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

26_October_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 intelligence and ability 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 14 cohorts of GCP participants who graduated high school in 1992-2005. We measure treatment effects on outcomes, ranging from high school to the labor market in their 30s and 40s. 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. The effect on the most chosen elective studies (computer science, physics, biology, and chemistry) is zero. The impact on the average composite score is negative, driven mainly by the effect on boys. However, all these estimates are relatively 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 MA and Ph.D. in Elite Universities is also positive; for the latter, it is statistically significant, with an effect size of about 50 percent increase. Examining choice of field of study shows that gifted children in GCP study more math, computer, and physical sciences but engage much less in engineering programs. The net effect on STEM degrees is, therefore, zero. However, among GCP participants, a much higher share 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. We examine marriage and family formation patterns as mediating effects and find no discerned GCP effects either.

As robustness check, we used different samples based on the age at which students took the intelligence and ability test to match a control group to the treatment group. Our results are fully robust to variations in the sample we use. In addition, as an alternative matching of a control group, we used 8th-grade national exams test scores instead of the intelligence/ability measures. The results are very similar, and the estimated effects on all university schooling outcomes are even identical.

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.

* 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 Psychometirc Entrance Test. We thank James Fensky, Emmch Duchini, and participants in seminars at Hebrew University, 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, 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) in high schools in Israel. The program tracks the most talented students into gifted children classes, starting 10th grade. They receive more resources, a unique and accelerated curriculum, access to high-quality teachers, and attend university courses. We use test scores in exams that measure intelligence and ability 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 14 cohorts of GCP participants who graduated high school in 1992-2005.

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 treated and control gifted children achieve this degree (99 percent). Still, we find a positive and relatively large impact on double majors and gaining advanced 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 have very similar earnings and employment rates in the knowledge economy sectors. We examine marriage and family formation patterns as mediating effects and find no discerned GCP effects either. However, we find that the program positively affects the "quality" of the partner, 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 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. Lastly, we estimated heterogeneous treatment effects by socio-economic status (SES). We found that the lower SES participants drive the minor adverse effects on *bagrut* test scores, while the longer-term effects are pretty similar between the groups. Our results of almost no variation in treatment effect by students' background stand in contrast to the abundant literature that show that the effect of many educational programs vary by socio-economic status of students. But our evidence are consistent with the fact that background characteristics of gifted children make little difference to their life-long outcomes.

As robustness check, we used different samples based on the age at which students took the intelligence and ability test to match a control group to the treatment group. Our results are fully robust to variations in the sample we use. In addition, as an alternative matching of a control group, we used 8th-grade national exams test scores instead of the intelligence/ability measures. The results are very similar, and the estimated effects on all university schooling outcomes are even identical.

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 the literature. 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 comprehensive set of outcomes in the medium and very 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 research and development firms and academic institutions. Another important 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.

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 direct gifted adolescents to math and computer and physical sciences program at the university is 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. In the next section, we describe 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. We then present the impact on labor market outcomes, including employment and earnings followed by evidence of the effects on carrier choices in knowledge-producing sectors, and marriage and parenting (section 5). Section 6 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 (ex 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 (defined[1] 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

[1] Self-concept tends to be more malleable at younger age and still going through the process of 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 attributes like 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).

responsibilities, including testing children 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), and 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. Students screened as gifted and who participated in GCP in earlier grades were not required to retake the test. Thus, we focus on the gifted classes for students in their 10th to 12th grades from 1992-2005. There were gifted classes in 12 schools in 10 localities in Israel, most in the major cities.

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. The number of special classes that operate in secondary schools has since then expanded to over 20.

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 seven universities (one of which confers only graduate and Ph.D. degrees) and over 50 colleges that confer academic undergraduate

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

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 available 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 includes 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 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 10^{th}-12^{th} 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 main analysis includes fourteen cohorts who completed high school (12^{th} grade) in 1992-2005. In 2018, the last year in the data we use, 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

In this subsection, we describe the outcomes in adulthood for students in the 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).

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 university field of study, we create dummy indicators for individual or grouped majors, for example, STEM. We 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 as outcomes earnings from salaried employment and earnings conditional on employment. 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 students' 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 academic, 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 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, Psychometric score, 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

[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. We do not use these data in our analysis because the sample is relatively small.

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, motivated by how GCP affects achievement for infra-marginal gifted children, we use an alternative identification strategy in this paper.[8] We chose as a control 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 control group. The first is the study program in high school 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 98 percent of gifted students), math (93 percent of boys, 81 percent of girls), physics (70 percent of boys, 31 of girls), chemistry (48 percent by girls and 43 by boys) and computer science (58 percent by boys, 31 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 only in cities with GCP, no such test scores are available for selecting a control group. We, therefore, opt for another ability measure. Applicants to universities in Israel have to take a psychometric test administered by The National Institute for Test and Evaluation (NITE).[9] The University Psychometric Entrance Test (UPET) is a tool for predicting academic success at higher education institutions in Israel.[10]

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

[8] A recently developed identification strategy allows to capture causal effects for children other than those in the immediate neighbourhood of admissions cut-offs, under a conditional independence assumption (CIA). Studying 9th grade applicants to Boston exam schools, Angrist and Rokkanen (2015) show that conditioning on baseline scores and demographic variables largely eliminates the relationship between the application score – the running variable in this contest - and studied outcomes. The authors claim that these results lay the foundation for a matching strategy that identifies causal effects for infra-marginal applicants.

[9] The National Institute for Testing and Evaluation was established in 1981 by the Association of University Heads in Israel to assist in the admissions and placement process for applicants to Israel's higher education institutions. NITE develops tools for testing and evaluation, especially admissions, placement, and accreditation tests for higher education institutions.

[10] With matriculation results and other assessment tools, the test is used to screen applicants to different faculties. The test enables the institutions to rank all applicants on a standardized assessment scale. Compared with other assessment tools, the test is less affected by each applicants' background or other subjective factors. A large body of research demonstrates the high predictive ability of the PET. Students who received high UPET scores are more successful in their academic studies than students who received low scores.

Remarkably, even though most students in Israel start college education at age 22-23[11], a high proportion of GCP participants take this test while in high school and even before midway into the program. Sixty-one percent take it during 10^{th} or 11^{th} grade, 12.5 percent take it during 12th grade, and 18.7 percent take it at age 22 or older. The rest, 7 percent, take it during their military service (age 18-21). We show that the UPET score's distribution of gifted students who took it in 10^{th}-11^{th} grade is identical to those who took the exam in 12^{th} grade. This similarity holds for the total and means scores in each UPET's three domains: verbal reasoning, quantitative reasoning, and English.

Further, the numeric part score of early takers (in 10^{th}-11^{th} grade and 12^{th} grade) is also similar to that of late takers after age 21. We think this pattern of little variation in UPET score by testing age permits viewing it as pre-determined to the GCP we study. We, therefore, will use it in selecting a control group from localities where the screening test of gifted children was not administered. We will first show results based on a sample of only 10^{th}-11^{th} UPET early takers. We will then expand the sample gradually to include 12^{th}-grade UPET takers and late takers and show that GCP treatment estimates are remarkably identical across these samples.

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. 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 in comparison to the gain experiences by non-gifted children.[13]

We provide additional evidence in support of this research strategy. First, we compare the test scores in national exams in English, Hebrew, math, and science, in the 8th grade, of GCP participants and students of the control group. We show that the two groups' density function of these pre-program test scores and the respective means are similar. These results are discussed in Section 4.2.

Another 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 kids had a higher mobility rate before 10th grade than families with gifted kids that did not participate

[11] 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.

[12] Evidence of high correlation between PET scores and IQ are 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 to 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.

in a GCP. We find no such differential mobility rate. These results are shown in the online appendix Table A1.

To compare our results to earlier studies that examined the effect of those identified at the giftedness border, we show evidence by three segments (thirds) of the UPET score distribution. The impact estimates for the lowest third can be viewed as representing the effect of the marginal group of students enrolled in GCP. We find that the GCP impact estimates for various outcomes do vary somewhat by the ability of the gifted children.

4.1 Evidence on Psychometric Scores by Testing Age

Figure 1 presents the UPET score distributions for two samples of GCP participants, those tested in 10th or 11th grade versus those in 12th grade. The first (upper left) panel 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.[14] The P values of this test for the composite score and the quantitative and verbal reasoning indicate that the pair-wise sample densities are not statistically different. The P-value for the total score is 0.619, for the quantitative reasoning 0.605, and verbal reasoning, it is 0.760. The p-value (0.0429) rejects that the two densities are statistically different only for English. This pattern holds also for the younger sample- cohorts 2006-2014 (see online Appendix Figure A1). Online Appendix Figure A2 shows the respective evidence for non-gifted children in our baseline sample with slightly larger differences.

Online appendix figure A3 show the densities for those tested in 12th grade versus those tested later. The figure shows that those tested later have little higher English and Hebrew scores. However, the numeric score density of these two groups is still very similar. Therefore, we will also check the robustness of our main results for the exclusion of the Hebrew and English UPET scores from the matching, as will be discussed later.

4.2 Constructing the Control Group

We use propensity score matching to choose the control group.[15] Then, we use the following three-step algorithm:[16]

---

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

[15] Rosenbaum and Rubin (1983) proposed an approach that circumvents the curse of dimensionality when using selection on observables for identification of causal effects. They provide a 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.

1. Estimating the propensity score using a Logit specification

(1) $\boldsymbol{P_1(X_i) \equiv \mathbf{Pr}\,(T_i = 1|\, X_i)}$

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 each of the three domains of the UPET, 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.[17]

3. Estimating the following controlled regression

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

This combination of propensity score matching and regression allows for enhanced robustness to misspecification. As long as the parametric model for either the propensity score or the regression functions is specified correctly, the resulting estimator for the average treatment effect is consistent. A notion discussed and termed as 'double robustness' in Robins and Ritov (1997) and Imbens (2004).[18] The standard errors of the program effects estimates were clustered at the school level.[19]

Figure 2 presents the propensity score distribution before and after the matching. The sample is based on UPET takers during the 10th-11th grades. We match 1769 students from this sample of 1949 gifted GCP participants. The unmatched are mostly from the very upper end of the propensity score distribution. The propensity score density graphs of the GCP participants and their matched counterparts are perfectly aligned and not distinguishable.[20]

In Figure 3, we present the density distributions of the UPET scores of the same two samples (total score, quantitative reasoning, verbal reasoning, and English). The KS test shows that the UPET total score distributions are not distinguishable statistically; the K.S. p-value is 0.857. An even closer similarity is seen in the two distributions in the UPET quantitative reasoning scores (p-value=0.900). Similar results are evident for the pair-wise distributions of the verbal reasoning and English parts of UPET. Figure A4 shows the density distributions of the UPET scores of the two samples before the matching, where there are substantial differences between GCP participants and the comparison pull

[16] 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.

[17] We further validate that our main results are not sensitive to the specification we choose by running alternative matching specifications. The results are shown in the online appendix Table A2and they are very similar to our main results.

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

[19] We provide also alternative calculation for the standard error in the online appendix Tables A3 and A4. The table shows that the clustered standard errors are almost identical to the correction specified by Abadie and Imbens (2008), and to bootstrapped standard errors.

[20] We also used a model where we imposed exact matches by cohorts. This propensity score regression specification yields a much smaller sample of matched gifted children, 1431. We report below results based on this sample as well. The estimates are not different from the ones we obtain when not imposing the cohort exact match restriction.

(before matching), as expected since the vast majority of students in the comparison are not gifted students.

Table 1, columns 1-2, presents detailed summary descriptive statistics for the girls' sample for the variables that we use in the propensity score estimation and other background variables by GCP participants and their matched counterparts which 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, we present the same evidence for the sample of boys. Panel A presents evidence for student's demographic and socio-economic characteristics. In panel B, we present evidence on the UPET total and three domains scores.

First, it is interesting to note that gifted children, GCP participants, and others come from a higher socio-economic background than regular students in Israel. For example, mean mother and father years of schooling are around 14.5 years for these two groups, higher than among non-gifted children (where years of education of the father is 14.0 and of the mother is 13.9). The difference between the gifted and the non-gifted samples in father's income is 23 percent, and in mother income, it is 22 percent.

Twenty-four 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 control group except for one. 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.

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 of parental income is particularly important. For example, the Mother's average yearly earnings are 102,000 (107,000) NIS in the girls (boys) treated group and 103,000 (100,000) in the comparison group. These differences of less than 10,000 NIS are not statistically different from zero. There is a small and statistically significant difference of about 31,000 NIS for fathers' annual earnings in the boys' sample when the average yearly earnings are 169,000 NIS in the treated group and 200,000 in the comparison group. However, there is a 5,000 insignificantly difference for fathers' annual earnings in the girls' sample. The four UPET mean scores are perfectly balanced, as expected, given the evidence shown in Figure 3. Table 2 presents descriptive statistics for the *bagrut* Study Program, with only one out of 22 outcomes that shows a statistically significant (95%) difference between treatment and control. The table shows that Gifted children have much higher participation rates in science advanced study programs. For example, in Math (93% of the boys and 81% of the girls), Physics (70% and 34%), Computer Sciences (58% and 31%). Gifted children also have many *bagrut* study credits (30.5 and 29.3, while the requirement is for 21).

Evidence on balancing in pre-treatment 8th-grade national exams scores

Figure 4 presents the test score distribution in the national exams ("Metzav") in English, Hebrew, math, and science. These tests started in 2002, so the test scores are available only for younger cohorts - high school graduates of 2006-2010. We present in the figure the distribution before and after propensity score matching. Most of the differences between the test score distributions before the matching are eliminated after matching. The Kolmogorov-Smirnov test p-value allows rejecting the null that the two distributions are different in three subjects (Hebrew, math, and science). The p-value for English is statistically insignificant. The mean differences after the matching are also small (see table A5 in the Online Appendix).[21]

This result supports the validity of our research design. The main caveat for our application could have been that the psychometric score is positively affected by participation in GCP. However, this would imply a negative difference between our treatment and control group in their pre-treatment abilities. Instead, we find a very small positive difference in a pre-treatment proxy for abilities, alleviating these concerns.

4.3 Simple Treatment-Control Differences in Outcomes

Table 3 shows the treatment-control contrast in mean test scores in all compulsory subjects and four most selected non-compulsory subjects. The unconditional differences are mostly very small, none larger than 3 percent relative to the control group mean. Interestingly, all the statistically significant negative differences (lower in the treatment group) are among boys. The statistically negative unconditional differences among boys are in math, bible studies, and Hebrew. However, GCP participants show positive and significant advantages in history and literature. These gaps accumulate and as a result the mean composite scores of boys in the treatment group is lower. As shown below, all these unconditional differences will survive the regression analysis when controls are added. Among girls, the only marginally statistically significant simple mean difference is in history.

Table 4 present the treatment-control simple differences in university outcomes. As with the *bagrut* outcomes shown in Table 3, whatever is significant in this table will remain in the regression analysis shown below. This pattern is a follow-up consequence of the perfect balancing when matching on the propensity score.

In contrast to the 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 identical for men and women. There are no significant gender differences in university degree attainment. An almost equal proportion of treated and control gifted individuals get an MA degree (about 54-57 percent) without considerable gender differences. More treated girls and treated boys get a Ph.D. degree (the difference is similar, 3-

[21] Table A6 in the online appendix reports the correlations between the UPET scores and the 8th grade national exam scores.

4 percent points).[22] We also find a largely positive and statistically significant difference in completing BA degrees with a double major.

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, and physical 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 statistically insignificant both for men and women.

Table 5 presents treatment-control groups means differences of the labor market and personal outcomes. Over 70 percent of our sample are employed, and there is a negative difference in employment among the girls.[23] More women than men are self-employed (10-11 percent versus 8-9 percent). Over a quarter of the women and about half of the men work in high-tech and knowledge-producing firms and institutions. Many men (35-36 percent) are in high-tech services (telecommunications, computer and related services, and research and development). More women work in universities (6-9% relative to 4-5%). The negative differences in employment in the knowledge economy and its sub-sectors among girls is consistent with the negative effect on employment for the girls.

In panel C, we present a means comparison of personal outcomes. Sixty-five percent of women and 57 percent of men are married by 2018. The divorce rate is 2-3 percent among women and 2 percent among men. Women marry earlier, on average at the age of 28.5, while men marry at the age of 30.5. Women and men have their first child about a year and a half after being married. The treatment-control groups' mean differences in these outcomes are minor, and none are statistically significant. Lastly, we find that GCP participants are more likely to marry other GCP participants. These differences come primarily from within program matches and induce a statistically significant increase in partner's psychometric scores for men (and a more minor and insignificant increase for women). However, we do not find a statistically significant effect on the partner's income.

5. Results: Short- and Long-Term Effect of Gifted Children Programs

We present results by the following outcome groups: (a) high school, (b) university schooling, (c) fields of university studies, (d) employment and earnings, (e) employment in sectors of the knowledge economy (f) marriage, divorce, and parenthood. We will also present results for various outcomes while gradually expanding the sample to include late takers of UPET.

