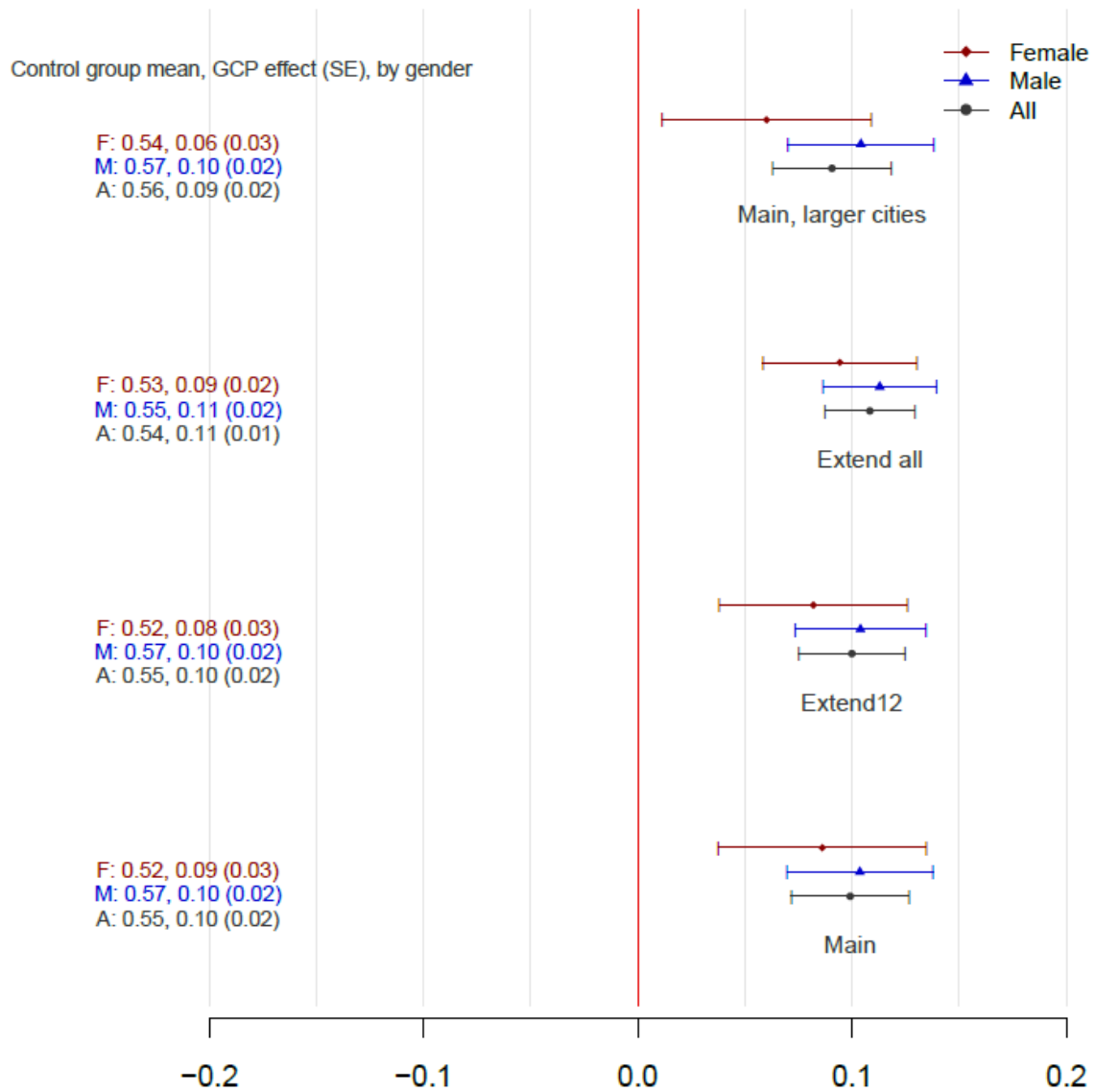
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on MA degree attainment, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A19: GCP Effects on Ph.D. Degree Attainment, By Sample



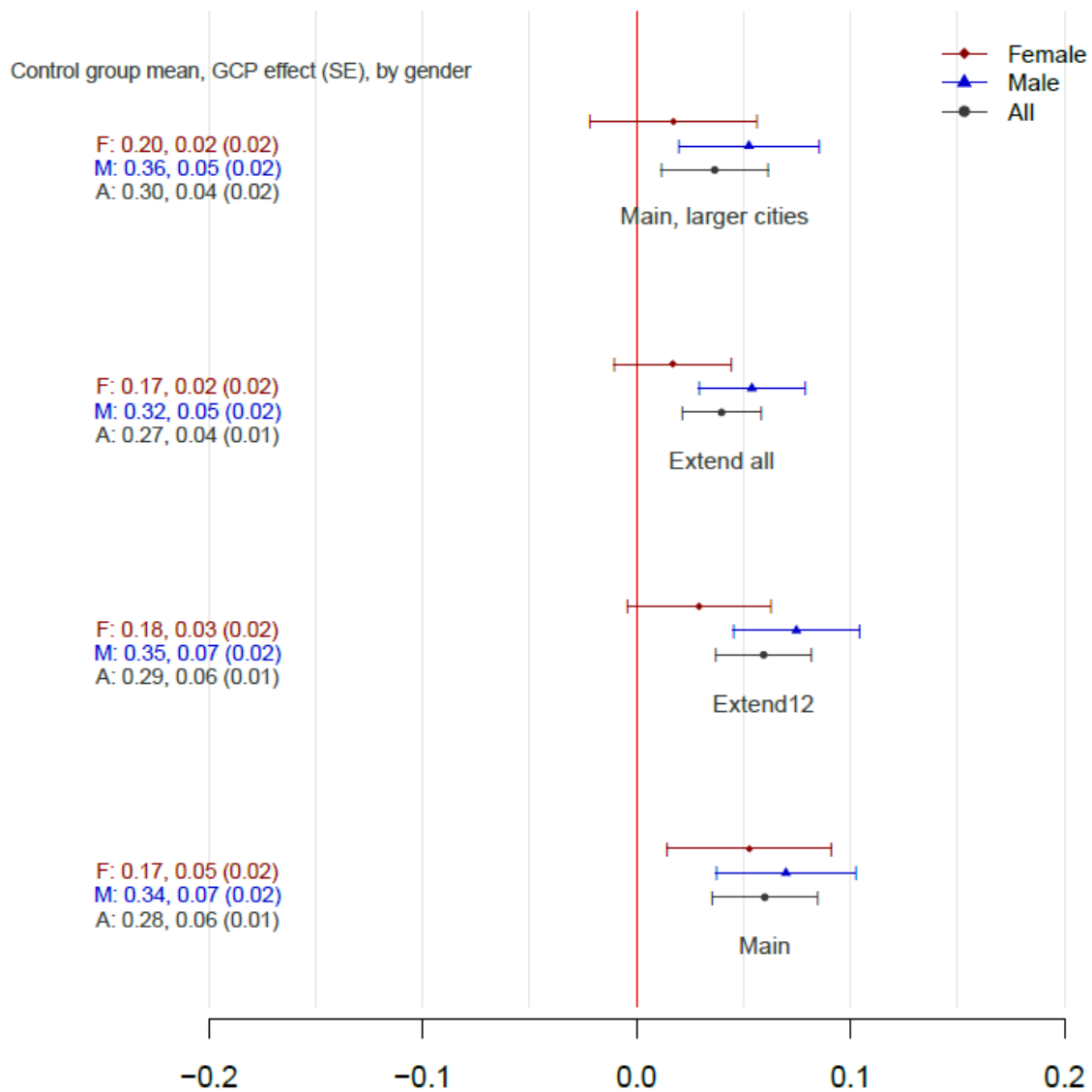
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Ph.D. degree attainment, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A20: GCP Effects on Double Major BA, By Sample