[22] The MA and Ph.D. attainment rate in the non-gifted population is much lower, 31% and 4% percent.

[23] We should be cautious in interpreting the evidence about employment (in Israel) because we do not observe in the sample those who left the country and perhaps work abroad. While we do not observe employment outside of Israel, we provide in section 5 a discussion of this issue.

5.1 Effect on High School Outcomes

Figure 5 presents the estimates and confidence intervals (10 percent level of significance) for the effect on matriculation certification and *bagrut* exams scores obtained from estimating equation (2). We show 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, physics, biology, and chemistry. We do not present the effect on obtaining a matriculation diploma because this outcome for both treatment and control groups is almost 100 percent. We also show the control group means, the point estimate, and its standard error in the figure. These are also presented for females and males separately.

Overall, the point estimates on test scores are negative in all compulsory subjects except history and literature, where they are positive, and English (zero). The estimated effects on the bible and civic studies, Hebrew and math, are negative and statistically different from zero. These effects are driven mainly by the adverse impact on male students. However, the effect on the mean composite score is statistically insignificant, and all the effects are very small in magnitude. The most considerable negative effect on the male sample is in bible studies, -0.07. However, relative to the control group mean (5.30), the effect size is only a 1.3 percentage point decrease. The largest positive estimated effect is for literature test scores, 0.12 against a mean of 4.84.

We note that these test scores are grouped into a seven-point scale (0-6). The test score data on a scale of 1-100 is available for later years (from 2006 and beyond). We estimated the effect on this measure of test scores using the sample of GCP participants in 2006-2014. We kept only GCP participants in the same schools we used in the earlier sample (1992-2005). We matched a control group anew from other cities that still did not have a GCP in any of the years in this later period. Since we also have the test scores for this period in the 0-6 scale, we used them and compared the two sets of results, presented in the online appendix Figure A5a and A5b. We find the same pattern of small negative effects on *bagrut* test scores for this sample and period. The effect on the mean composite score is negative and a little bit larger than our main specification (-0.12). However, the relative decrease is still very small (2.1%). The estimates based on the 1-6 scale test score for these younger cohorts are entirely aligned with the evidence obtained using the 1-100 score scale.

Estimates from extended samples

Next, we estimated the effect on these high school outcomes while varying the sample's definition based on the age at which students took the UPET. These are the alternative samples that we use: (1) the baseline sample (2) the baseline sample, where we omit the Hebrew and English UPET scores from the matching (include only the numeric score) (3) adding to the baseline sample UPET 12th grade test takers (4) adding to the baseline sample UPET test-takers at all ages, (5) sample 4, where we omit the Hebrew and English UPET scores from the matching.

The estimation with five 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 high school outcomes are similar regardless of which sample we use. This result will repeatedly appear for all other medium and long-term outcomes that we will examine in this paper. 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 6 for the sample of girls and Table 7 for boys (in the online Appendix A6-A27, we show these estimates graphically, too).

Column 1 in Tables 6 and 7 show that all estimates for the effect of GCP on the *bagrut* average test scores of boys and girls are tiny (between -0.04 and 0.11), with minor variation in the point estimates. Column (2) in these tables shows that all the estimates for the effect on the *bagrut* score in math are negative (both for boys and girls). Note also that the estimates of the alternative samples are more precisely estimated than in the benchmark sample because of the larger number of observations.

Discussion of mechanisms: gifted children program learning environment

The pattern of mostly a small adverse effect of GCP on test scores in *bagrut* exams is interesting. It is especially intriguing given the abundance of educational inputs that GCP participants enjoy relative to the control group that we use. In the online appendix Table A7, 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. 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?

Two mechanisms can be offered 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 put 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 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; potentially more talented people in the new peer group. When two students of the same ability or achievement level are put 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.).[24] 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 student's 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 prior ranking in primary school.[25]

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 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

[24] See Marsh et al. (2008) for a review.

[25] 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).

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 supported 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 a gender peer effect. There is growing evidence that a higher proportion of female students in the classroom improve the learning environment and raise achievements of boys and girls (Lavy and Schlosser 2012). The proportion of female students in a gifted children classroom is lower on average than in regular classes. Therefore, a gender peer effect in this case will lower learning outcomes of both genders.[26]

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 the following university outcomes: getting a BA, MA, and Ph.D. degrees. As expected, GCP does not affect BA attainment because the control group mean, 98 percent, is already near the maximum possible. However, the program has a significant positive effect on getting a double major. The control group average rate is 50 percent, and the estimated impact is 0.12, implying a statistically significant 24 percent increase. The effect on men is larger than on females though the two estimates are not different statistically.

We noted earlier that 66 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. More than half of all double majors have two STEM subjects (26 out of 50 percent). GCP enhances this ratio. Its overall effect is to increase double major with two STEM subjects by 5 percentage points relative to a baseline of 26 percent. Both genders drive this effect and when estimated separately by gender, both estimates are statistically different from zero.

[26] The literature on tracking students by ability is of limited relevancy to the context of this study. This literature refers mostly to grouping students based on median ability, the upper half having much lower mean than that of 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).

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 leads to a small decrease in MA degree attainment among girls and a small increase among boys, but these effects are not precisely estimated. The most significant effect size of GCP on academic achievement is on Ph.D. degrees. Again, the average effect is positive and statistically significant, and so is the impact on men and women when estimated separately. The average effect reflects a 25 percent increase (3 percent relative to a baseline of 12 percent). The effect is driven entirely by a dramatic increase of almost 50 percent in the probability of attaining a Ph.D. degree in an Elite University (3.3 percent points relative to 7 percent).[27]

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 sciences, with a substantial effect on men and women. Second, it causes a sharp decline in engineering degrees, both for men and women.

The increase in math/computer science is 19 percent for men (6 percent point increase relative to control group mean of 32 percent) and 50 percent among women (6 percent point increase relative to control group mean of 12 percent). The absolute decline in engineering degrees is of similar (absolute) magnitude. It is a 21 percent decline for men (7 percent point decline relative to the control group mean of 34 percent) and a 25 percent decrease among women (5 percent point relative to 22 percent). Another dramatic effect on studying science is the increase among women majoring in physics and related subjects. The counterfactual is 10 percent, and the GCP effect is a significant 4 percent points increase, implying a massive 40 percent increase. We see a similarly significant rise in studying medicine among women, a four percentage point increase relative to 9 percent otherwise.

[27] The figure also shows that most degrees that gifted students get are from elite universities. Although we do not find effect on M.A. degrees, the effect on M.A. degrees from an *elite university* is positive, and statistically significant for boys. However, the effect on B.A. in elite university is positive for girls and negative for boys. This pattern could be driven by field of study preferences. It could also reflect Israeli Defence Forces programs joint with universities that encourage gifted children to enrol in university schooling during their military service. During the 1990s/2000s years this collaboration was mostly with non-elite universities.

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.[28] 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.[29]

The results almost did not change when we extended the analysis to the other three samples outlined above. We demonstrate this pattern in Tables 6 and 7, where we present the point estimates for seven higher education outcomes in columns 5-11 of panel A (and in figures A10-A15 in the Online Appendix). All estimated effects on double major and Ph.D. are positive and nine out of ten are statistically significant. Similarly, for math and computer science as a field of university studies, the positive and significant effect among men is replicated in all the other samples. In the girls' sample, the point estimates are smaller, most positive, and statistically insignificant. The positive and significant effect on physical sciences is replicated in all the other samples (girls and boys). Additionally, in all the samples, we estimate a similar negative effect on enrollment of girls and boys in engineering study programs (all estimates are statistically significant).

Figure 8 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.

[28] 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.

[29] 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.

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 insignificant for women due to the smaller sample. 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 2.7pp relative to a baseline of 5% in the comparison group. This effect is statistically significant for both the boys' and the girls' samples.

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).

5.3 Matching a control group based on middle school test scores

We test the sensitivity of our results to the intelligence measure we use in our matching model. As discussed in Section 4, we observe the national standardized 8th grade test scores for the cohorts of 2006-2010. For the sample of students who participated in these tests and in the psychometric test (at any age), we could run three different matching specifications: 1) include all variables that are included in our main specification (denoted by version I); 2) omit the UPET scores and include 8th grade test scores (II); 3) include both UPET and 8th grade test scores (III). Appendix figures B28-A36 (and figure 4) show the distribution of each test score distribution before and after each version of the matching, and the distribution of the propensity scores before and after.

We are able to follow cohorts of 2006-2010 for up to 12 years since high school graduation. Ninety six percent of them received a BA degree within this time period and so effectively we can examine GCP effect on BA outcomes and choice of field of study. However, individuals in this sample are too young and therefore we cannot examine the GCP effect on advanced academic degrees (PhD) and labor market outcomes. Generally, the results on Bagrut test scores for this sample are stable across these three specifications. Appendix figures A37-A44 show the estimates of the GCP Effects on the mean composite, and on the scores of each compulsory subject. We also provide the estimates for the full sample in table 8. Again, we find mostly small adverse effects on Bagrut outcomes, with little variation in the point estimates between the three different matching models. In most subjects, the point estimates are very similar across the matching specifications, and in few there are small and insignificant differences. The estimated effects on university schooling are also presented in Table A8 (and in figures A45-A49). We find that the results for these cohorts based on matching on 8th grade test scores are identical to the results when matching on the university admission psychometric test scores. For example, the two estimated effects on double major are 0.20 and 0.18, on double major in STEM 0.21 and 0.16, and on engineering -0.07 and -0.10. These results indicate that the two alternative methods of matching a control group trace exactly the same treatment effect of GCP.

5.4 Impact on Labor Market Outcomes

Eighty percent of gifted women and 75 percent of gifted men are employed in 2018. These proportions are almost unchanged relative to the rates in 2016. Figure 9 shows the estimated effect of GCP on these rates is negative but very small and not statistically significant (larger and still statistically insignificant for women).[30] 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 very small positive, and statistically insignificant.

The mean income of GCP participants and gifted children in the control group is 50 percent higher than those of non-gifted students from the cities with no GCP program. However, the wage income differences between the two gifted children groups are minor and not statistically different from zero (As shown in Figure 10, there is a difference of -0.04 relative to a control group mean of 2.38 hundred thousand Israeli Shekels (IS)). The male and female mean wage earnings differences were negative but small and insignificant (-0.04 and -0.02, respectively). The results for average salaried earnings for the entire sample, and also for males and females sub-samples, are very similar. Restricting the sample to those employed only leaves the results unchanged, as expected, given that GCP did not significantly affect employment. Figure A50 shows the estimated effect on the average annual income from 2016-2018, and the results are identical. Figures A51 and A52 show the results on the natural log and the rank of the income, providing very similar results.

Extending the analysis of labor market outcomes to the other samples leaves the main conclusions unchanged. In columns 3-5, panel B of Tables 6 and 7, we show the estimated effect of GCP on earnings in 2018. These estimates are probably more precise because they are based on much larger sample sizes. There are some small statistically significant negative effects on earnings. However, none of the estimates obtained from the other sample are significantly different from the estimate obtained based on the baseline sample,

To further validate this point, we estimated the effects on the labor market outcomes with two 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. The first sample allows us to validate that the results here are not due to the different mix of cohorts in each group. The second sample focuses on individuals in the ages 36 and above (already on a much more stable stage of their

[30] The negative (though insignificant) effect on employment for girls could be due to relocation and employment outside of Israel. While we could not test this channel directly, we can estimate the effect of GCP on an indicator variable for a household with two unemployed (for the married individuals). We find a positive effect of 0.04 for girls, relative to a baseline rate of 0.08 in the comparison group (0 effect relative to 0.13 in the boys' sample). This positive effect suggests that GCP positively affects the probability that female participants and their spouses will not be employed in Israel. We interpret this result as supporting the channel of relocation.

career). The results are shown in online appendix Tables A8 and A9, and they as well are very similar to our main results. We, therefore, conclude that GCP has no significant positive effects on earnings. If anything, there could be small adverse effects.

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

Thirty-one percent of gifted children are employed in 2018 in high-tech services industries. This rate is 36 percent among men, and among women, it is 21 (Figure 9). These include telecommunications, computer and related services, and research and development. The effect on these rates is a very small negative and not statistically different from zero.[31]

Five percent of the women and 7 percent of the 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 men is a two-percentage-point decline relative to 7 percentage points in the control group. The effect on women is a two-percentage-point decline relative to 5 percentage points in the control group.

When focusing on academic institutions where 4 percent of men and 9 percent of women are employed, we still find a negative effect of GCP on this outcome among women and zero effect on men (Figure 9). Estimating the effect on employment in the knowledge economy (each of these sectors) we find a negative effect, statistically significant for women only. Using each of the other samples for estimating the impact of GCP on these employment outcomes yields almost similar results (see Tables 6, 7, A4, A5). The conclusion is that GCP does not enhance talented individuals' employment in the economy's knowledge-producing sectors. The negative effects estimated for girls in some samples could be due to the negative effect on employment (probably due to relocation), and the positive effects on studying medicine.

5.6 Effect on Family Formation

Figure 11 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 60 percent of men are 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 psychometric score of the partner (see Table 5). The effect on marriages with the same GCP participants is particularly interesting in light of the recent work by Mogstad et al. (2021).

[31] Table A10 in the online appendix provides details on the relationship between studying a STEM degree and working in the knowledge economy.

5.7 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. For sample size considerations, we divided the sample into thirds 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. We pooled together in the baseline sample (including students from cohorts 1992-2005 who took UPET during 10^{th}-11^{th} 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. We also present results by gender while using the sample of all GCP participants regardless of grade/age of taking the UPET.

We present these results in Table 9. The omitted group is the lowest giftedness ability. 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 only for the two lower Giftedness thirds (statistically significant only for the lowest third), while the estimated effect on the highest third is zero.[32] On the other hand, the negative effect on Hebrew *bagrut* test scores is large and significant only for the upper two-thirds.

Another substantial heterogeneity in treatment effect is the positive estimated effect of GCP on the attainment of PhD degrees that is evident only for the upper two-thirds of the UPET score distribution, with very large effect on the highest third (5.5pp increase relative to baseline of 18%, implying a relative increase of about 30%).

Analyzing the heterogeneity of the effects on choice on field of study, we find that the positive estimated effect of GCP on university degrees in physical sciences is evident for the lower two-thirds with only a small and statistically insignificant effect on the highest third. In contrast, the positive estimated effect of GCP on university degrees in math and computer science is evident for the higher two-thirds with only a small and statistically insignificant effect on the lowest third. Finally, we note three outcomes for which the estimated effect is very similar for all thirds of giftedness rank, attaining a MA degree (zero effect on all thirds), university degrees in engineering (negative impact), and double major (positive).

The heterogeneity in the effects on labor market outcomes is like the heterogeneity pattern in *bagrut* test scores. The negative and significant treatment effect on employment in the knowledge

[32] 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.

economy is evident only for the lower two-thirds. Interestingly, we find a positive effect on self-employment for the higher third.

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.

5.8 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 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.[33] We also stratified the sample by father's income. The following table presents the means of these two SES variables for the treatment and control group below and above the median. The mean of father's years of schooling among GCP participants above the median is 17.3, and below the median, it is 11.55, a huge difference of 5.2 years. The respective means in the control group are very similar. The comparison based on mather's year of schooling reveals a similar pattern. The statistics on parental income also shown in the table reveal equally significant differences between students above and below the median. 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		Income 2019 (100k NIS)		Psychometric Score		
		Father	Mother	Father	Mother	Numeric	Hebrew	English
Gifted in GCP (main sample)	**Father education => 15**	17.29	16.23	2.03	1.2	135.63	130.11	136.68
	Father education < 15	11.55	12.25	1.39	0.88	134.18	128.61	133.7
Gifted in comparison (main sample)	**Father education => 15**	17.1	15.9	2.37	1.09	136.3	130.2	136.93
	Father education < 15	12.13	12.7	1.5	0.91	133.27	128.19	133.76
Non-Gifted (comparison pull)	**Father education => 15**	16.81	15.68	1.98	1.04	120.14	111.31	122.34
	Father education < 15	11.85	12.58	1.14	0.76	114.52	106.74	115.09

Table 10 shows the results. Overall, GCP effects, particularly on long-term outcomes, do not vary by student's SES background. For example, the estimated effect on academic degrees attainment and field of studies are very similar, with little and no significant differences in the point estimates between the SES groups. In addition, the impact on the labor market and personal outcomes are

33 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.

almost entirely statistically insignificant, again not indicating any major difference between the groups.[34]

The insensitivity of the estimated effects of GCP to SES variation is in sharp contrast to the effects of many other schooling inputs, which varies by student's background.[35] However, it is important to note that the low SES group of GCP participants does not represent the low SES populations in Israel. For example, the average income for fathers in this "low SES" sample is 135,000 NIS, much higher than the corresponding average among the group of low SES (father education is less than 15 years) in the comparison pull (113,000 NIS).

6. 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 relatively 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

[34] Interestingly, the small negative effects on bagrut test scores are more pronounced for the higher SES group. For example, the effect on the average composite score is -0.041 for this group (above median father's schooling), with only -0.002 for the lower SES group (below median father's schooling). One potential explanation for this pattern is that the improvements in peers' quality are more important for the low SES students.

[35] 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).

‘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. Our sectoral classification is perhaps too coarse for capturing this potential effect, and we should instead use occupational categories within these sectors.

Against the benefit and gains accruing to participants in gifted children programs, we should note the potential loss to other students in the education system. There is some evidence that suggest that non-gifted children benefit from having high achievers and gifted children as peers (Lavy, Olmo, and Weinhardt, 2012, Balestraa, Aurélien and Stefan Wolterc. 2021). Thus, there is a concern that excluding gifted children from regular classes might have adverse effects on otherwise 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.

Angrist Joshua, and Miikka Rokkanen (2015) “Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cut-off,” *Journal of the American Statistical Association*, 110:512, 1331-1344.

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. 2021. “Does Remedial Education at Late Childhood Pay Off After All? Long-Run Consequences for University Schooling, Labor Market Outcomes and Inter-Generational Mobility.” Forthcoming, *Journal of Labor Economics*.

Leung , S. A. , Conoley , C . W. , & Sfheel , M.J . 1994. Th e caree r an d educationa l aspirations o f gifted hig h schoo l students : A retrospective studv. *Journal of Counseling & 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 Nijmengen. 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 rasing 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 o 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.

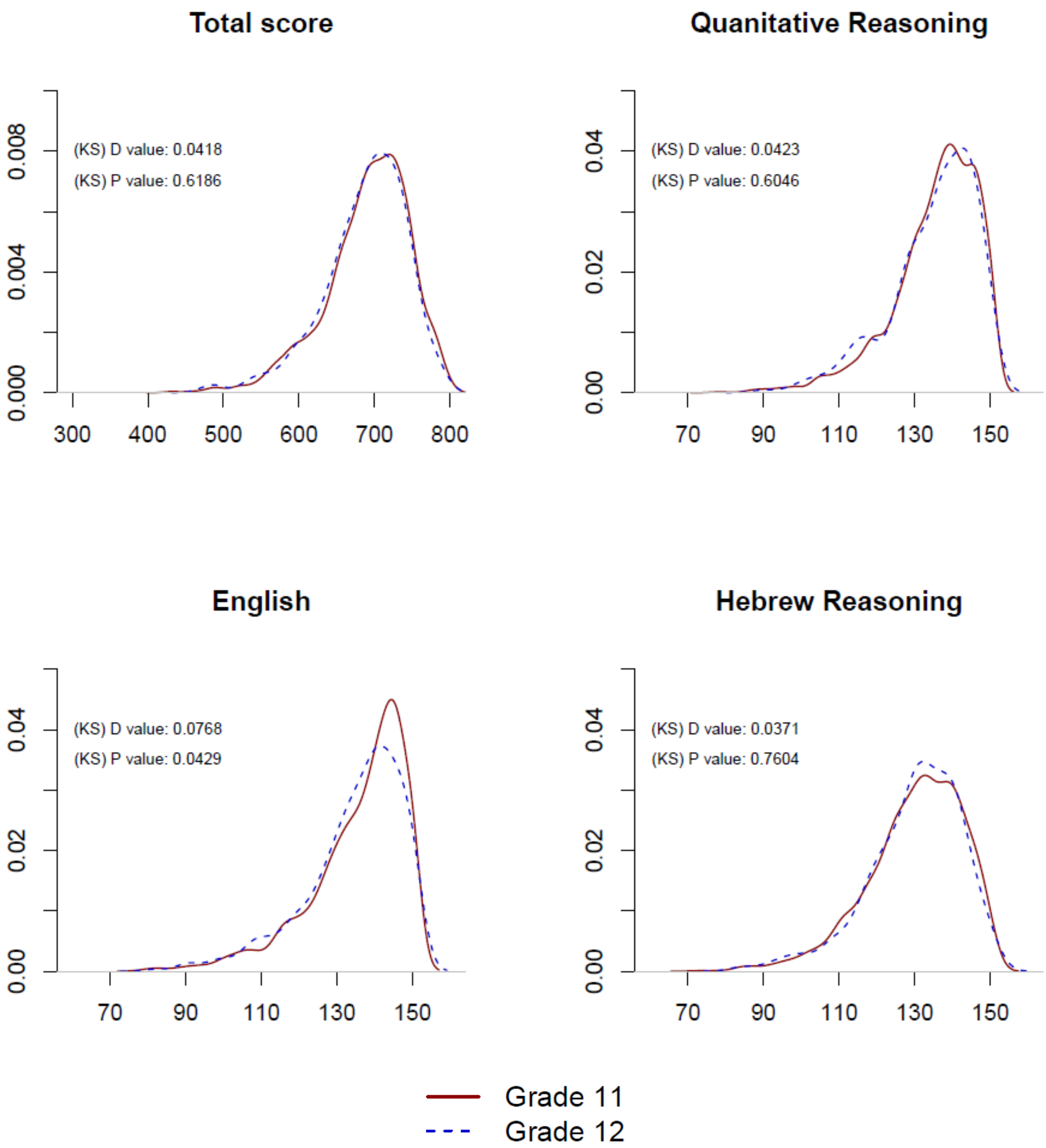
Zeidner, Moshe, and Inbal Shani-Zinovich. (2013) .“Research on Personality and Affective Dispositions of Gifted Children: The Israeli Scene.” Gifted and Talented International, 28:1-2, 35-50.

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.

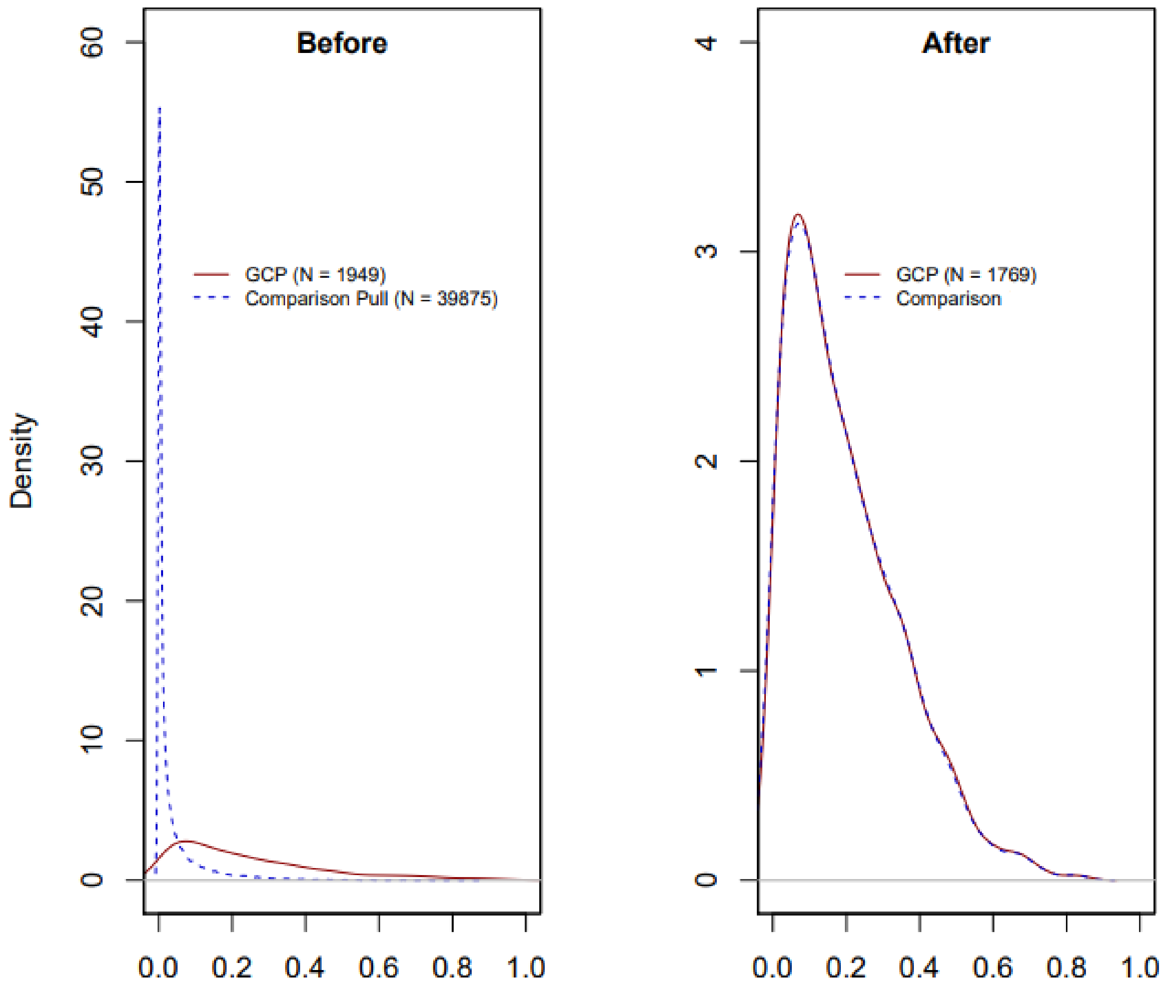
Zeidner, Moshe, and E. J. Schleyer. “The Big-Fish-Little-Pond Effect for Academic Self-Concept, Test Anxiety, and School Grades in Gifted Children.” *Contemp Educ Psychol*. 1999 Oct;24(4):305-329.

Figure 1: Psychometric Scores Distributions, by Grade of Testing



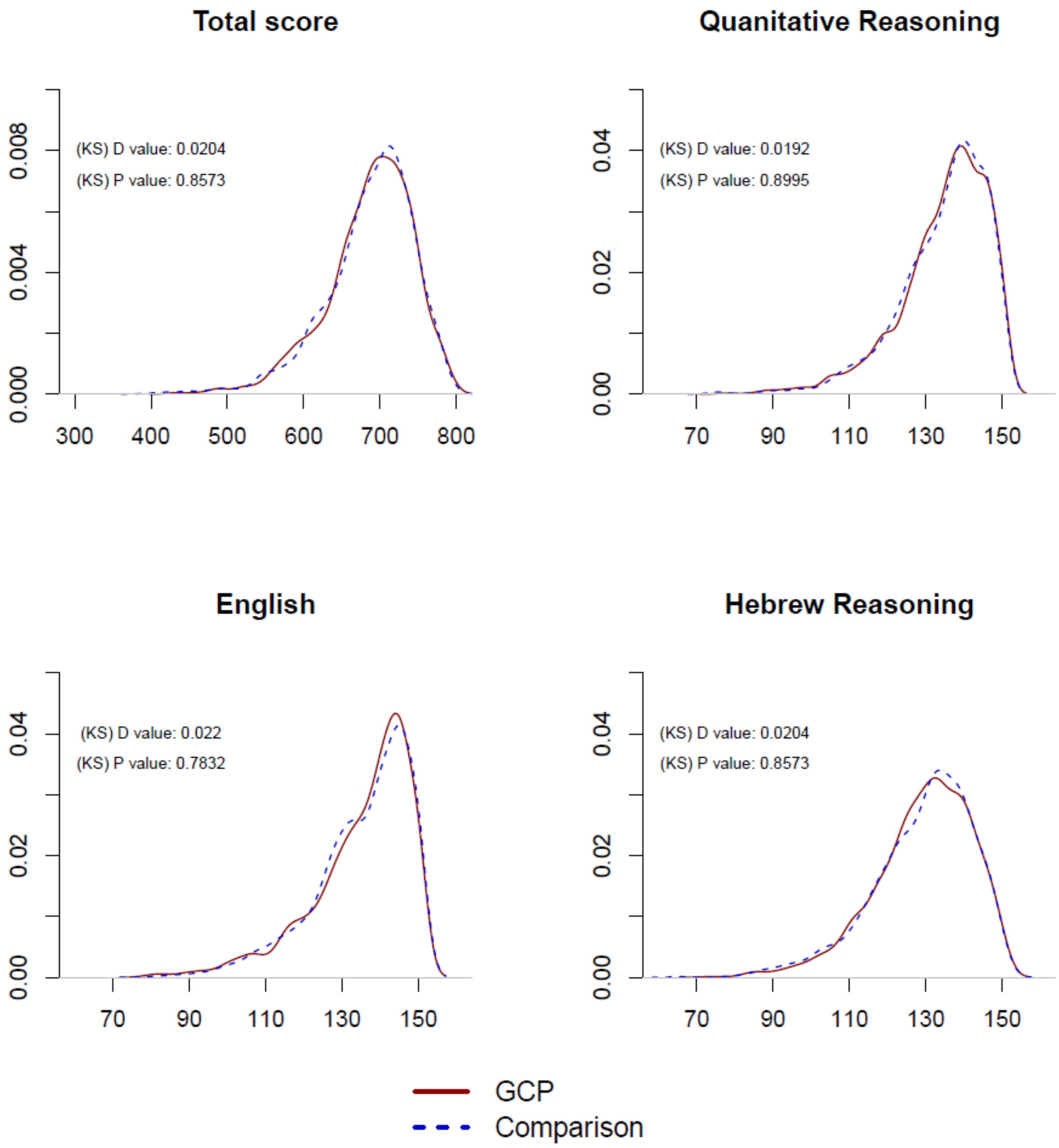
Notes: This figure plots the distribution of psychometric scores, by grade of taking the test- the red solid line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. 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 graduetes in 1992-2005.

Figure 2: Propensity Score Distributions, Before and After Matching



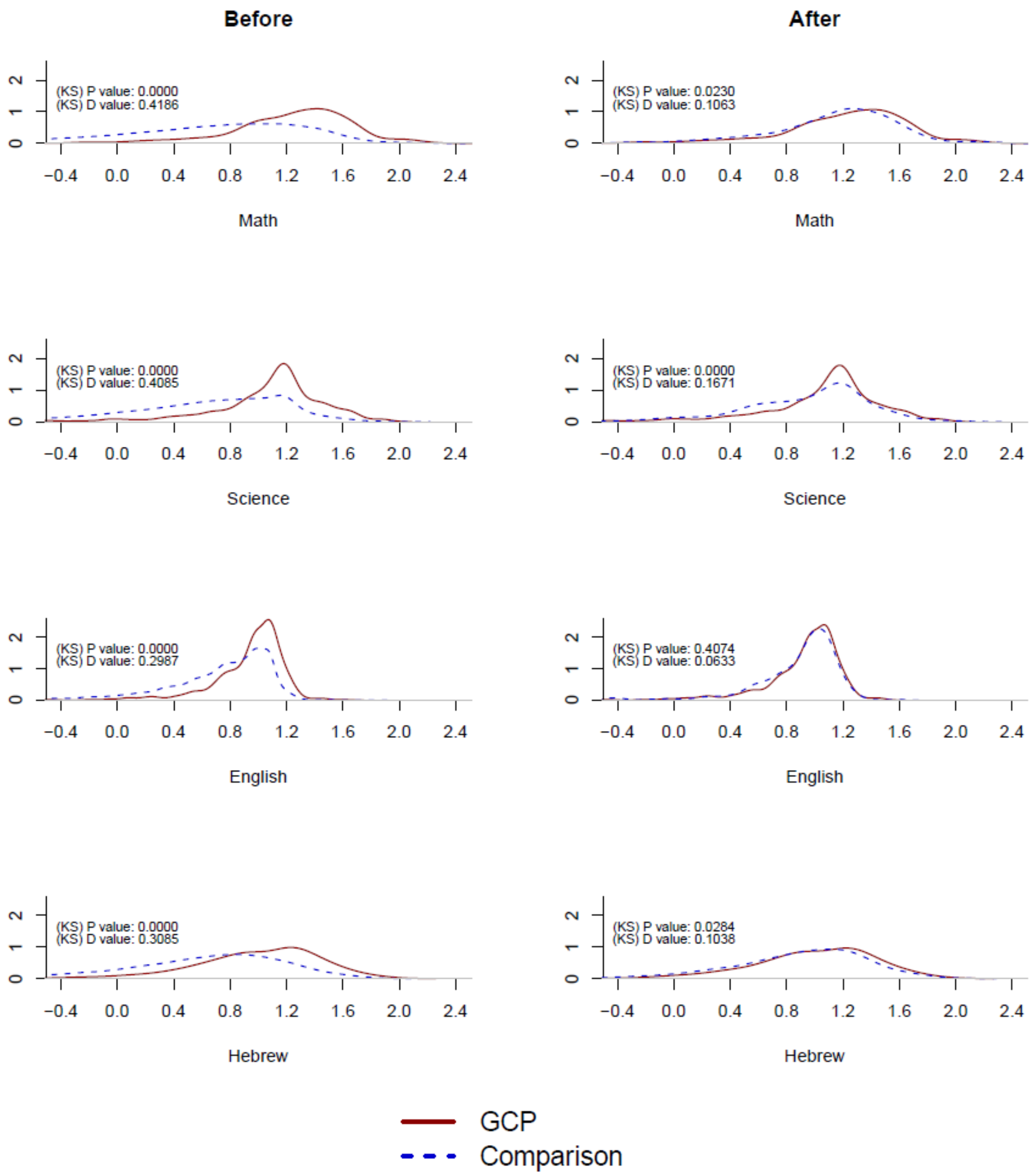
Notes: This figure plots the distribution of the propensity score, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching,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