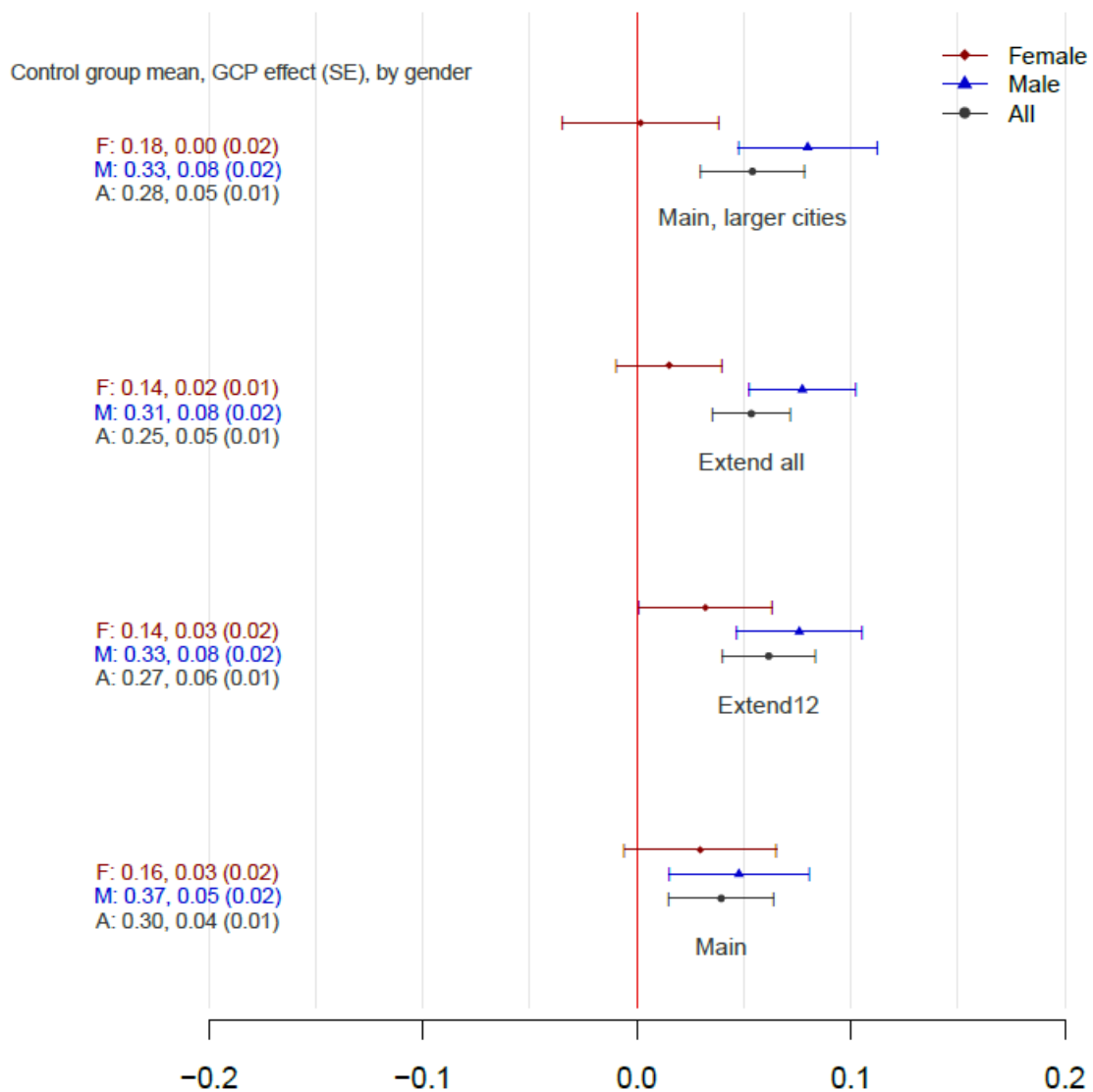
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on double Major in BA degree attainment, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A21: GCP Effects on STEM Double Major BA, By Sample



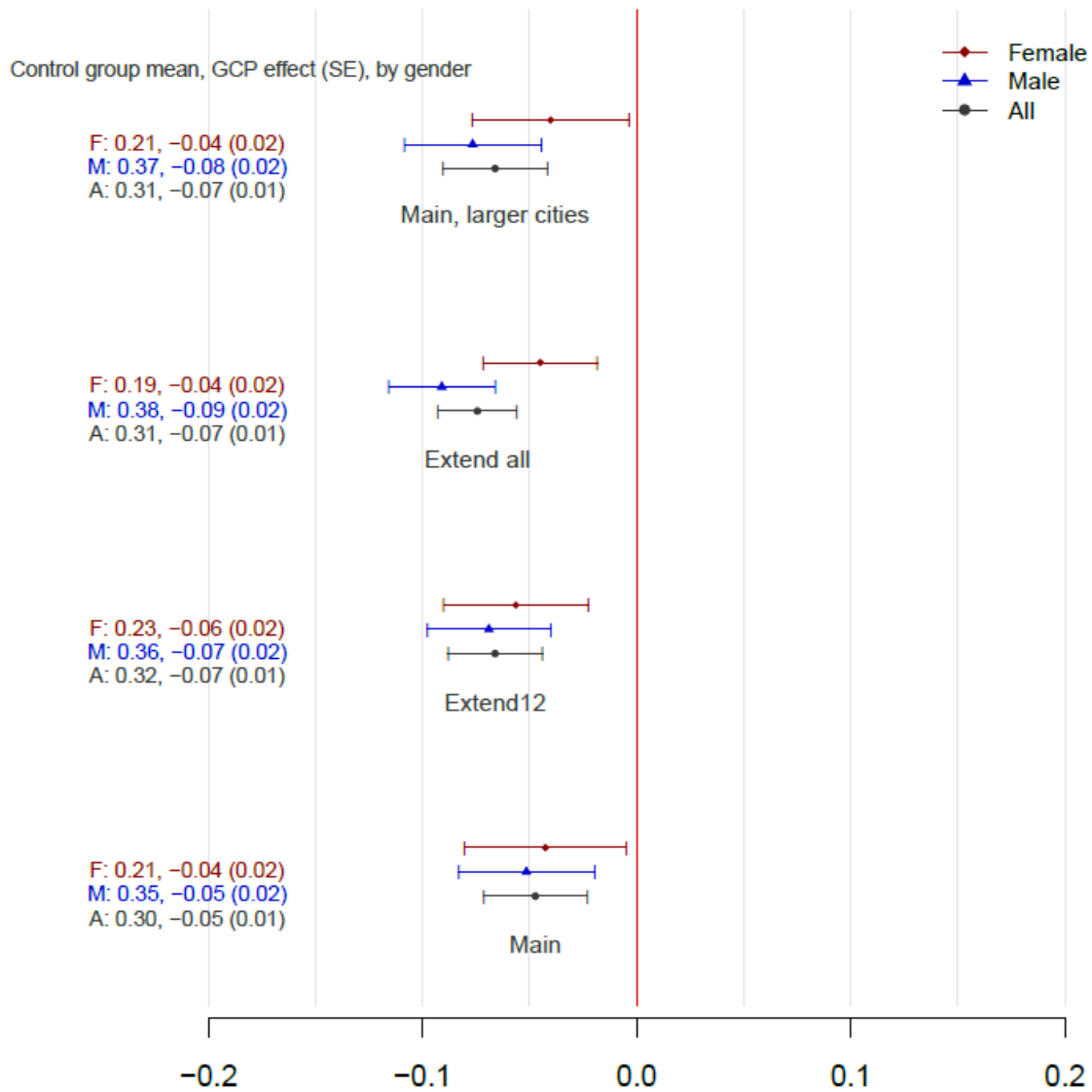
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on stem double major BA degree attainment, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A22: GCP Effects on BA in Math, Statistics, and Computer Sciences, By Sample