Figure 3: Psychometric Scores Distributions, GCP Participants and Comparison Group



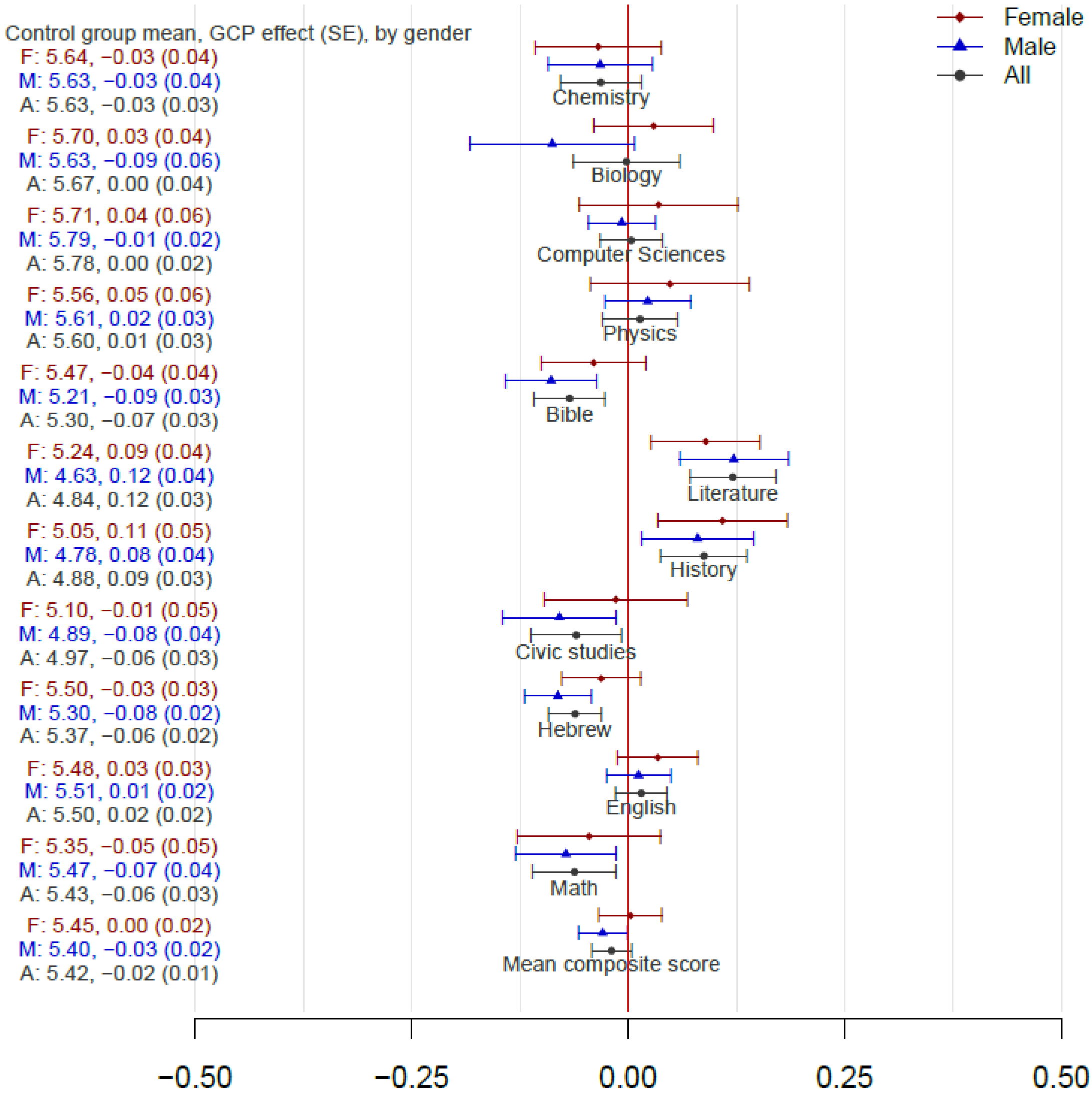
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the matched comparison group. The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions. The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

Figure 4: Pre-treatment Middle-school Test Scores, Before and After Matching I



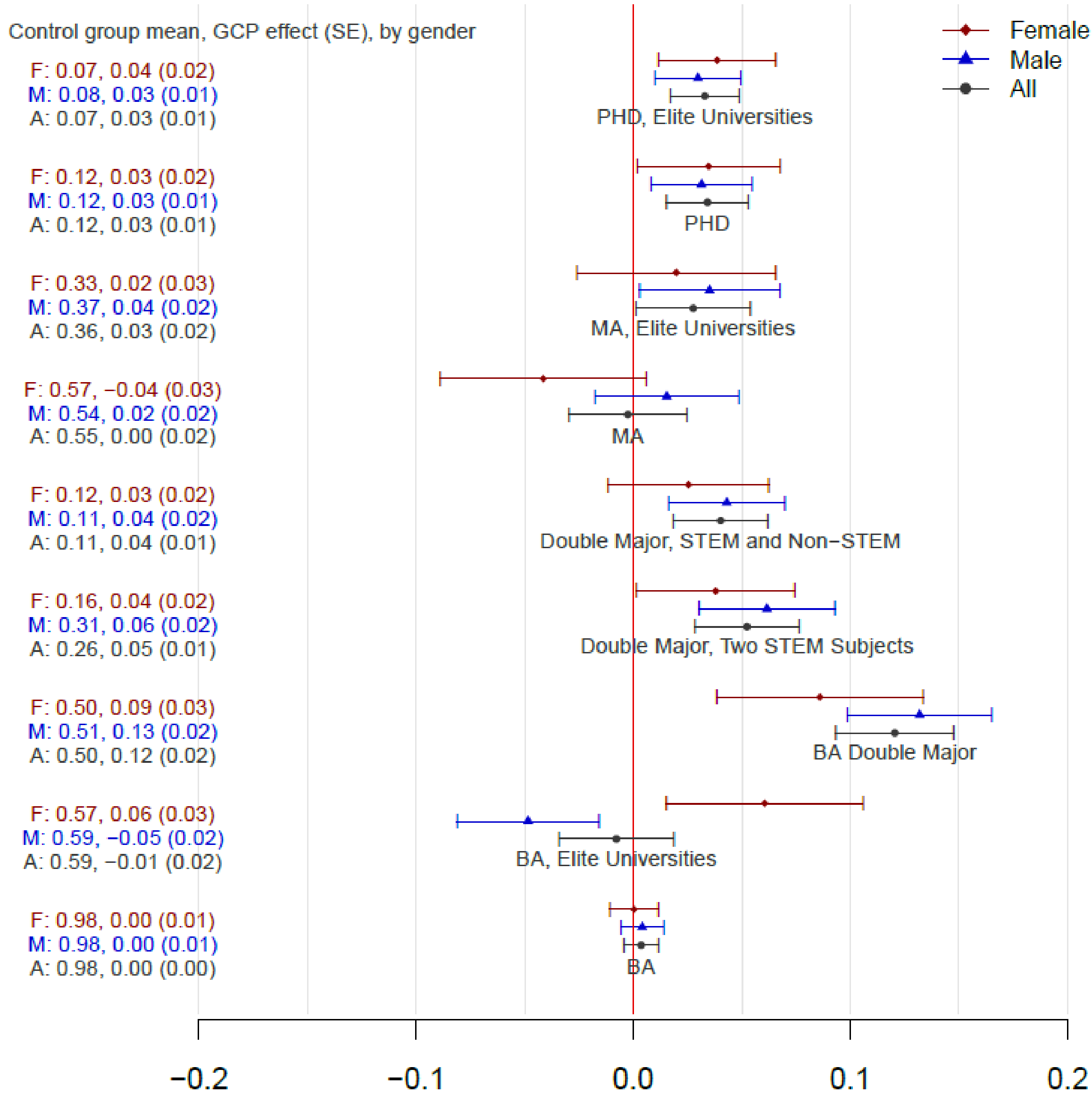
Notes: This figure plots the distribution of the Pre-treatment Middle-school Test (Metzav) test scores, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graphs on the left show the distributions before the matching, and the graphs on the right shows the distributions after the matching (version I, which does not include these tests in the logit regression).. 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, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Figure 5: GCP Effects on Bagrut Test Scores



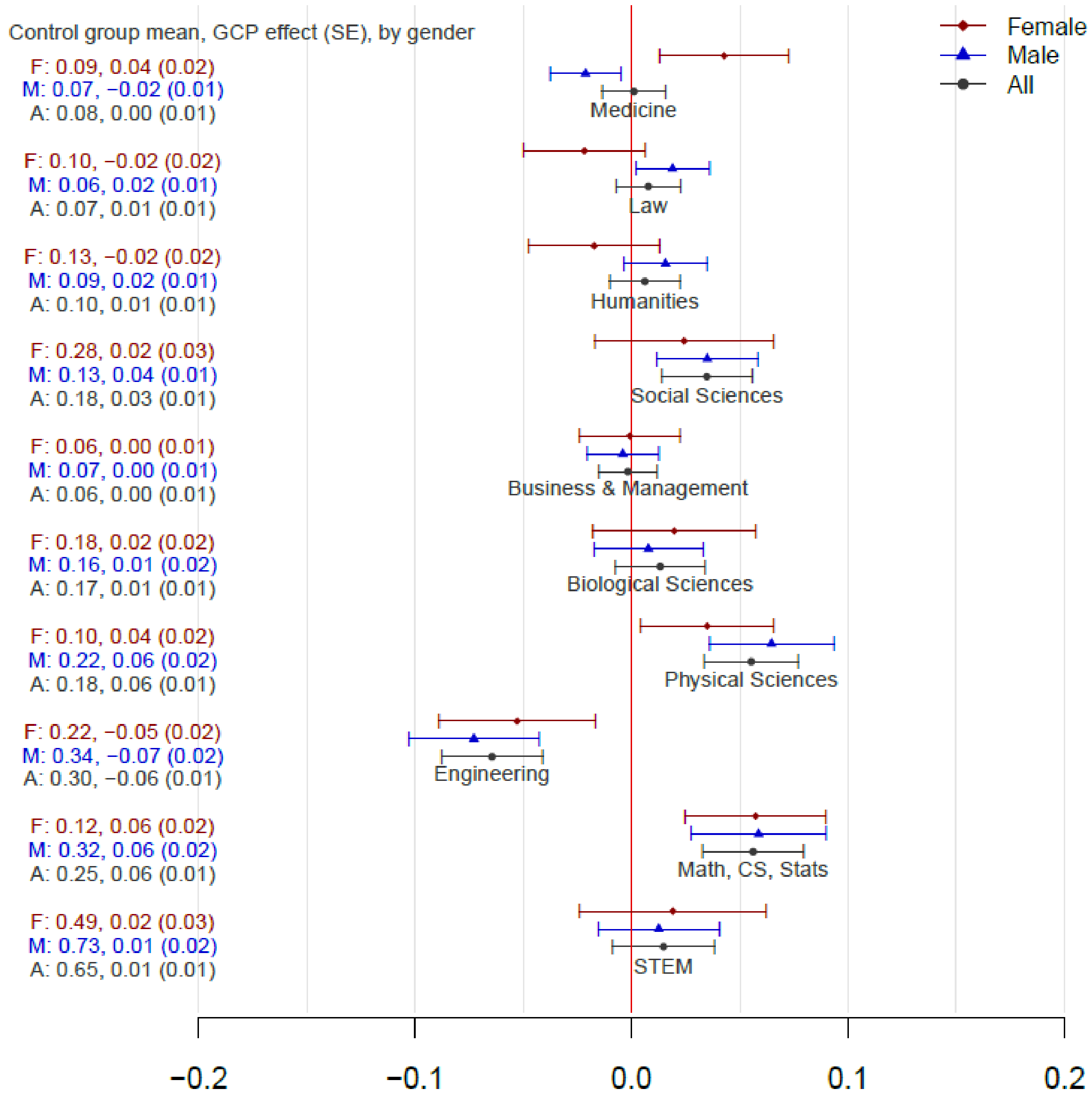
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory and most chosen elective subjects). These scores are normalised on a scale of 0-6. Red dots and lines represesnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The sample includes students from the cohorts of high-school graduetes in 1992-2005, who took the psychometric test during their 10-11th grade.

Figure 6: GCP Effects on University Degree Attainment



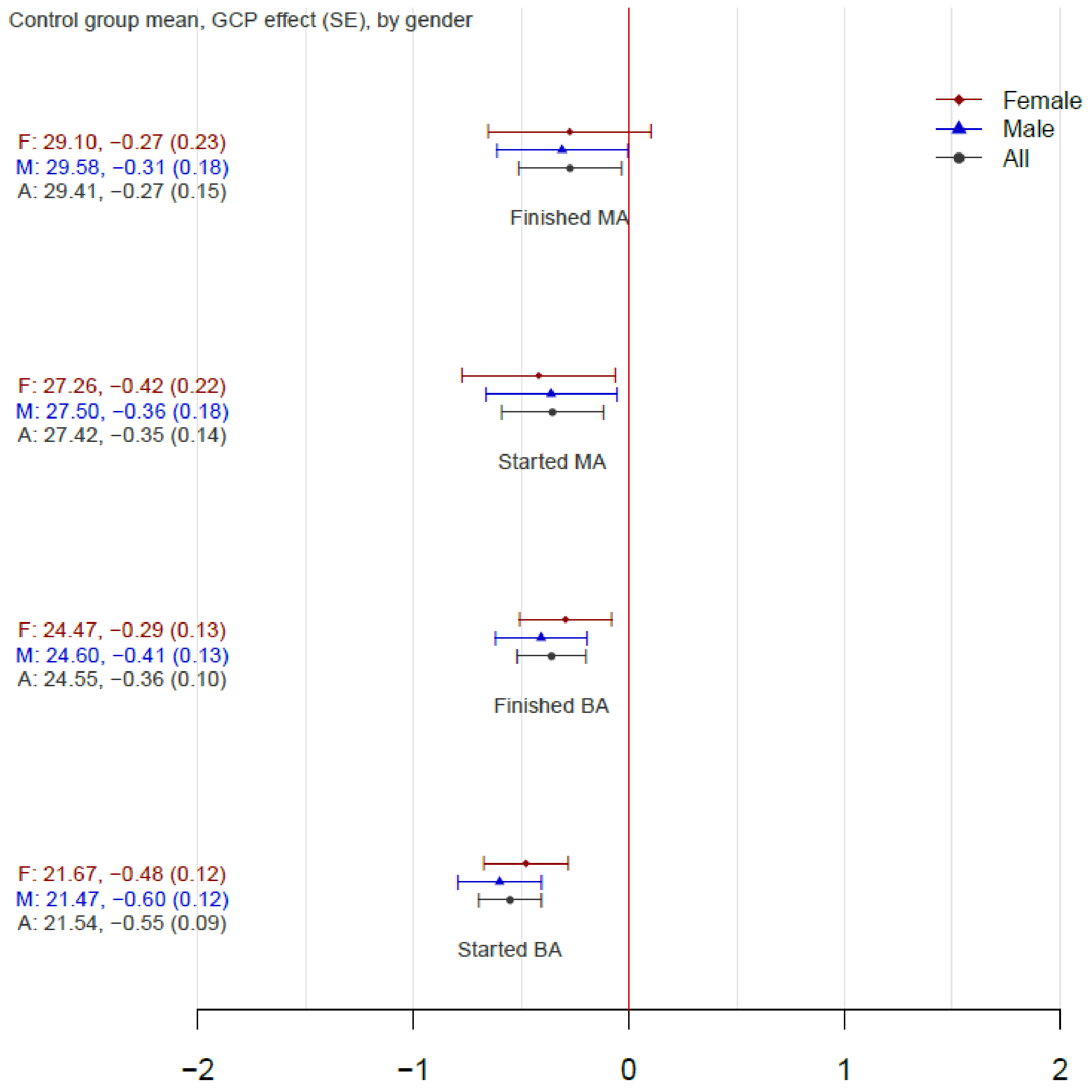
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on university degree attainment. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and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

Figure 7: GCP Effects on University Field of Studies



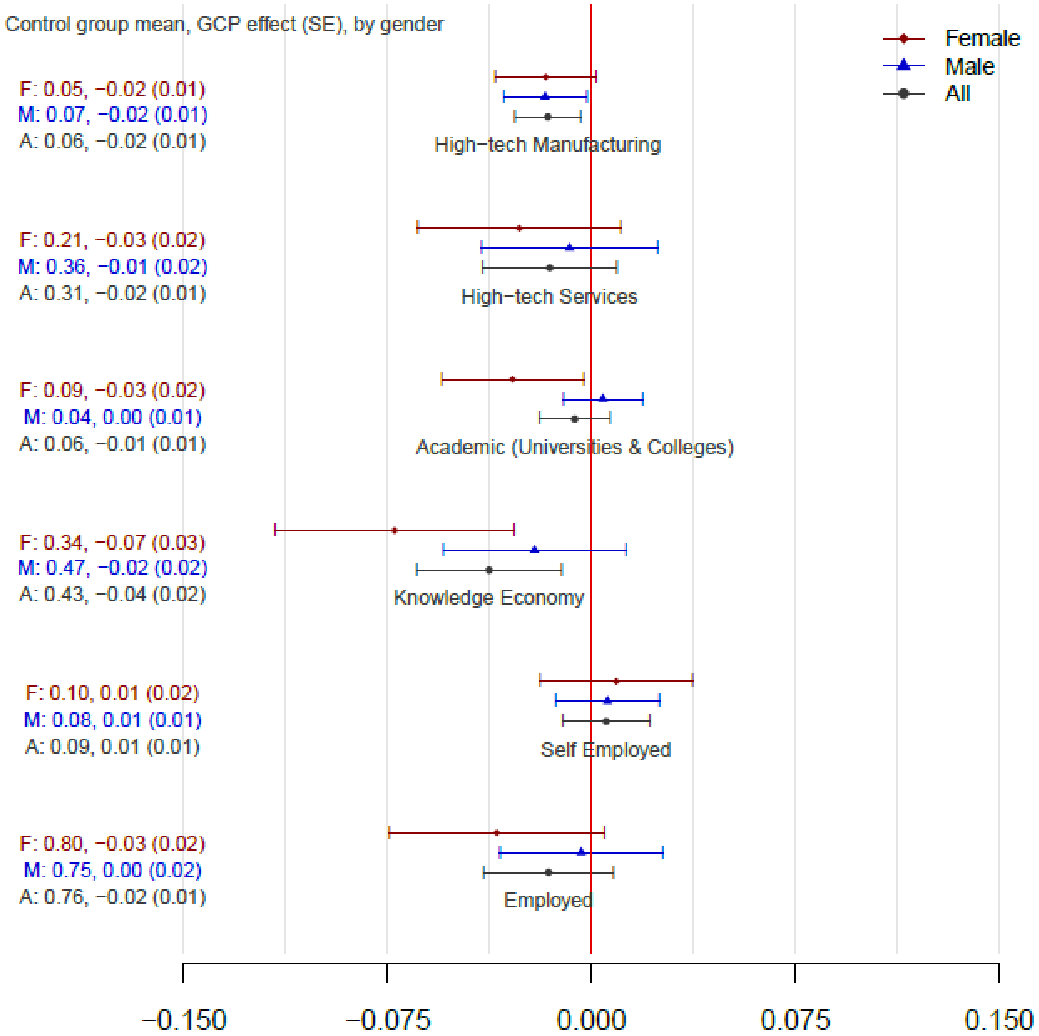
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on university field of studies. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females).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

Figure 8: GCP Effects on Studies Timing



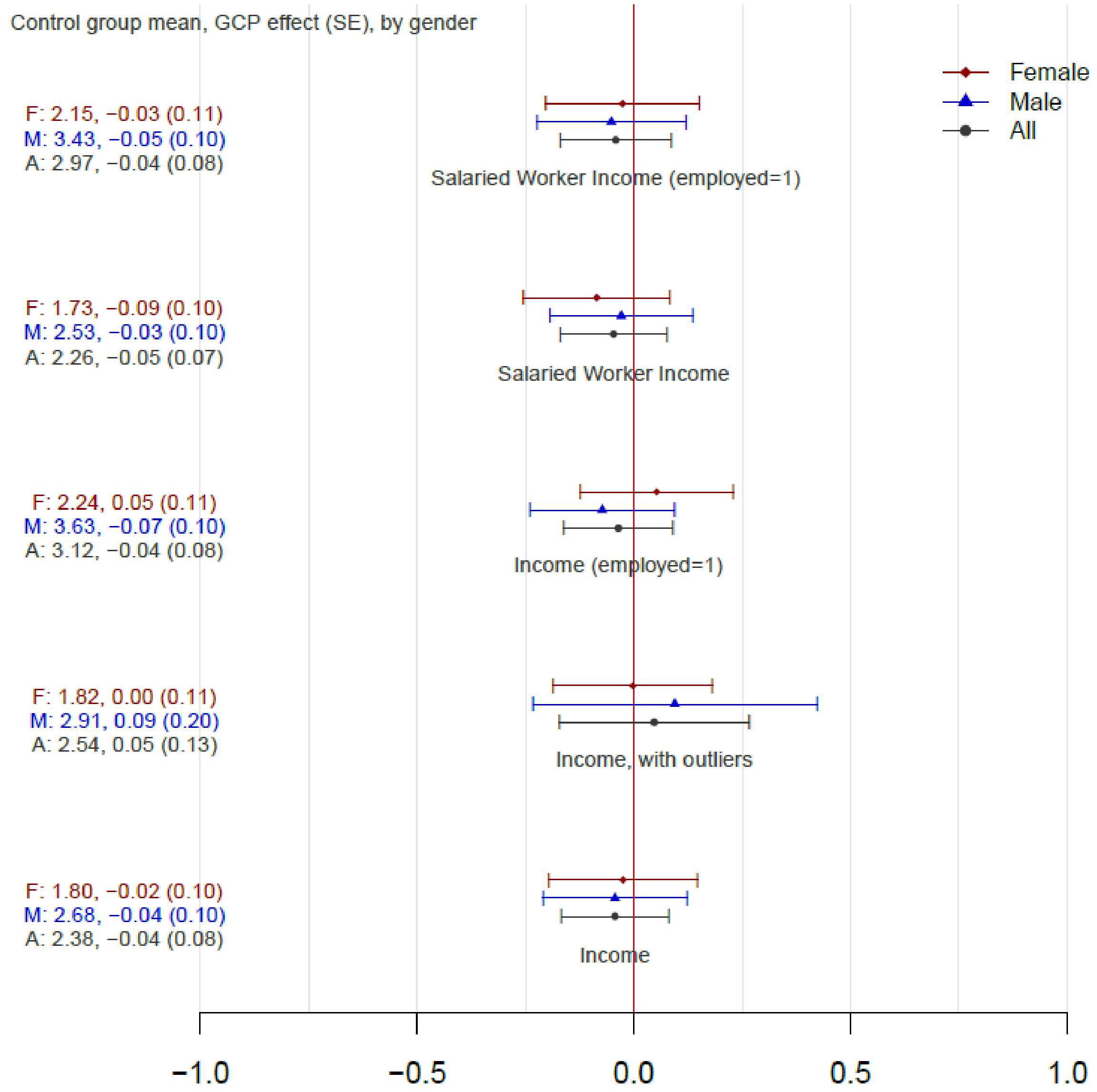
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on studies timing. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The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

Figure 9: GCP Effect on Employment, by Sector in 2018



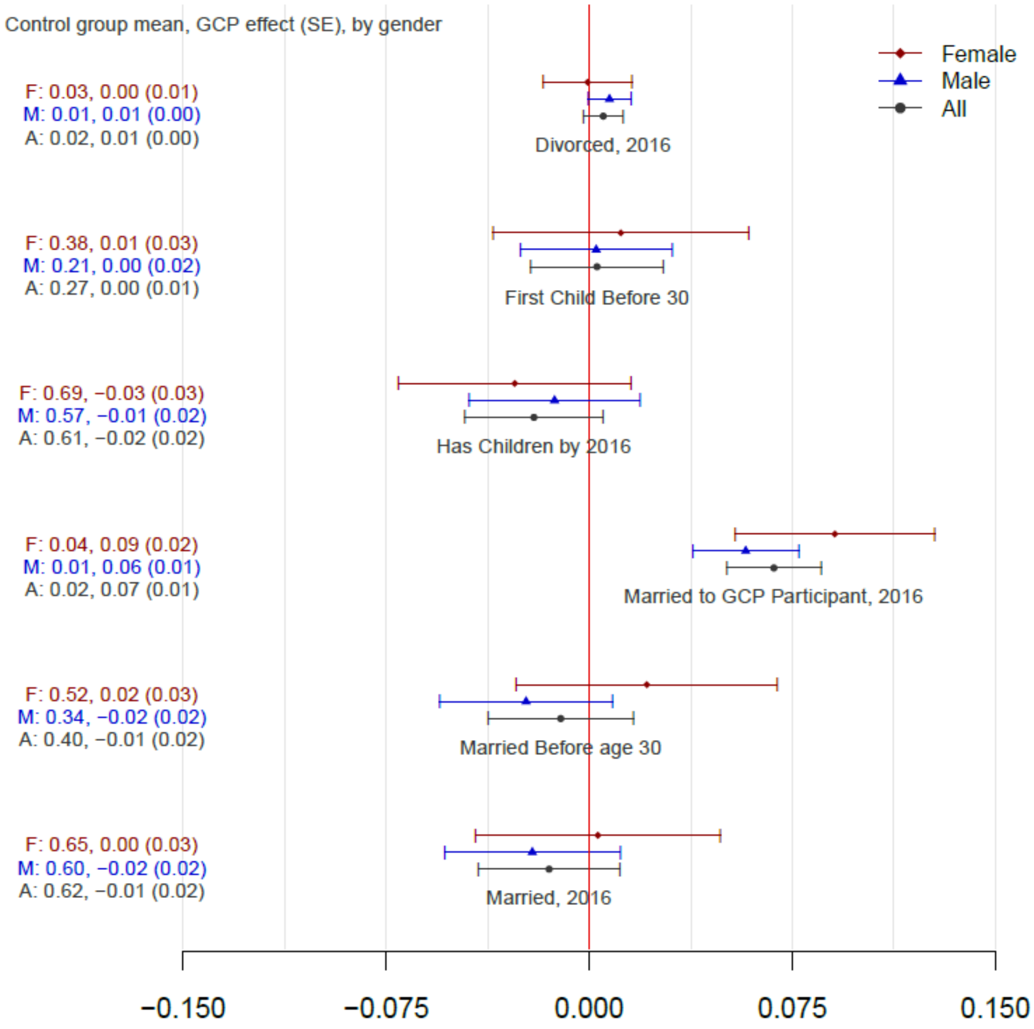
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on employment, by sector in 2018. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and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

<u>Figure 10</u>: GCP Effect on Annual Income, in 2018



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

Figure 11: GCP Effect on Personal Outcomes



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

Table 1: Descriptive Statistics, Demographics and Psycometric Scores, by Gender

	Girls				Boys			
	GCP	Control	difference	p-value	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Family Background:								
Father education	14.68	14.66	0.02	0.93	14.65	14.82	-0.17	0.30
Mother education	14.52	14.42	0.10	0.62	14.35	14.39	-0.04	0.80
Born in israel	0.86	0.86	0.00	1.00	0.87	0.87	0.00	1.00
Father Born in israel	0.52	0.52	0.00	1.00	0.52	0.53	-0.01	0.63
Mother born israel	0.55	0.55	0.00	1.00	0.58	0.59	-0.01	0.62
Father income 2009	1.83	1.88	-0.05	0.69	1.69	2.00	-0.31	0.00*
Mother income 2009	1.02	1.03	-0.01	0.90	1.07	1.00	0.07	0.23
Siblings	1.85	1.95	-0.10	0.16	1.95	1.86	0.09	0.10
B. Psychometric score:								
Total	672.17	669.28	2.89	0.43	694.77	695.59	-0.82	0.70
Quantative reasoning	129.46	128.42	1.04	0.18	129.40	129.69	-0.29	0.58
Verbal reasoning	130.26	130.19	0.07	0.93	137.43	137.33	0.10	0.81
English	131.70	131.39	0.31	0.69	137.21	137.56	-0.35	0.47
Number of Observations	608	608			1,161	1,161		

Notes : This table presents descriptive statistics for the Girls sample (columns 1-4) and the Boys sample (columns 5-8). Columns (1) and (5) show the means among the treatment group, Columns (2) and (6) show the means among the comparison group, Columns (3) and (7) and (4) and (8) show the difference between the means and the corresponded p-value. Income is measured in 100K NIS. The psychometric score (total) is between 200 and 800, and the psychometric score (numeric/hebrew/english) is between 50 and 150. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Table 2: Distribution of Bagrut Subjects at Advanced Level

	Girls				Boys			
	GCP	Control	difference	p-value	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.81	0.82	-0.01	0.65	0.93	0.92	0.01	0.35
Physics	0.34	0.38	-0.04	0.15	0.70	0.72	-0.02	0.29
Computer Sciences	0.31	0.24	0.07	0.01*	0.58	0.58	0.00	1.00
Chemistry	0.48	0.48	0.00	1.00	0.43	0.45	-0.02	0.33
Biology	0.29	0.29	0.00	1.00	0.12	0.12	0.00	1.00
Arabic	0.11	0.11	0.00	1.00	0.08	0.07	0.01	0.36
Literature	0.15	0.19	-0.04	0.06*	0.04	0.03	0.01	0.21
Social Sciences	0.11	0.12	-0.01	0.58	0.04	0.03	0.01	0.19
History	0.04	0.05	-0.01	0.40	0.05	0.04	0.01	0.22
Art	0.04	0.06	-0.02	0.11	0.02	0.02	0.00	1.00
Total credits	29.34	29.07	0.27	0.15	30.50	30.52	-0.02	0.90
Number of Observations	608	608			1,161	1,161		

Notes : This table presents descriptive statistics for the Girls sample (columns 1-4) and the Boys sample (columns 5-8). Columns (1) and (5) show the means among the treatment group, Columns (2) and (6) show the means among the comparison group, Columns (3) and (7) and (4) and (8) show the difference between the means and the corresponded p-value. Total credits refer to the total credits acuumulated in the bagrut. All other variables are indicator variables for 5 credits bagrut in the subject. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Table 3: Bagrut Exams Test Scores in Compulsory and Selective Electives Subjects

	Girls				Boys			
	GCP	Control	difference	p-value	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Cumpolsory subjects:								
Mean composite score	5.45	5.45	0.01	1.00	5.37	5.40	-0.03	0.16
English	5.52	5.46	0.06	0.14	5.51	5.51	0.00	1.00
Math	5.32	5.35	-0.03	0.60	5.40	5.47	-0.07	0.08*
Bible	5.43	5.47	-0.04	0.34	5.10	5.21	-0.11	0.00*
Literature	5.30	5.24	0.06	0.16	4.74	4.63	0.11	0.01*
History	5.13	5.05	0.08	0.10*	4.88	4.78	0.10	0.02*
Civic studies	5.08	5.10	-0.02	0.72	4.83	4.89	-0.06	0.15
Hebrew	5.46	5.50	-0.04	0.23	5.22	5.30	-0.08	0.00*
B. Elective subjects:								
Physics	5.55	5.56	-0.01	0.88	5.63	5.61	0.02	0.54
Computer Sciences	5.70	5.71	-0.01	0.86	5.79	5.79	0.00	1.00
Chemistry	5.60	5.64	-0.04	0.47	5.63	5.63	0.00	1.00
Biology	5.77	5.70	0.07	0.19	5.61	5.63	-0.02	0.79
Number of Observations	608	608			1,161	1,161		

Notes: This table presents descriptive statistics for the Girls sample (columns 1-4) and the Boys sample (columns 5-8). Columns (1) and (5) show the means among the treatment group, Columns (2) and (6) show the means among the comparison group, Columns (3) and (7) and (4) and (8) show the difference between the means and the corresponded p-value. All scores are normalised between 1 and 6. We included the four most popular elective subject among gifted children in our data. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Table 4: Descriptive Statistics, University Degrees and Field of Study

	Girls				Boys			
	GCP	Control	difference	p-value	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Degrees:								
BA	0.99	0.98	0.01	0.18	0.98	0.98	0.00	1.00
BA, Double Major	0.60	0.50	0.10	0.00*	0.64	0.51	0.13	0.00*
MA	0.54	0.57	-0.03	0.29	0.55	0.54	0.01	0.63
PHD	0.16	0.12	0.04	0.04*	0.15	0.12	0.03	0.03*
B. Field of Study:								
STEM	0.51	0.49	0.02	0.49	0.73	0.73	0.00	1.00
Math, Computer Sciences, Statistics	0.19	0.12	0.07	0.00*	0.38	0.32	0.06	0.00*
Engineering	0.16	0.22	-0.06	0.01*	0.26	0.34	-0.08	0.00*
Physical Schiences	0.14	0.10	0.04	0.03*	0.28	0.22	0.06	0.00*
Biological Sciences	0.20	0.18	0.02	0.38	0.17	0.16	0.01	0.52
Business and Management	0.06	0.06	0.00	1.00	0.06	0.07	-0.01	0.33
Social Sciences	0.31	0.28	0.03	0.25	0.17	0.13	0.04	0.01*
Humanities	0.12	0.13	-0.01	0.60	0.11	0.09	0.02	0.10
Law	0.09	0.10	-0.01	0.56	0.08	0.06	0.02	0.05*
Medicine	0.14	0.09	0.05	0.01*	0.05	0.07	-0.02	0.04*
Number of Observations	608	608			1,161	1,161		