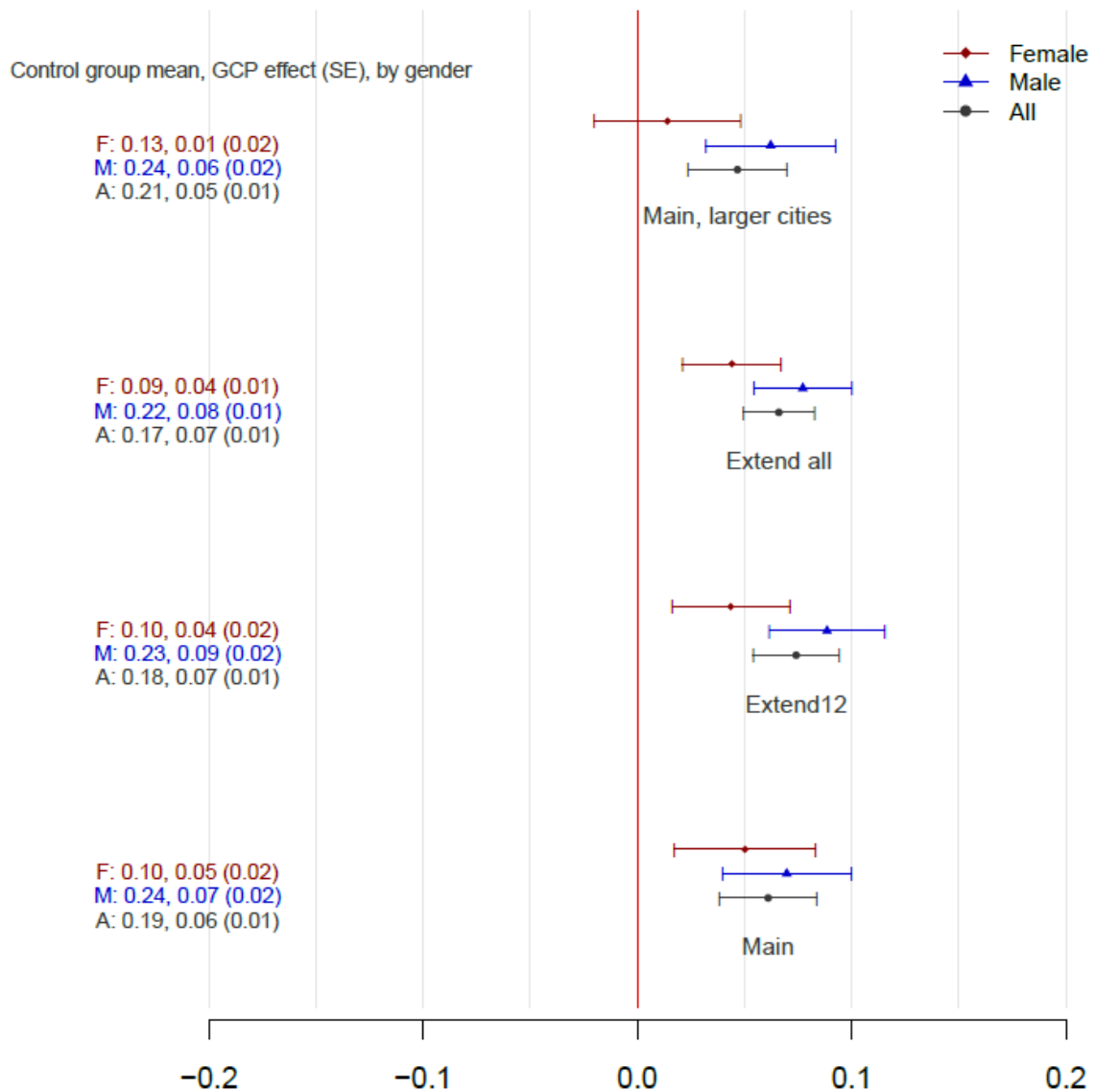
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Math, Statistics, and Computer Sciences, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A23: GCP Effects on BA in Engineering, By Sample



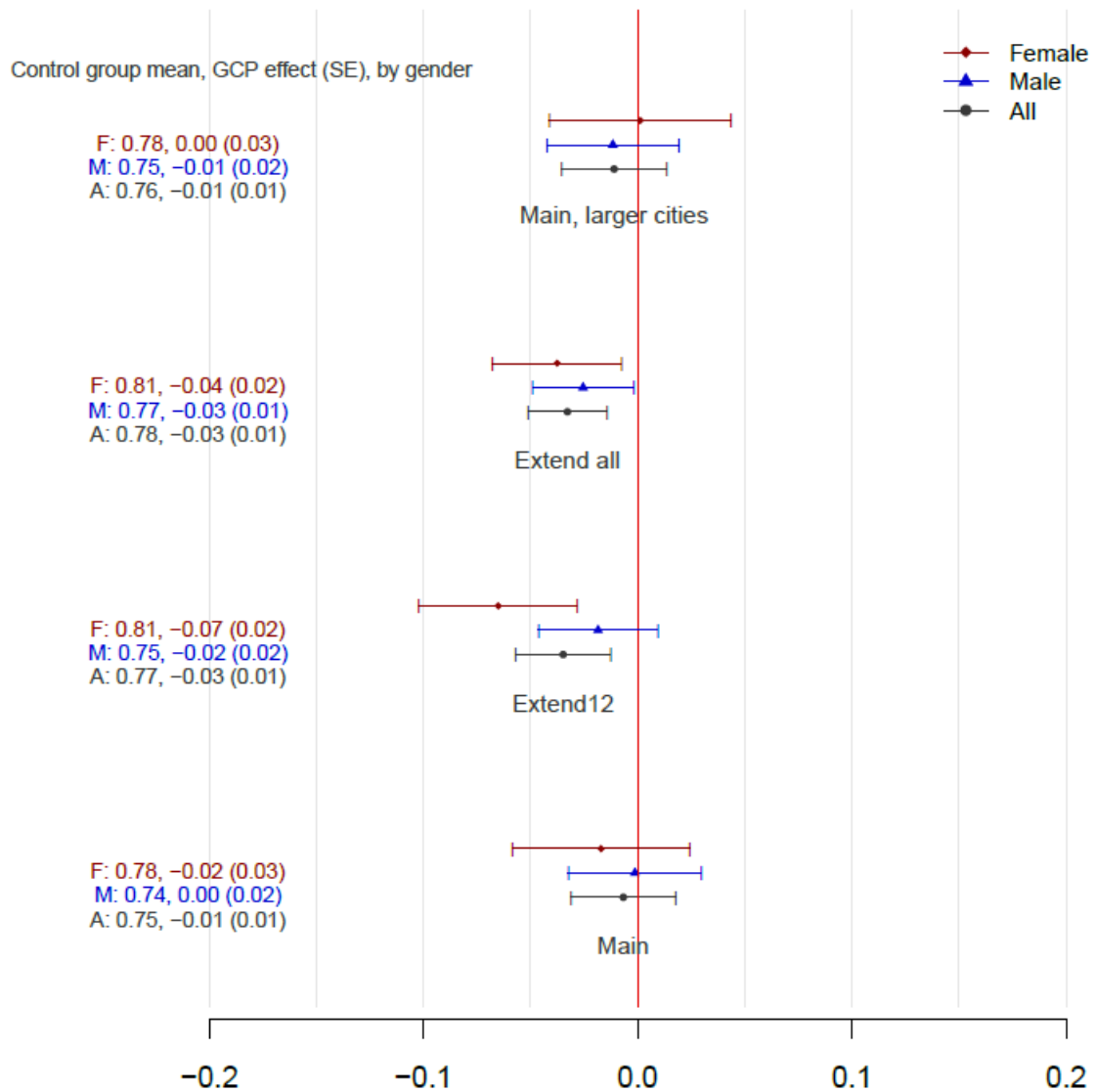
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Engineering, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A24: GCP Effects on BA in Physical Sciences, By Sample