Notes: This table presents descriptive statistics for the Girls sample (columns 1-4) and the Boys sample (columns 5-8). Columns (1) and (5) show the means among the treatment group, Columns (2) and (6) show the means among the comparison group, Columns (3) and (7) and (4) and (8) show the difference between the means and the corresponded p-value. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Table 5: Descriptive Statistics, Labor Market and Personal Outcomes

	Girls				Boys			
	GCP	Control	difference	p-value	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Labor:								
Income Salaried Worker	1.65	1.73	-0.08	0.43	2.48	2.53	-0.05	0.63
Income Salaried Worker (employed=1)	2.15	2.15	0.00	1.00	3.36	3.43	-0.07	0.53
Income (employed=1)	2.30	2.24	0.06	0.59	3.53	3.63	-0.10	0.36
Income (employed=1), with outliers	2.34	2.26	0.08	0.51	3.89	3.87	0.02	0.93
Employed as salaried worker	0.71	0.77	-0.06	0.02*	0.70	0.71	-0.01	0.60
Self employed	0.11	0.10	0.01	0.57	0.09	0.08	0.01	0.39
Employed	0.76	0.80	-0.04	0.09*	0.74	0.75	-0.01	0.29
B. Employment Secotrs:								
Knowledge economy	0.28	0.34	-0.06	0.02*	0.45	0.47	-0.02	0.33
High tech manufacturing	0.03	0.05	-0.02	0.07*	0.05	0.07	-0.02	0.04*
High tech services	0.18	0.21	-0.03	0.19	0.35	0.36	-0.01	0.62
Academic	0.06	0.09	-0.03	0.05*	0.05	0.04	0.01	0.26
C. Personal:								
Married	0.65	0.65	0.00	1.00	0.57	0.60	-0.03	0.14
Married Before Age 30	0.53	0.52	0.01	0.73	0.32	0.34	-0.02	0.45
Divorced	0.03	0.03	0.00	1.00	0.02	0.01	0.01	0.04*
Number of children	1.45	1.46	-0.01	0.89	1.12	1.20	-0.08	0.13
First Child Before 30	0.40	0.38	0.02	0.24	0.22	0.21	0.01	0.46
Age at marriage	28.39	28.61	-0.22	0.31	30.32	30.25	0.07	0.67
Age at first child	30.13	30.28	-0.15	0.48	31.53	31.46	0.07	0.68
GCP Partner	0.13	0.04	0.09	0.00*	0.07	0.01	0.06	0.00*
Same School Partner	0.13	0.07	0.06	0.00*	0.10	0.09	0.01	0.90
Same Class Partner	0.07	0.01	0.07	0.00*	0.04	0.03	0.01	0.62
Partner's Psychometric Score	650.43	645.05	5.38	0.43	612.39	599.75	12.64	0.02*
Partner's Income 2018	2.39	2.47	-0.08	0.69	1.23	1.23	0.00	0.45
Number of Observations	608	608			1,161	1,161		

Notes: This table presents descriptive statistics for the Girls sample (columns 1-4) and the Boys sample (columns 5-8). Columns (1) and (5) show the means among the treatment group, Columns (2) and (6) show the means among the comparison group, Columns (3) and (7) and (4) and (8) 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 graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Table 6: Treatment Effect Estimates, Girls, by Alternative Samples

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	0.00	-0.05	-0.03	-0.04	-0.04	0.03*	0.09*	0.04*	0.06*	-0.05*	0.04*
N: 1,216	(0.02)	(0.05)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Main only numeric	0.06*	-0.09*	0.04	0.04	0.01	0.04*	0.13*	0.04*	0.01	-0.07*	0.06*
N: 1,492	(0.02)	(0.05)	(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Extend 12	0.00	-0.07	0.00	-0.05*	-0.01	0.01	0.07*	0.05*	0.06*	-0.05*	0.05*
N: 1,492	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Extend all	0.01	-0.07	-0.01	-0.07*	0.00	0.03*	0.09*	0.03*	0.01	-0.04*	0.05*
N: 2,120	(0.02)	(0.05)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Extend all only numeric	0.11*	-0.10*	0.06*	0.09*	0.03	0.04*	0.14*	0.02	0.01	-0.06*	0.06*
N: 2,120	(0.02)	(0.05)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(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	-0.03	0.01	-0.09	-0.03	0.05	-0.03	-0.02	-0.07*	-0.03*	0.02	0.01
N: 1,216	(0.02)	(0.02)	(0.10)	(0.11)	(0.11)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)
Main only numeric	-0.03	0.02	-0.17	-0.16	-0.10	-0.03	-0.03*	-0.05*	0.02	-0.02	-0.07*
N: 1,216	(0.02)	(0.02)	(0.10)	(0.11)	(0.11)	(0.02)	(0.01)	(0.03)	(0.01)	0.03	0.03
Extend 12	-0.06*	-0.01	-0.14	-0.03	0.02	-0.03	-0.02	-0.06*	-0.02*	-0.01	-0.04
N: 1,492	(0.02)	(0.02)	(0.09)	(0.10)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)
Extend all	-0.05*	0.00	-0.16*	-0.10	-0.10	-0.03	0.00	-0.03*	0.00	-0.03	-0.04*
N: 2,120	(0.02)	(0.01)	(0.07)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	0.02	0.02
Extend all only numeric	-0.05*	0.00	-0.21*	-0.15*	-0.13*	-0.02	-0.02*	-0.03	0.01	-0.04*	-0.05*
N: 2,120	(0.02)	(0.01)	(0.07)	(0.07)	(0.07)	(0.02)	(0.01)	(0.02)	(0.01)	0.03	0.02

*Notes: This table presents the treatment effect estimated on girls, by sample. Each row represent different sample, and each column represent different outcome variable. * represents statistical significance at the 90% level.*

Table 7: Treatment Effect Estimates, Boys, by alternative Samples

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	-0.03*	-0.07*	-0.08*	-0.09*	0.02	0.03*	0.13*	0.06*	0.06*	-0.07*	0.06*
N: 2,322	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Main only numeric	0.03	-0.06*	-0.05*	0.01	0.02	0.04*	0.12*	0.06*	0.07*	-0.10*	0.09*
N: 2,322	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Extend 12	-0.04*	-0.08*	-0.08*	-0.10*	0.02	0.04*	0.09*	0.07*	0.05*	-0.07*	0.08*
N: 2,844	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Extend all	-0.02	-0.10*	-0/07*	-0.08*	-0.01	0.03*	0.11*	0.06*	0.07*	-0.08*	0.08*
N: 3,768	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Extend all only numeric	0.07*	-0.04	0.02	0.04	0.01	0.03*	0.13*	0.05*	0.06*	-0.09*	0.10*
N: 2,120	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(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	0.00	0.01	-0.03	-0.05	-0.07	-0.01	-0.02*	-0.02	0.00	-0.02	0.00
N: 2,322	(0.02)	(0.01)	(0.10)	(0.11)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Main only numeric	-0.02	0.01	-0.09	-0.06	-0.06	-0.01	-0.01	-0.02	0.00	-0.02	0.01*
N: 2,322	(0.02)	(0.01)	(0.10)	(0.11)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Extend 12	-0.03*	0.00	-0.06	0.02	-0.03	0.00	-0.01	-0.01	-0.01	-0.03	0.00
N: 2,844	(0.02)	(0.01)	(0.09)	(0.09)	(0.09)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Extend all	-0.02	0.00	-0.08	-0.04	-0.07	-0.01	-0.02*	-0.02	0.00	0.00	0.02
N: 3,768	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Extend all only numeric	-0.03*	0.02*	-0.17*	-0.09	-0.06	0.00	-0.02*	-0.01	0.01	-0.02	-0.01
N: 3,768	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)

*Notes: This table presents the treatment effect estimated on girls, by sample. Each row represent different sample, and each column represent different outcome variable. * represents statistical significance at the 90% level.*

Table 8: Treatment Effect Estimates, by Matching Versions (2006-2010 Cohorts Sample)

Panel A.	**Bagrut Exam Scores**								
Outcome:	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Matching I	-1.40*	-3.20*	-1.45*	-2.21*	0.20*	0.21*	0.06*	-0.07*	0.06*
N: 790	(0.31)	(0.78)	(0.36)	(0.51)	(0.04)	(0.12)	(0.03)	(0.03)	(0.03)
Matching II	-0.08	-1.69*	-0.09	-1.04*	0.18*	0.16*	0.06*	-0.10*	0.12*
N: 778	(0.35)	(0.86)	(0.44)	(0.54)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Matching III	-1.02*	-3.45*	-0.94*	-1.68*	0.17*	0.14*	0.11*	-0.08*	0.07*
N: 786	(0.32)	(0.74)	(0.42)	(0.50)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)

*Notes: This table presents the treatment effect estimated on boys and girls, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age. Each column represent different outcome variable. * represents statistical significance at the 90% level.*

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

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:											
Control Group Mean	5.127	5.070	5.111	5.047	0.483	0.078	0.431	0.177	0.164	0.311	0.120
GCP Effect	0.001	-0.115*	-0.004	-0.041	-0.007	0.017	0.139*	0.059*	0.024	-0.091*	0.047*
	(0.027)	(0.061)	(0.032)	(0.046)	(0.028)	(0.016)	(0.028)	(0.023)	(0.021)	(0.024)	(0.020)
Middle:											
Control Group Mean	5.483	5.501	5.423	5.363	0.554	0.114	0.511	0.255	0.246	0.302	0.144
GCP Effect	-0.058*	-0.078	-0.104*	-0.164*	-0.006	0.033*	0.125*	0.061*	0.079*	-0.058*	0.086*
	(0.025)	(0.052)	(0.035)	(0.044)	(0.029)	(0.020)	(0.029)	(0.026)	(0.026)	(0.025)	(0.023)
Upper:											
Control Group Mean	5.658	5.740	5.585	5.509	0.615	0.178	0.571	0.349	0.359	0.287	0.272
GCP Effect	-0.002	0.012	-0.081*	-0.003	0.007	0.054*	0.096*	0.054*	0.068*	-0.043*	0.036
	(0.019)	(0.036)	(0.029)	(0.039)	(0.028)	(0.023)	(0.028)	(0.027)	(0.027)	(0.025)	(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 manufacturin g 2018	Knowledge 2018	Academic 2018	Married Before 30	First Child Before 30
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Lower:											
Control Group Mean	0.798	0.098	1.983	2.474	2.586	0.267	0.055	0.385	0.063	0.420	0.298
GCP Effect	-0.025	-0.010	-0.153	-0.115	-0.078	-0.005	-0.011	-0.039	-0.025*	0.009	0.023
	(0.023)	(0.017)	(0.105)	(0.107)	(0.105)	(0.025)	(0.012)	(0.027)	(0.013)	(0.027)	(0.026)
Middle:											
Control Group Mean	0.801	0.104	2.435	3.045	3.233	0.315	0.070	0.443	0.058	0.413	0.280
GCP Effect	-0.039	-0.004	-0.007	0.083	0.077	-0.009	-0.024*	-0.036	-0.003	-0.009	0.005
	(0.025)	(0.018)	(0.137)	(0.144)	(0.140)	(0.027)	(0.013)	(0.029)	(0.014)	(0.029)	(0.026)
Upper:											
Control Group Mean	0.692	0.062	2.374	3.498	3.667	0.349	0.055	0.462	0.058	0.363	0.221
GCP Effect	0.016	0.031*	0.028	-0.086	-0.103	-0.012	-0.018	-0.014	0.014	-0.032	-0.020
	(0.027)	(0.016)	(0.145)	(0.157)	(0.152)	(0.027)	(0.012)	(0.029)	(0.015)	(0.028)	(0.025)
Number of Observations	2,368										

Notes: This table presents the treatment effect heterogeneity, by gifted levels. Gifted levels are "Lower", "Middle", and "Higher". Lower includes all gifted students with lower psychometric score (below 675). Middle includes all gifted students with psychometric score between 676 and 721. Lower includes all gifted students with higher psychometric score (above 722). Each column represent different outcome variable. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test during their 10-11th grade. The effects are calculated based on the estimates of the heterogeneity estimation equation described in the text (omitted group: the lowest third). * *represents statistical significance at the 90% level.*

Table 10: Treatment Estimated Effect Heterogeneity, by Socio-Economic Background

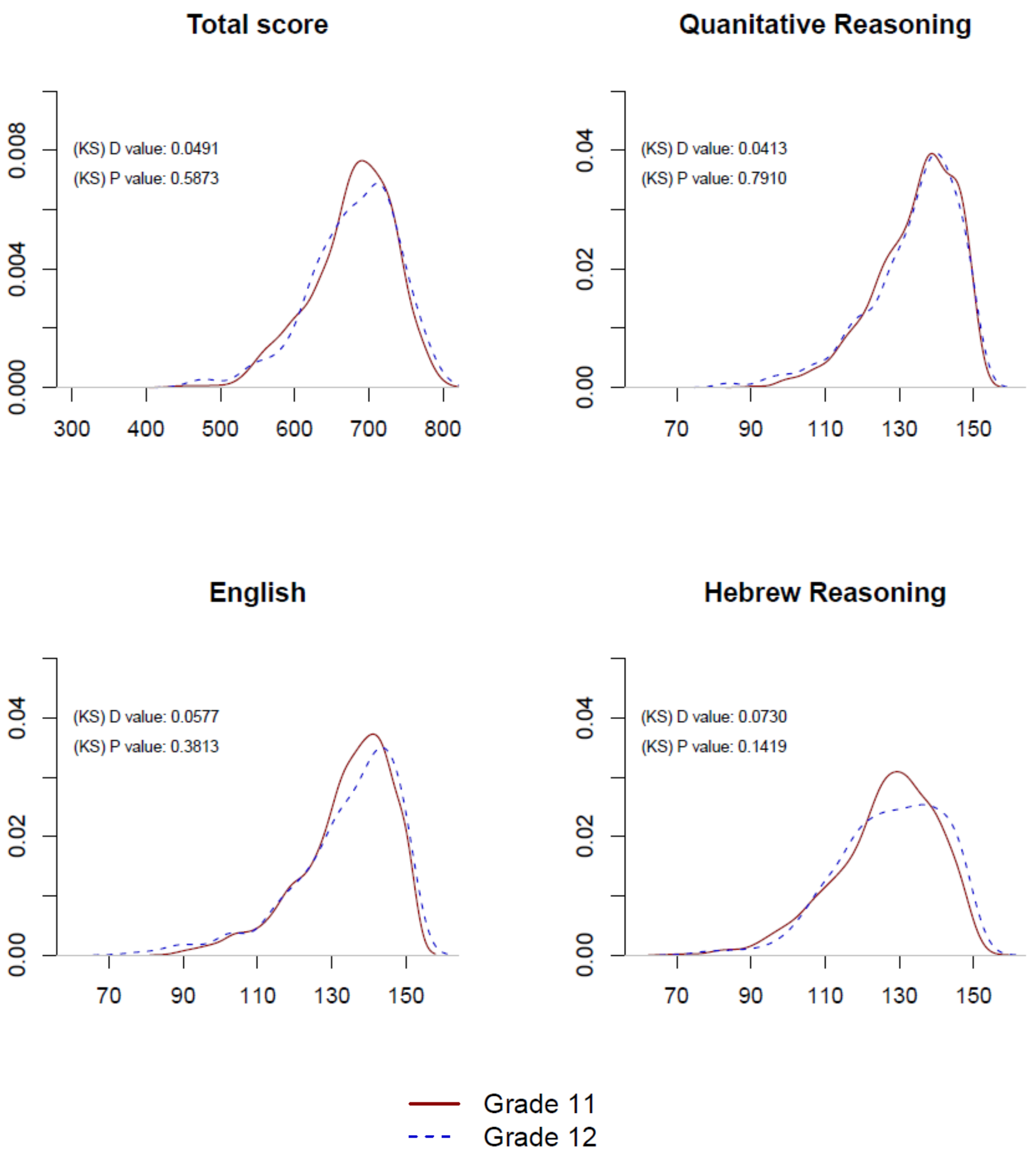
Panel A. Outcome:	Mean Bagrut	Math Bagrut	Hebrew Bagrut	Bible Bagrut	MA	PHD	BA Double Major	Double Major in STEM	Math, CS, Stats	Engineering	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Father Education < 15y											
Control Group Mean	5.340	5.310	5.320	5.260	0.550	0.120	0.490	0.240	0.240	0.300	0.160
GCP Effect	-0.002	-0.041	-0.046*	-0.054*	-0.002	0.035*	0.147*	0.048*	0.037*	-0.067*	0.044*
	(0.021)	(0.045)	(0.028)	(0.037)	(0.025)	(0.017)	(0.025)	(0.022)	(0.021)	(0.021)	(0.019)
Number of Observations	1,640										
Father Education ≥ 15y											
Control Group Mean	5.480	5.530	5.410	5.330	0.550	0.120	0.520	0.270	0.270	0.300	0.190
GCP Effect	-0.041*	-0.089*	-0.085*	-0.084*	0.001	0.029	0.102*	0.064*	0.070*	-0.058*	0.060*
	(0.019)	(0.040)	(0.026)	(0.035)	(0.023)	(0.016)	(0.023)	(0.020)	(0.020)	(0.020)	(0.019)
Number of Observations	1,898										

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 manufacturin g 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											
Control Group Mean	0.780	0.090	2.240	2.870	3.010	0.310	0.060	0.430	0.060	0.420	0.310
GCP Effect	-0.006	0.016	-0.023	-0.024	0.015	-0.030	-0.007	-0.049*	-0.013	-0.016	0.011
	(0.021)	(0.015)	(0.107)	(0.110)	(0.108)	(0.022)	(0.011)	(0.024)	(0.011)	(0.025)	(0.023)
Number of Observations	1,640										
Father Education ≥ 15y											
Control Group Mean	0.750	0.090	3.230	3.060	2.270	0.310	0.060	0.420	0.060	0.380	0.240
GCP Effect	-0.007	-0.006	-0.115	-0.086	-0.056	-0.010	-0.019	-0.030	-0.001	-0.001	0.006
	(0.020)	(0.013)	(0.113)	(0.115)	(0.108)	(0.021)	(0.009)	(0.022)	(0.011)	(0.022)	(0.020)
Number of Observations	1,898										

Notes: This table presents the treatment effect heterogeneity, by socio-economic status, proxied by father education. Each column represent different outcome variable. The sample includes students from the cohorts of high-school graduetes in 1992-2004, who took the psychometric test 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 sepearely. * *represents statistical significance at the 90% level.*

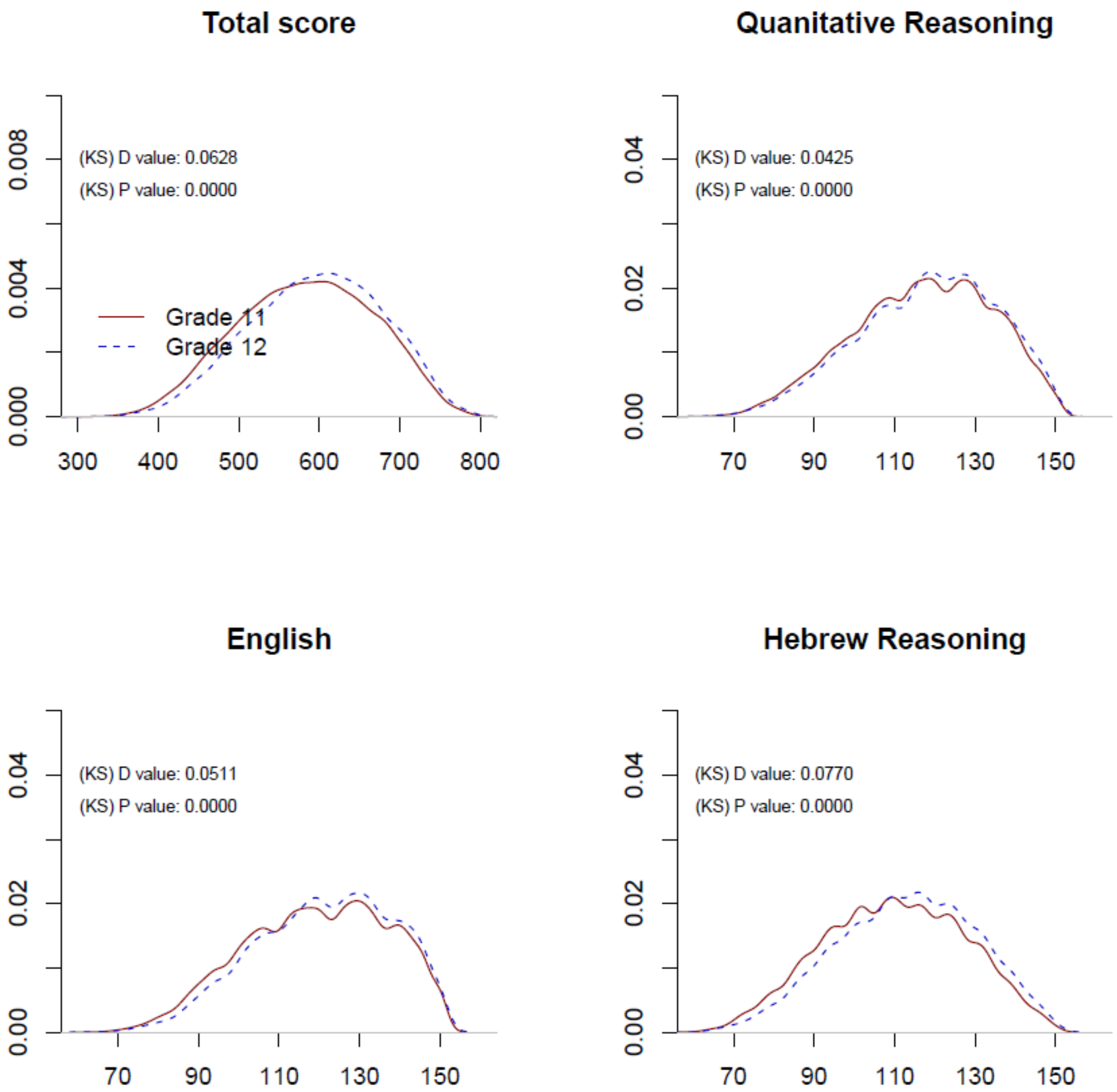
Online Appendix

Appendix Figure A1: Psychometric Scores Distributions, by Grade of Testing (2006-2014 Sample)



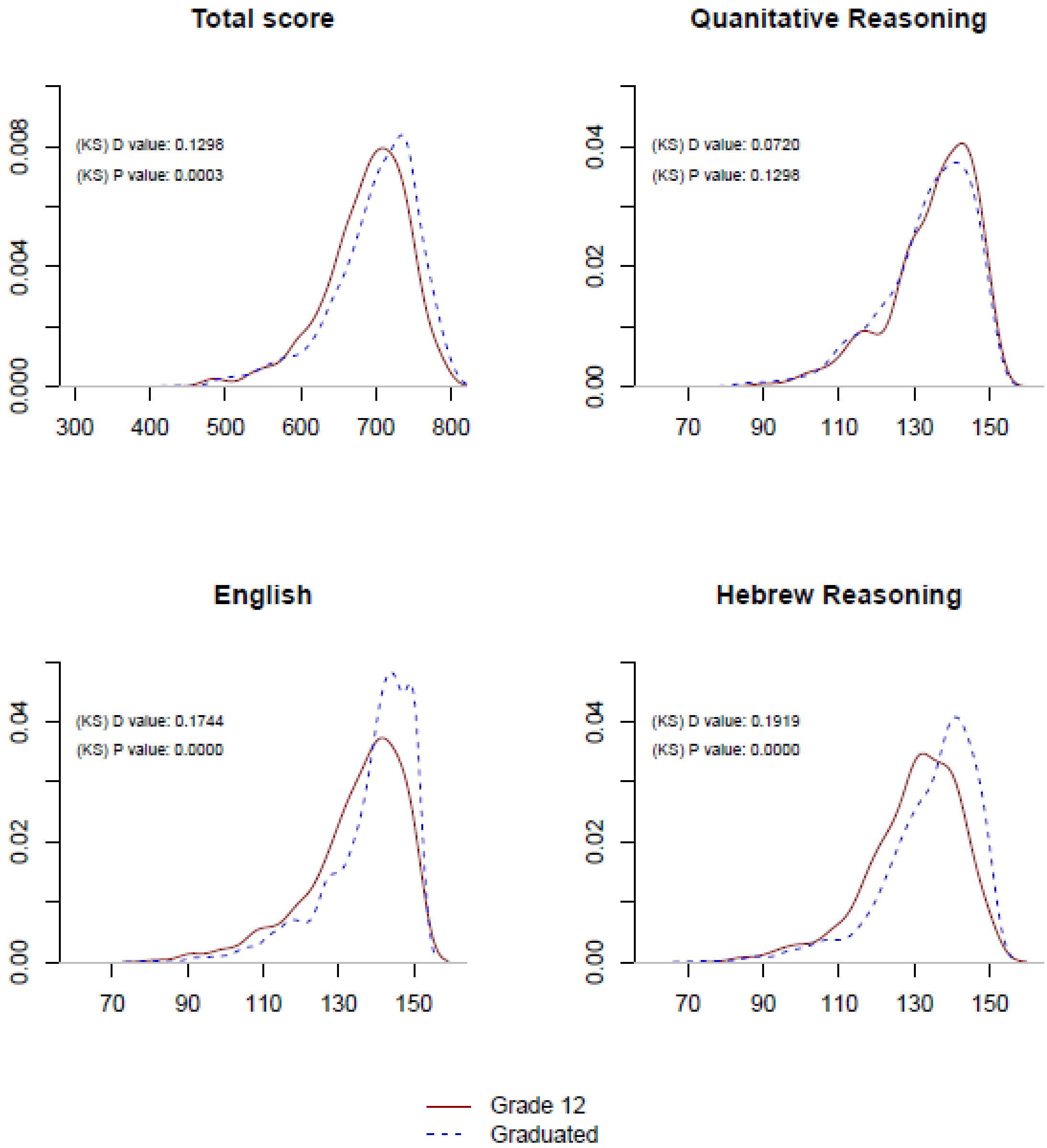
Notes: This figure plots the distribution of psychometric scores, by grade of taking the test- the red solid line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade. The 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 graduetes in 2006-2014.

Appendix Figure A2: Psychometric Scores Distributions, by Grade of Testing (1992-2005 cohorts, non-gifted)



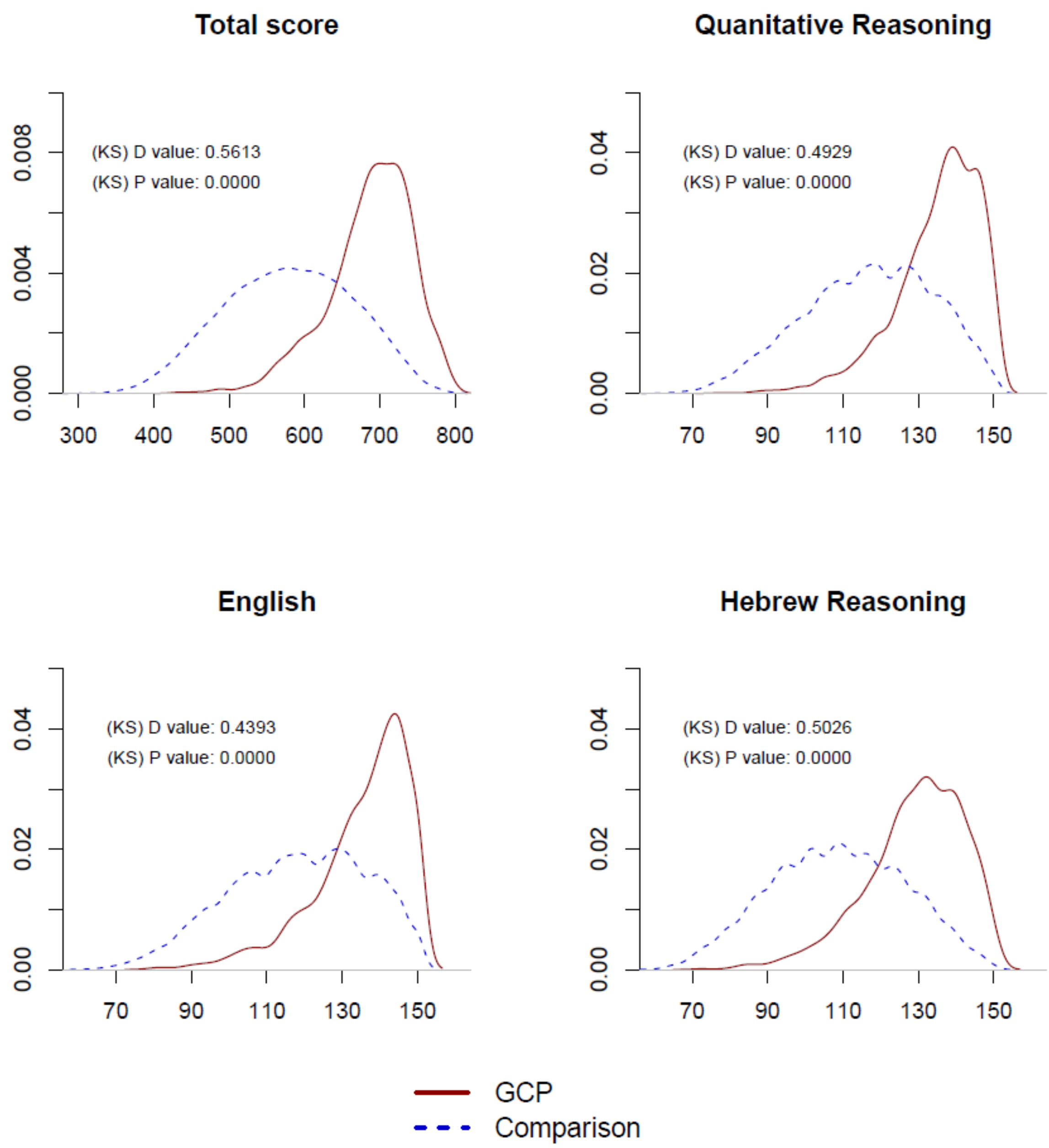
Notes: This figure plots the distribution of psychometric scores, by grade of taking the test- the red solid line represents the sample of students who took the test during their 10th or 11th grade, and the blue dashed line represents the sample of students who took the test during their 12th grade.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

Appendix Figure A3: Psychometric Scores Distributions, by Grade of Testing (1992-2005 cohorts)



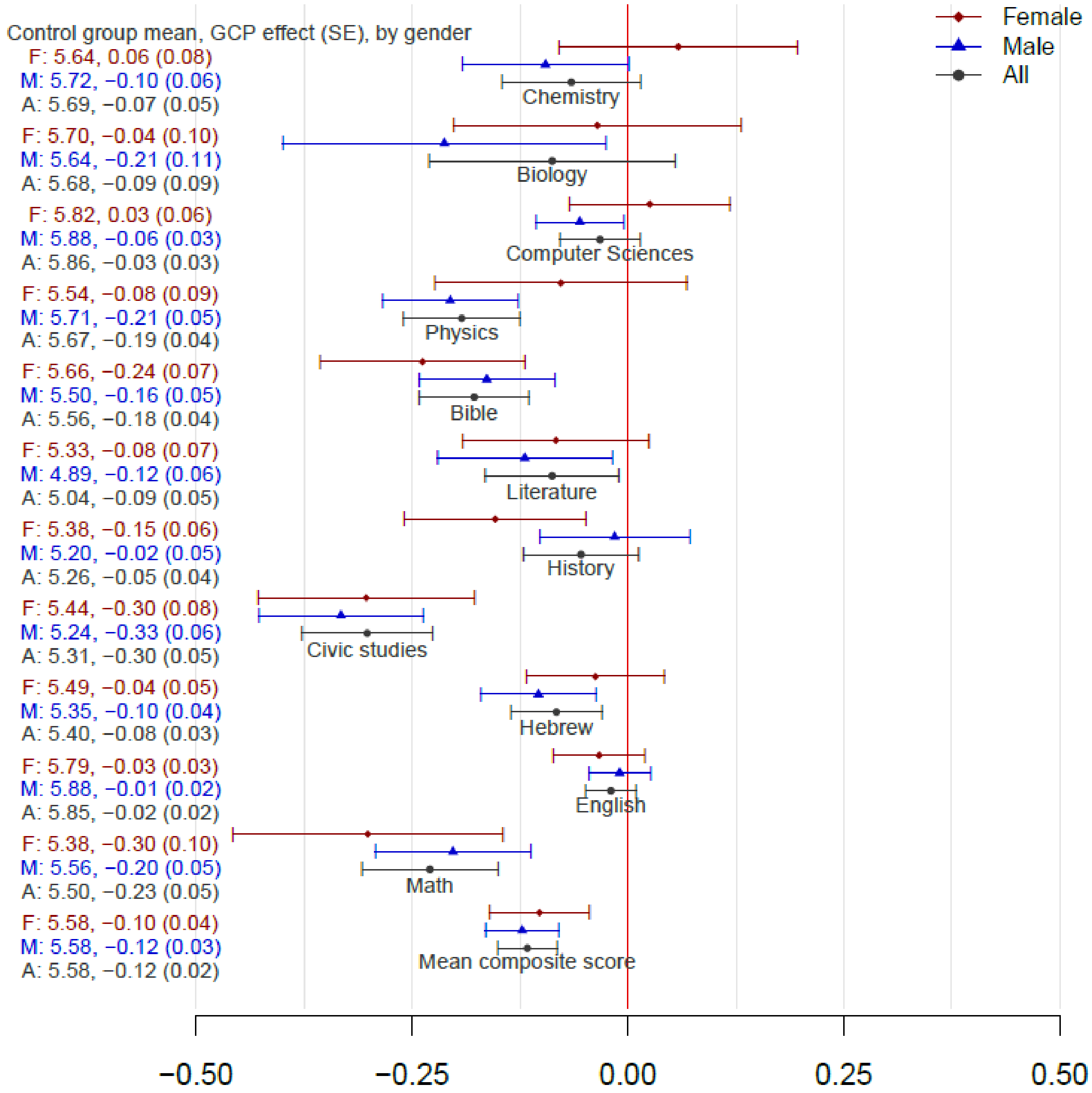
Notes: This figure plots the distribution of psychometric scores, by grade of taking the test- the red solid line represents the sample of students who took the test during their 12th grade, and the blue dashed line represents the sample of students who took the test during after their 12th grade (during 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 from the cohorts of high-school graduetes in 1992-2005.