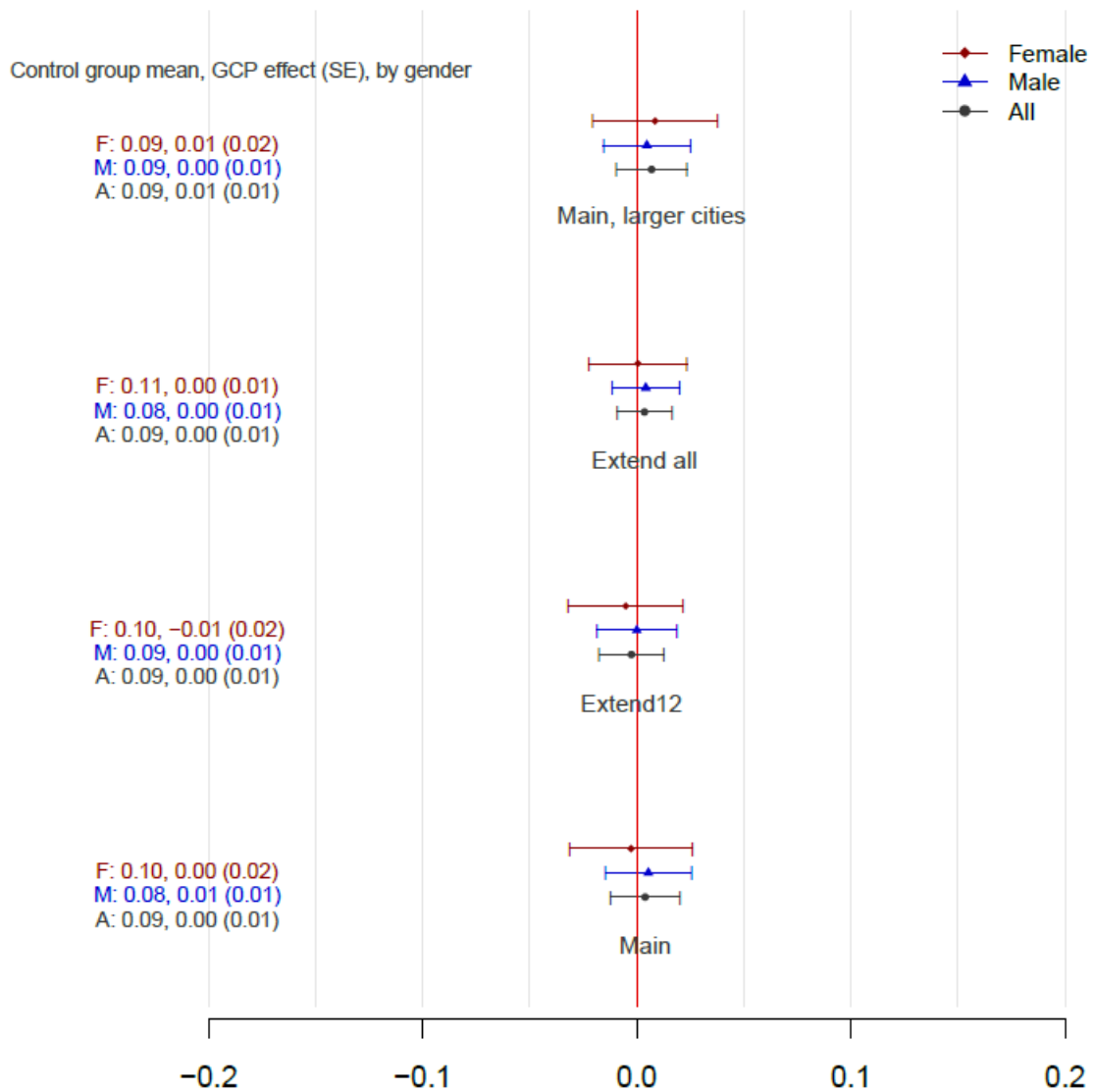
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Physical Sciences, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A25: GCP Effects on Employment in 2018, By Sample



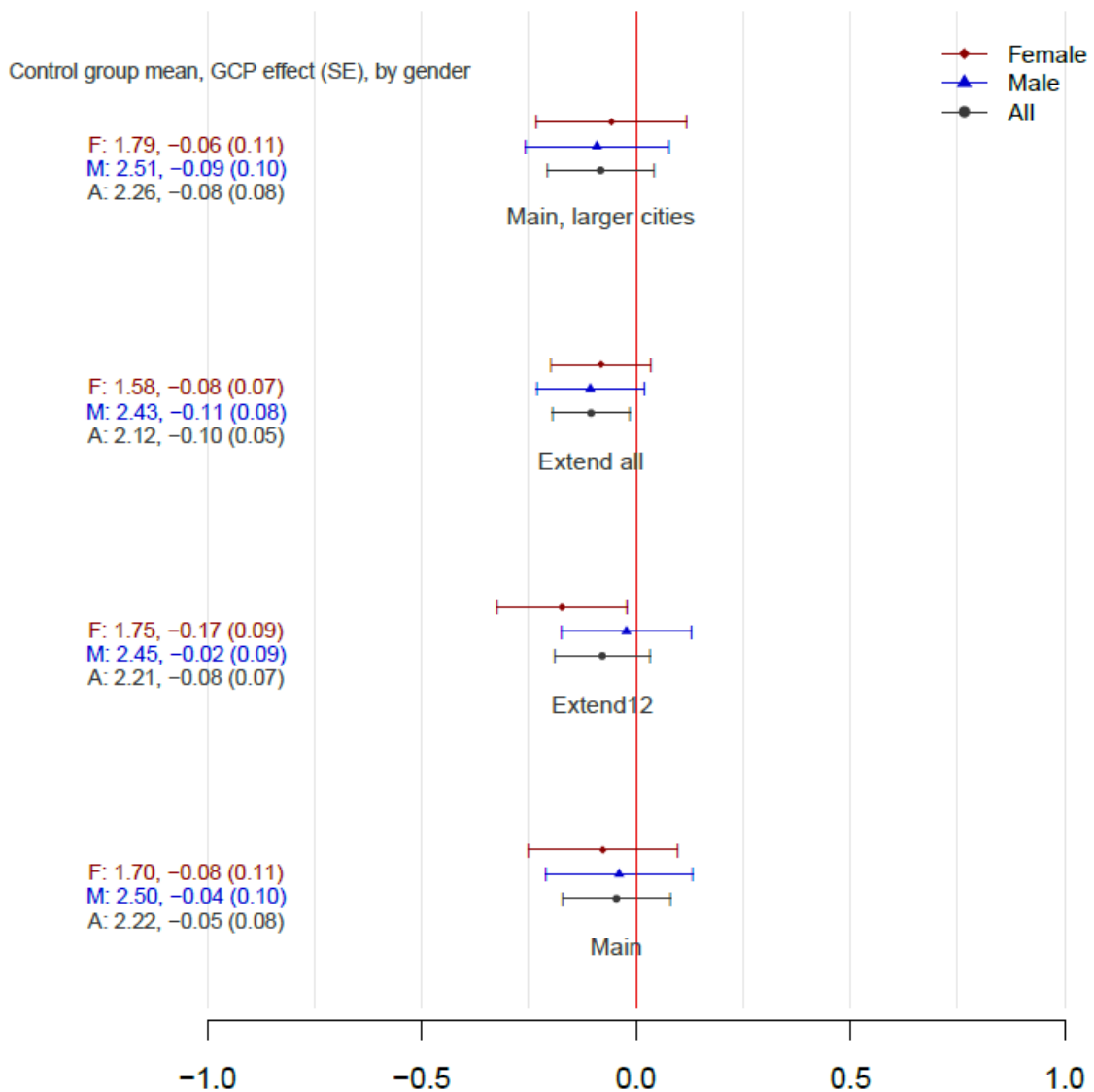
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employment in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A26: GCP Effects on Self Employment in 2018, By Sample



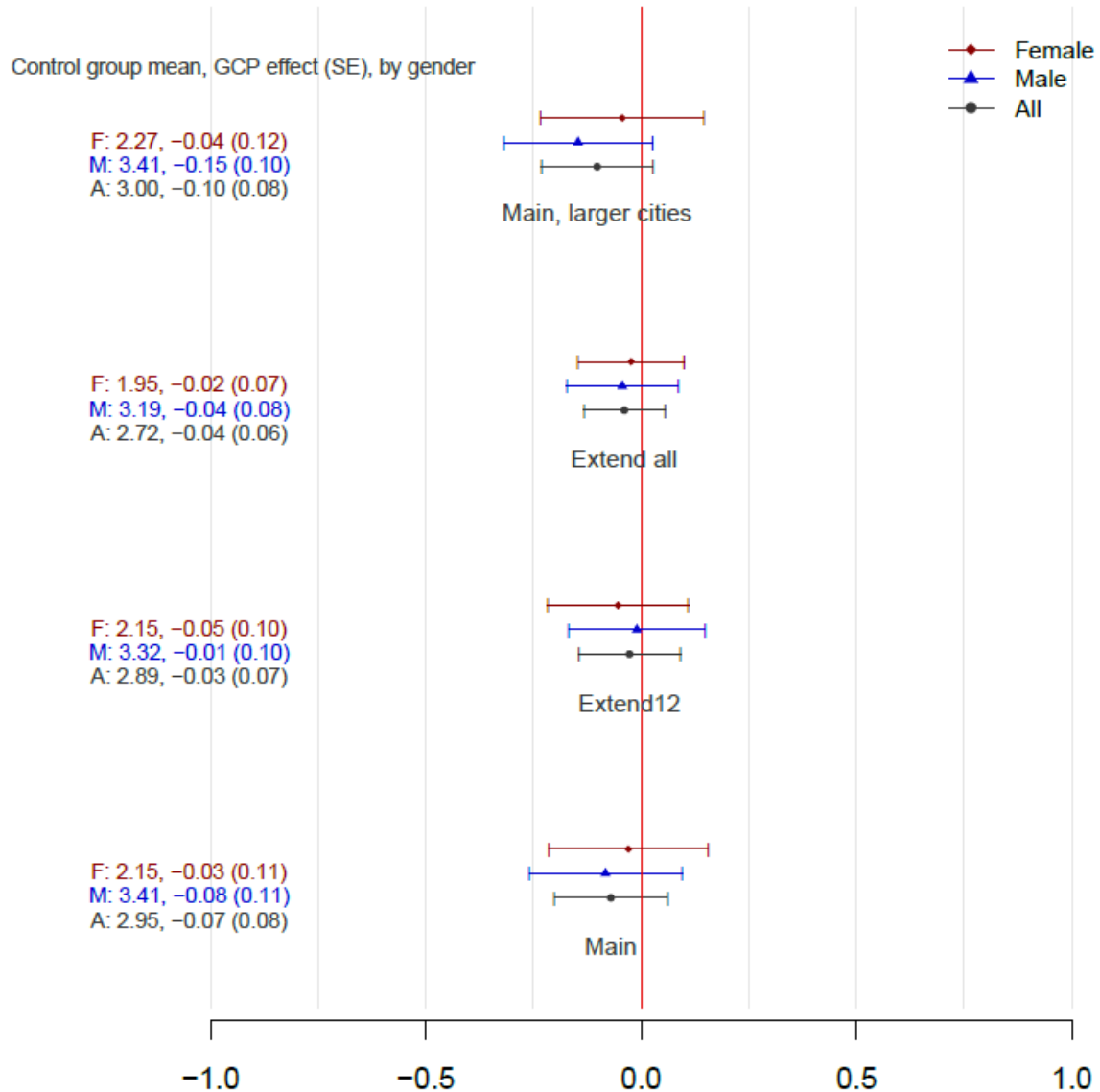
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on self-employment in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A27: GCP Effects on Salaried Worker Income in 2018, By Sample



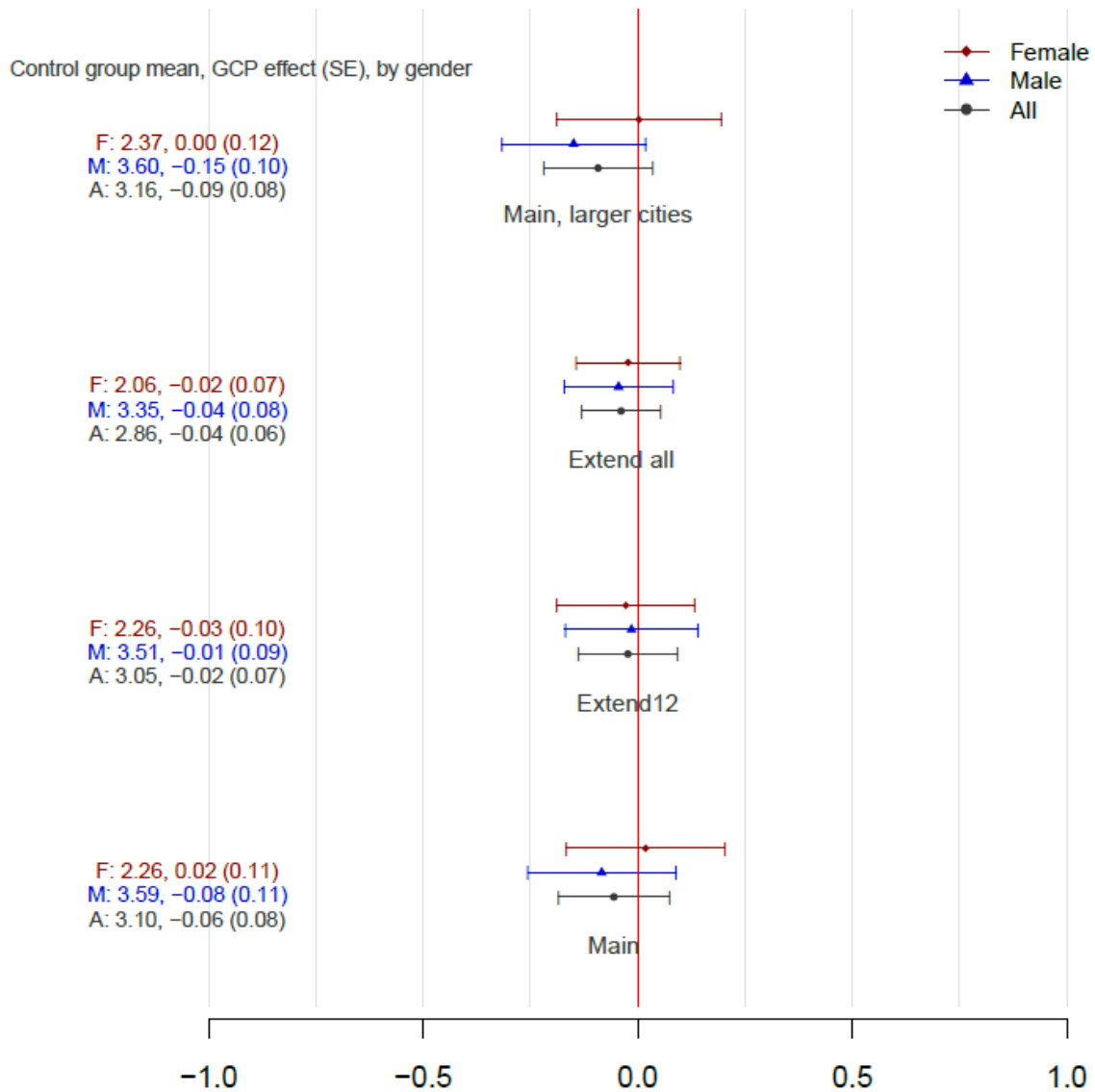
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on salaried workers' income in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A28: GCP Effects on Salaried Worker Income (employed=1) in 2018, By Sample



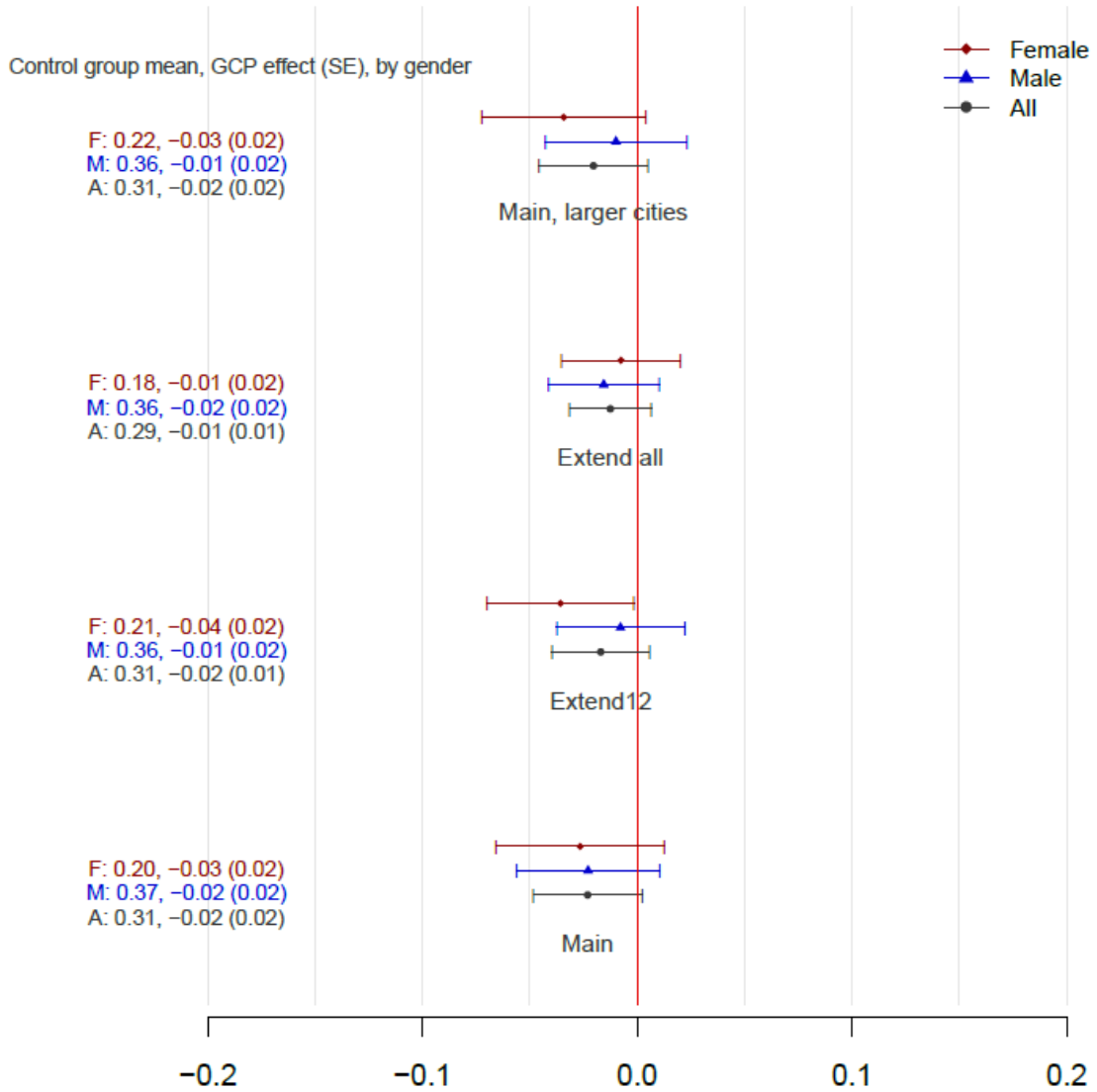
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on salaried workers' income (employed=1) in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A29: GCP Effects on Income (employed=1) in 2018, By Sample



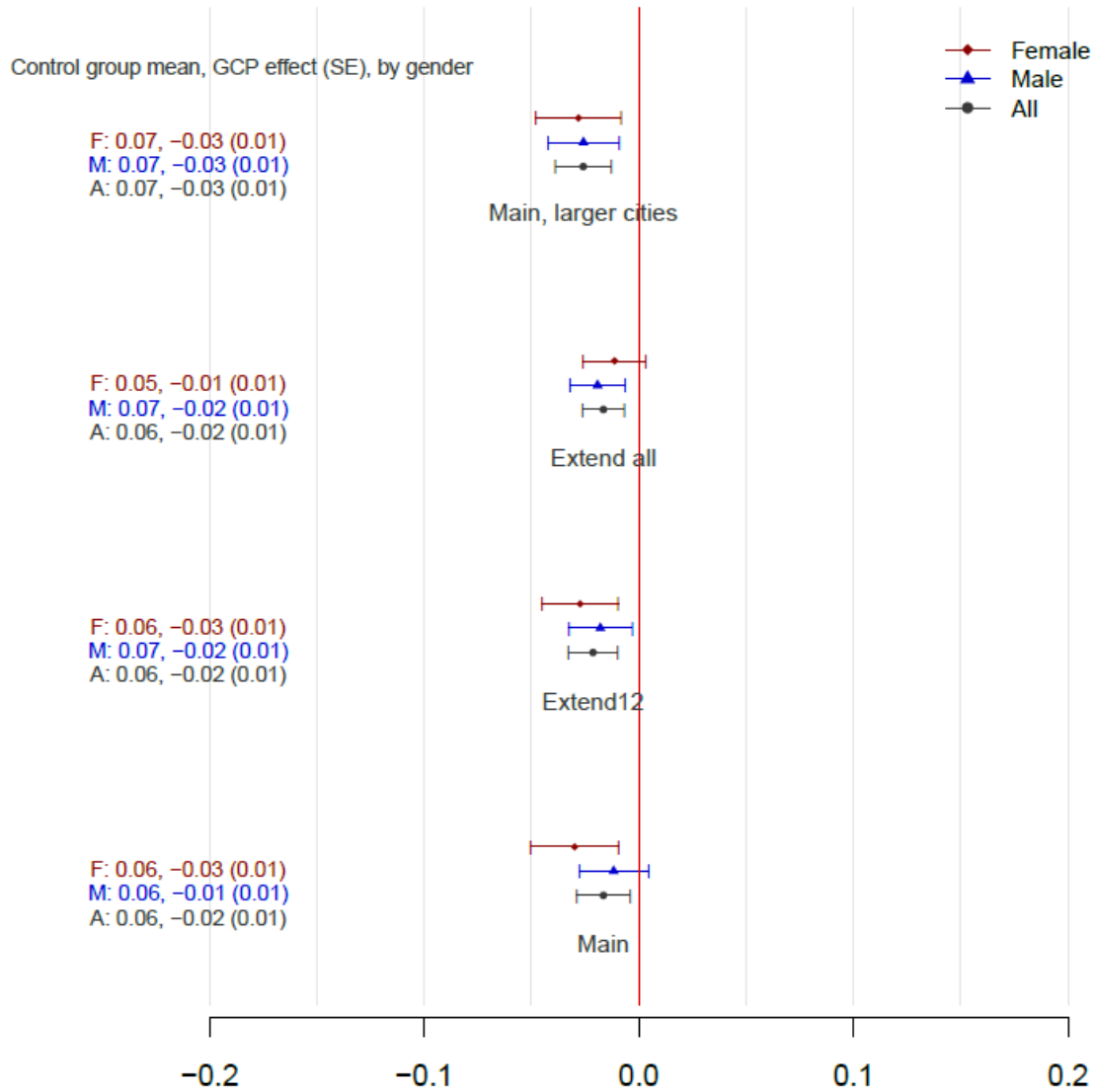
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Annual Income (employed=1) in 2018, by sample by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A30: GCP Effects on Employment in High-Tech Services Sector in 2018, By Sample



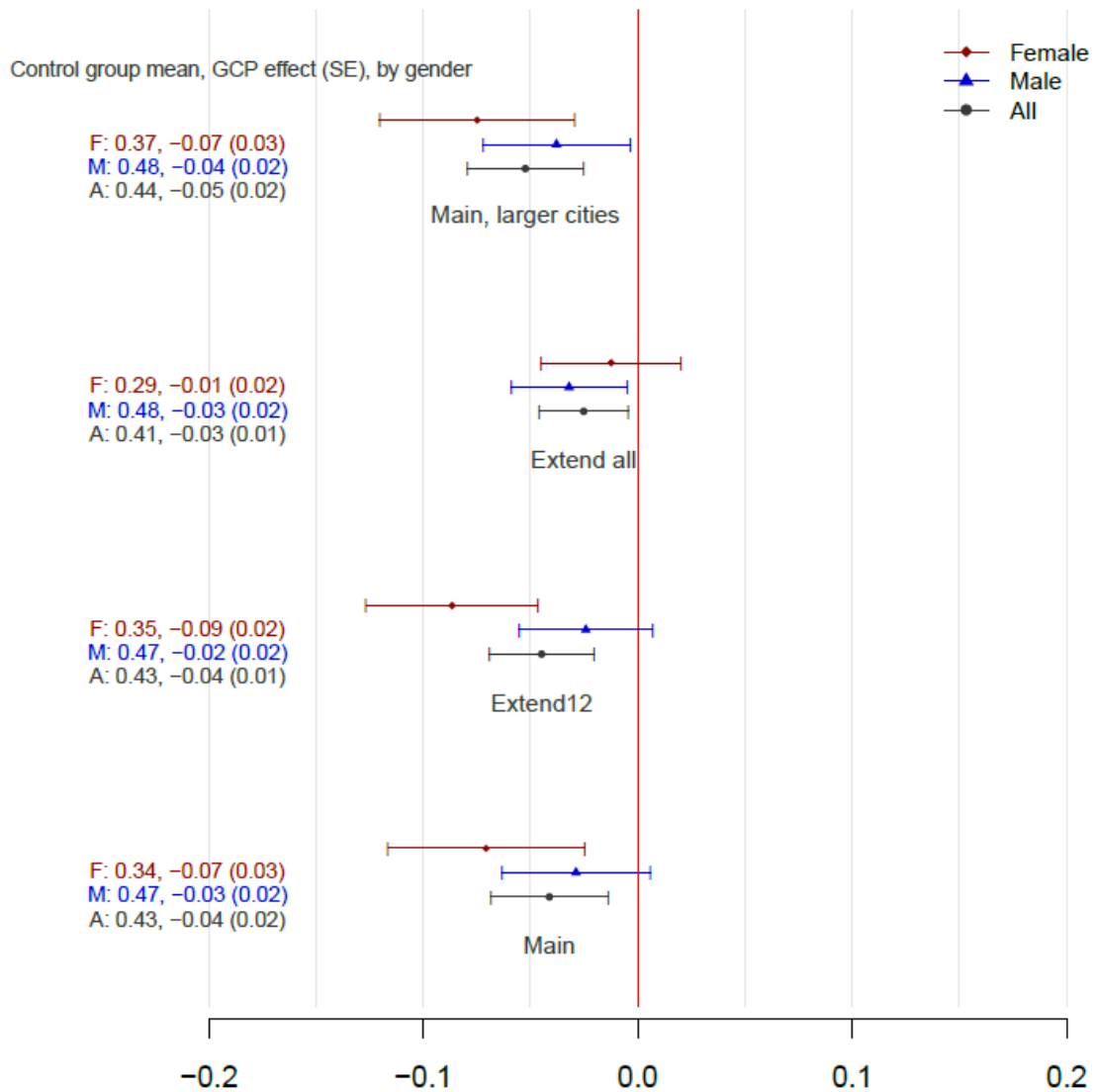
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employment in the High-Tech services sector in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A31: GCP Effects on Employment in High-Tech Manufacturing Sector in 2018, By Sample



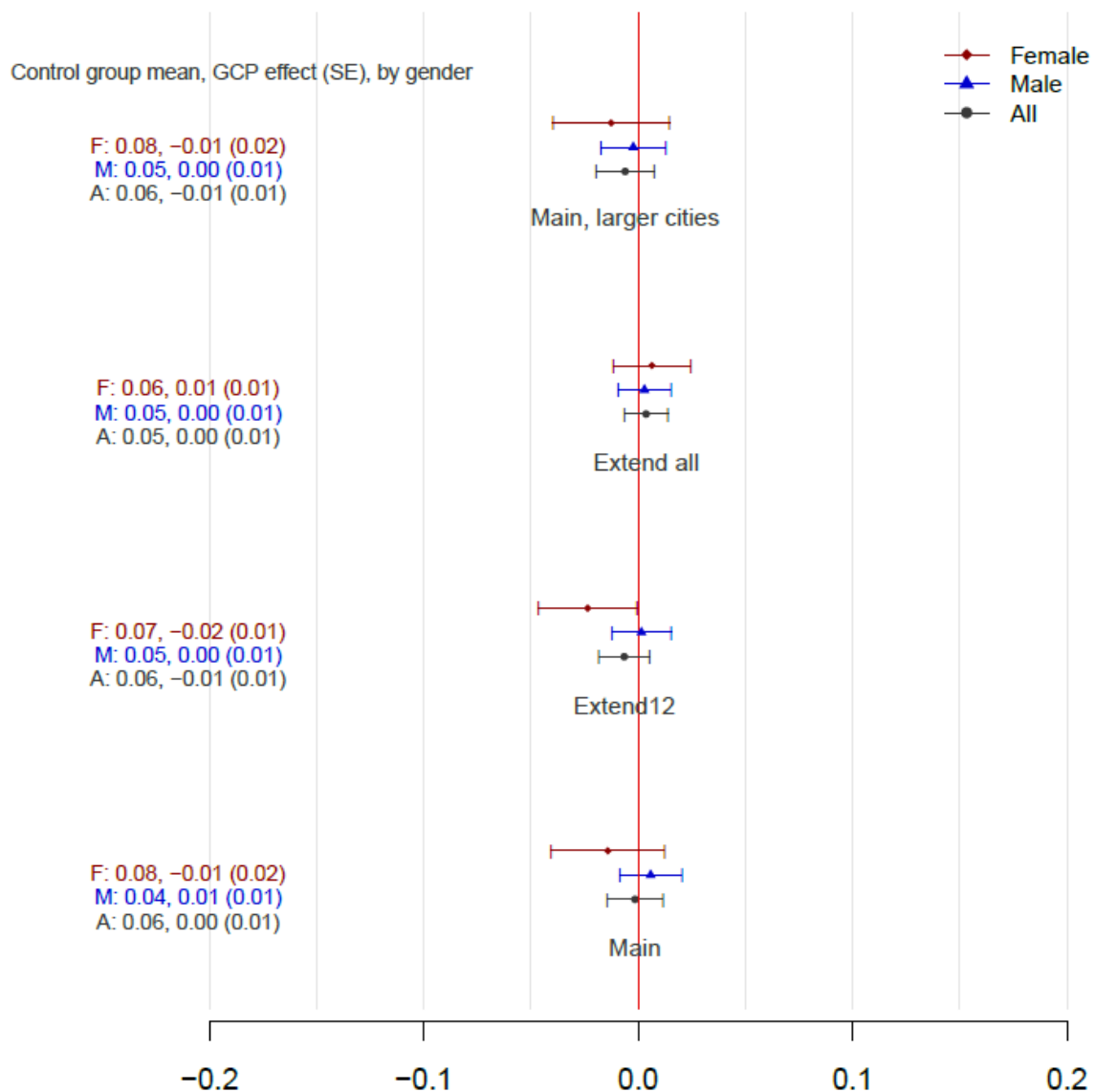
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employment in the High-Tech Manufacturing Sector in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A32: GCP Effects on Employment in the Knowledge Economy in 2018, By Sample



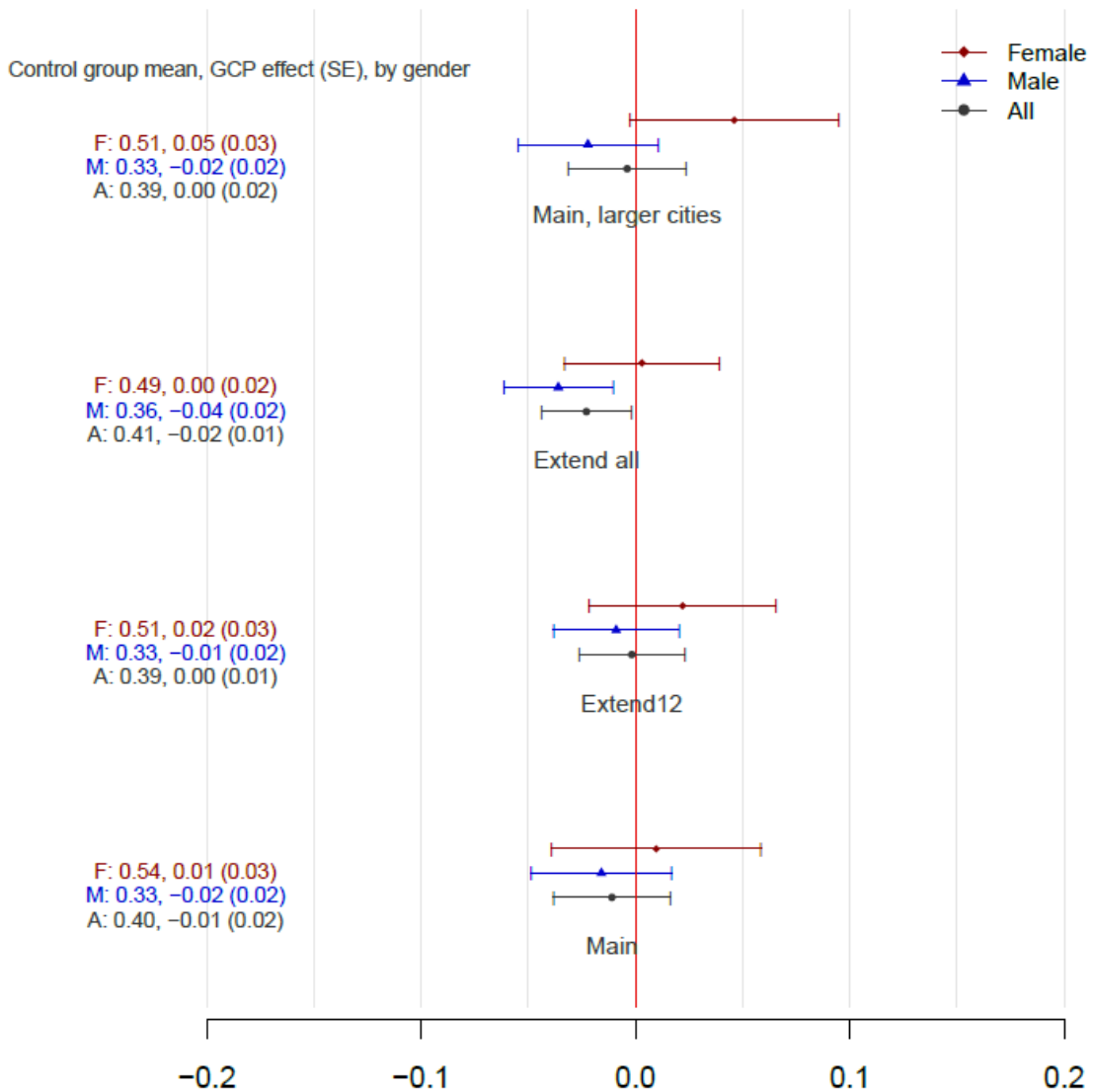
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employment in the knowledge economy in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A33: GCP Effects on Employment in the Academic Sector in 2018, By Sample



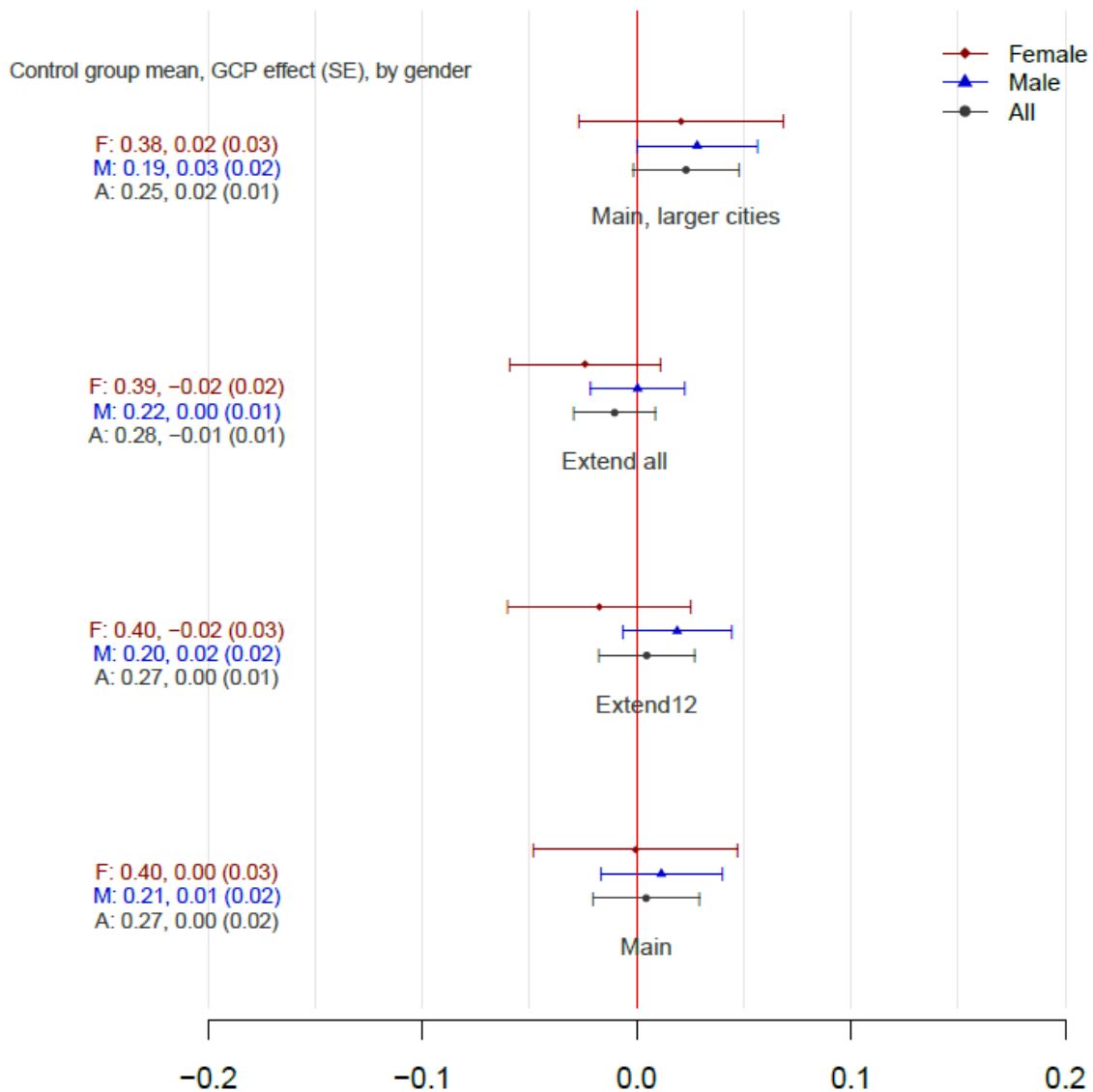
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employment in the academic Sector in 2018, by sample. *Main* denotes the main sample of 1992-2005 graduates, described in the text. *Main, larger cities* denote a sample that restricts the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test-takers. *Extend all* denotes a sample that adds to the main sample UPET test-takers at all ages.

Appendix Figure A34: GCP Effects on Marriage Before 30, By Sample



Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on marriage before 30, by sample. *Main* is the main sample of 1992-2005 graduates, described in text. *Main, larger cities* is restricting the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test takers. *Extend all* is adding to the main sample UPET test takers at all ages.

Appendix Figure A35: GCP Effects on Having Children Before 30, By Sample



Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on having children before 30, by sample by sample. *Main* is the main sample of 1992-2005 graduates, described in text. *Main, larger cities* is restricting the comparison group to include only students from larger cities (with 200 or more students in the cohort). *Extend12* is adding to the main sample UPET 12th grade test takers. *Extend all* is adding to the main sample UPET test takers at all ages.