Appendix Figure A4: Psychometric Scores Distributions, GCP and Comparison Group Pull (Before Matching)



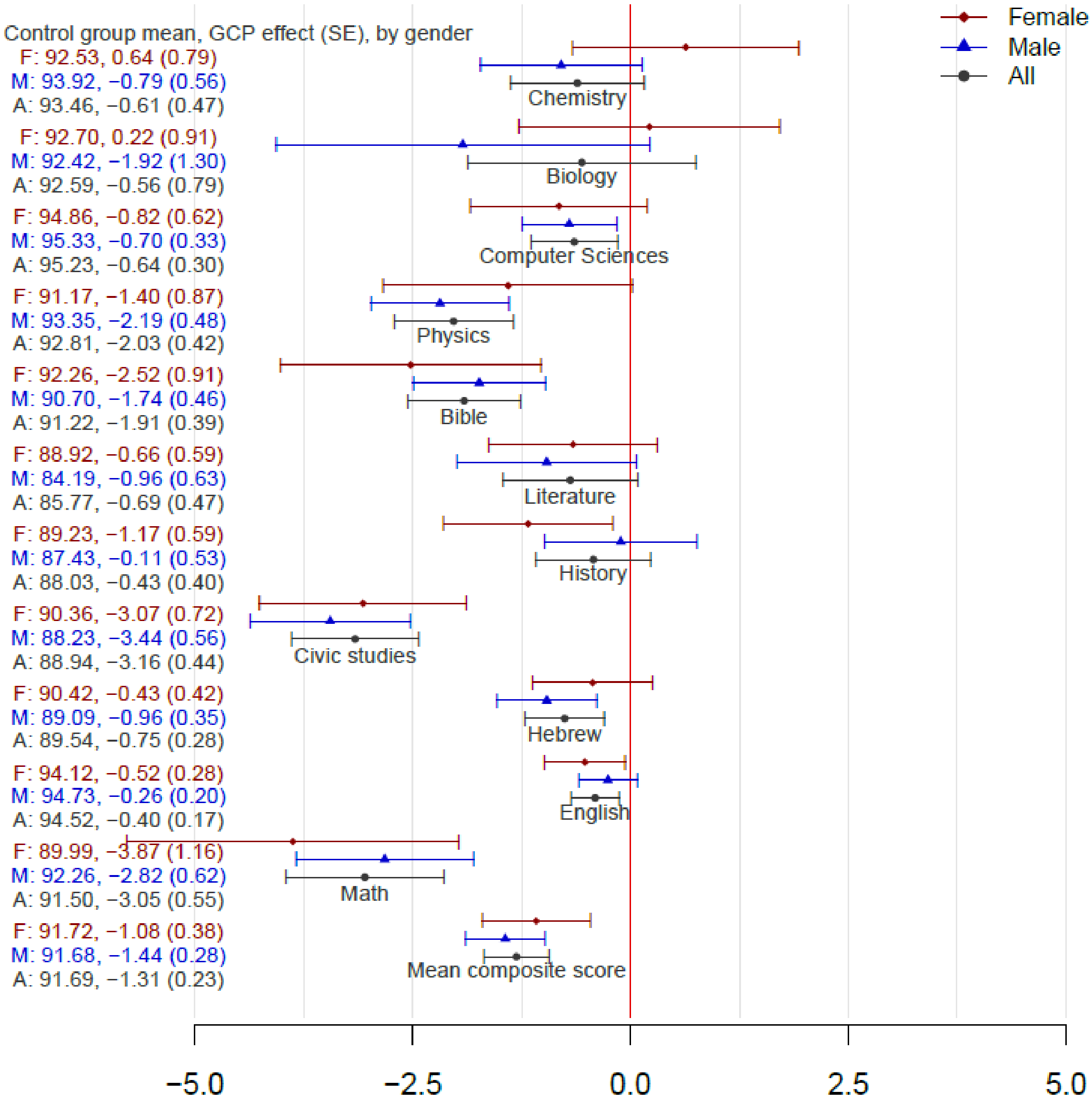
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group pull (before matching). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions.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

Appendix Figure A5a: GCP Effects on Bagrut Test Scores, Compulsory and Elective Subjects



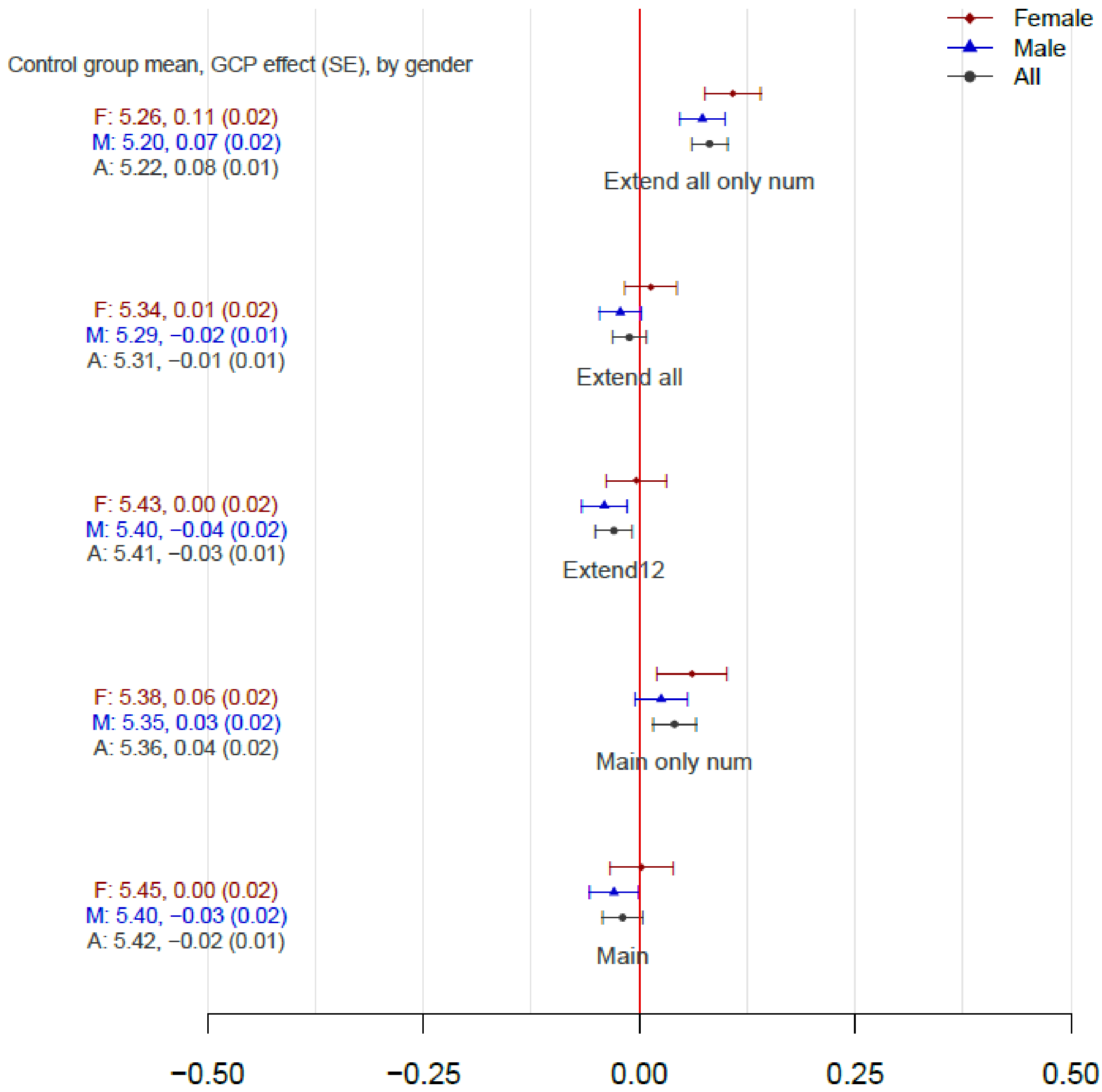
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory and most chosen elective subjects). These scores are normalised on a scale of 0-6. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The sample includes students from the cohorts of high-school graduetes in 2006-2014, who took the psychometric test during their 10-11th grade.

Appendix Figure A5b: GCP Effects on Bagrut Test Scores, Compulsory and Elective Subjects



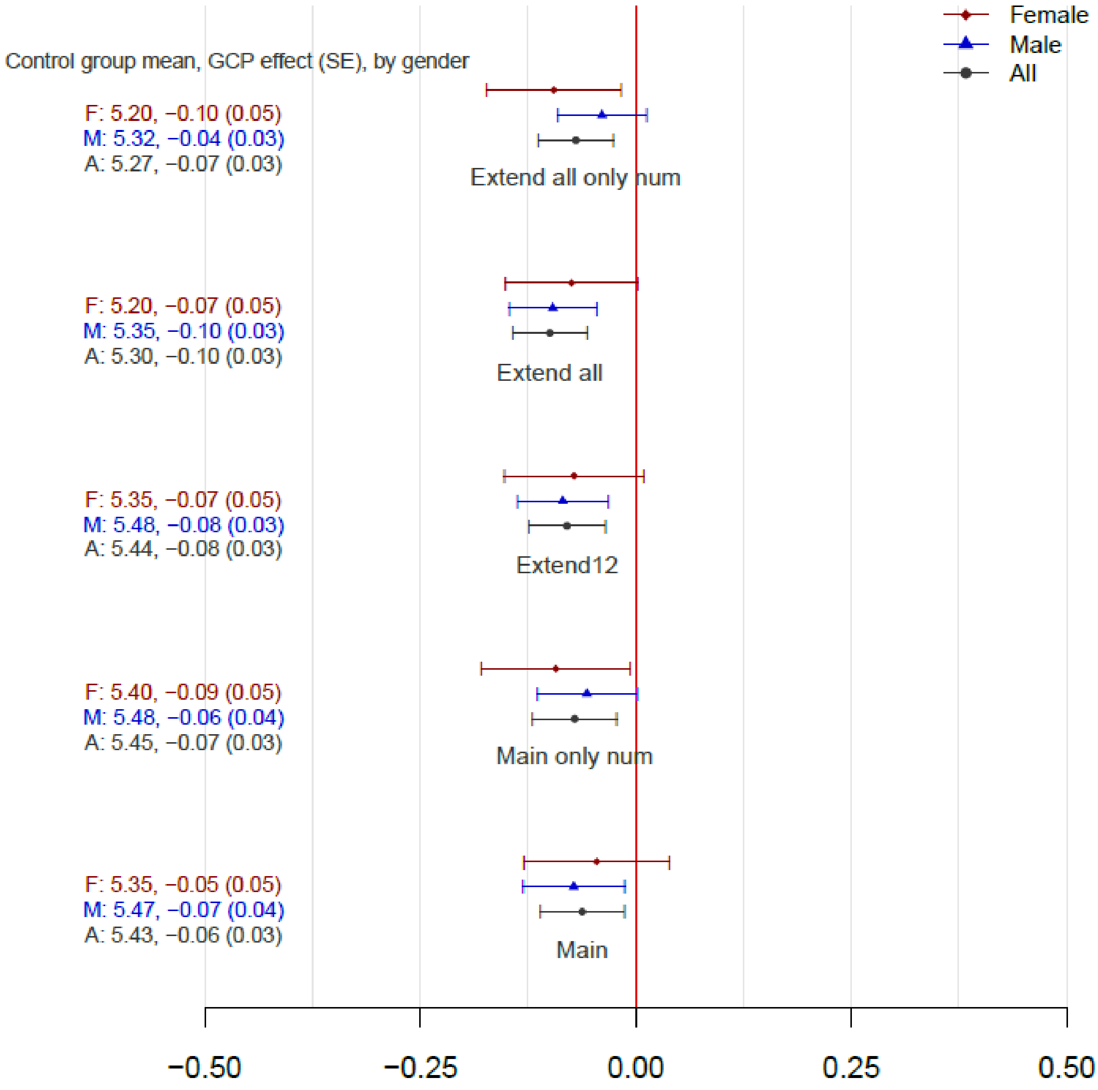
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on Bagrut test scores (compulsory and most chosen elective subjects). These scores are the full test scores, on a scale of 1-100. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The sample includes students from the cohorts of high-school graduetes in 2006-2014, who took the psychometric test during their 10-11th grade.

<u>Appendix Figure A6</u>: GCP Effects on Bagrut Mean Compsite Score, By Sample



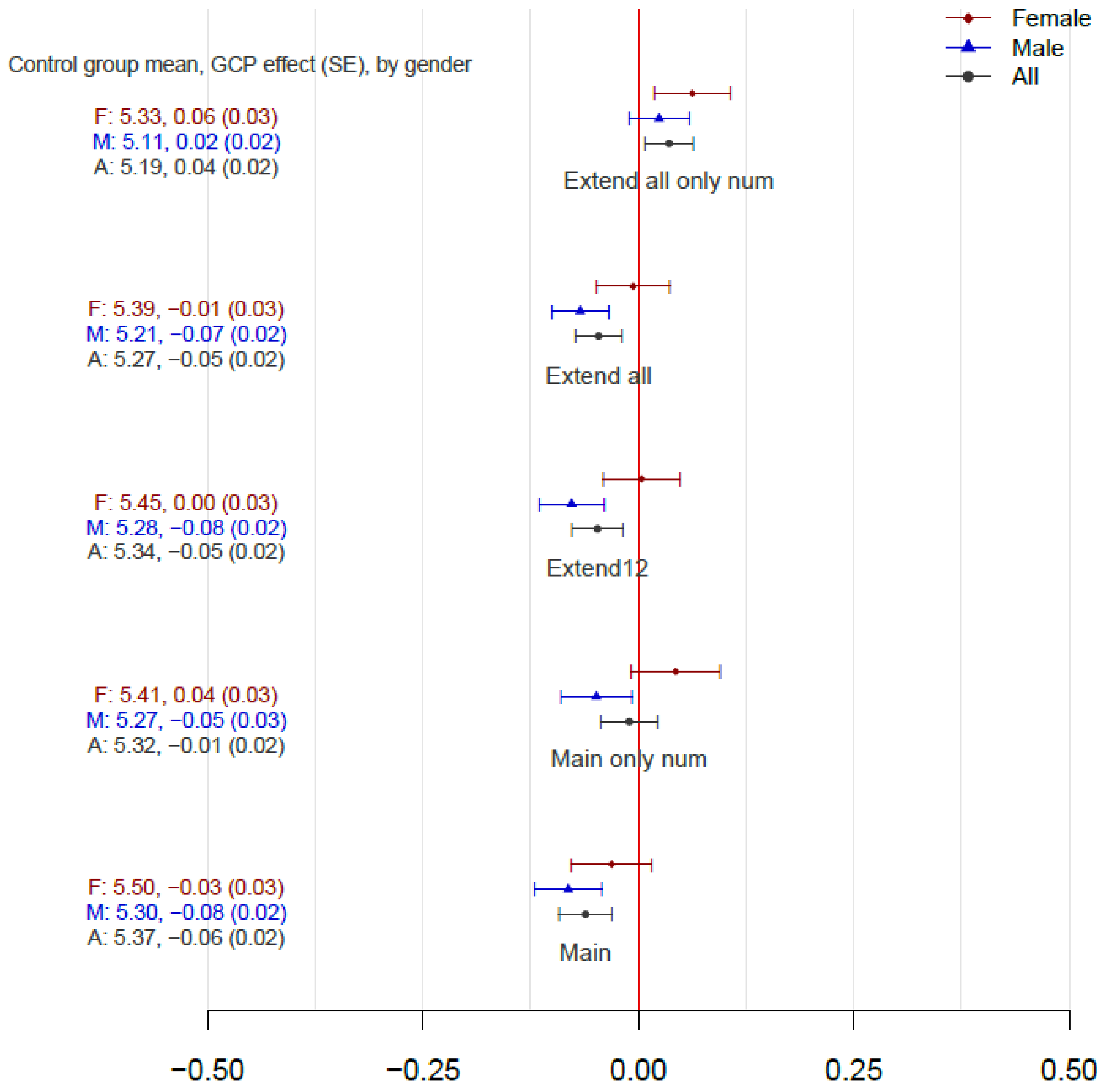
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut mean composite score, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A7: GCP Effects on Bagrut Math Score, By Sample



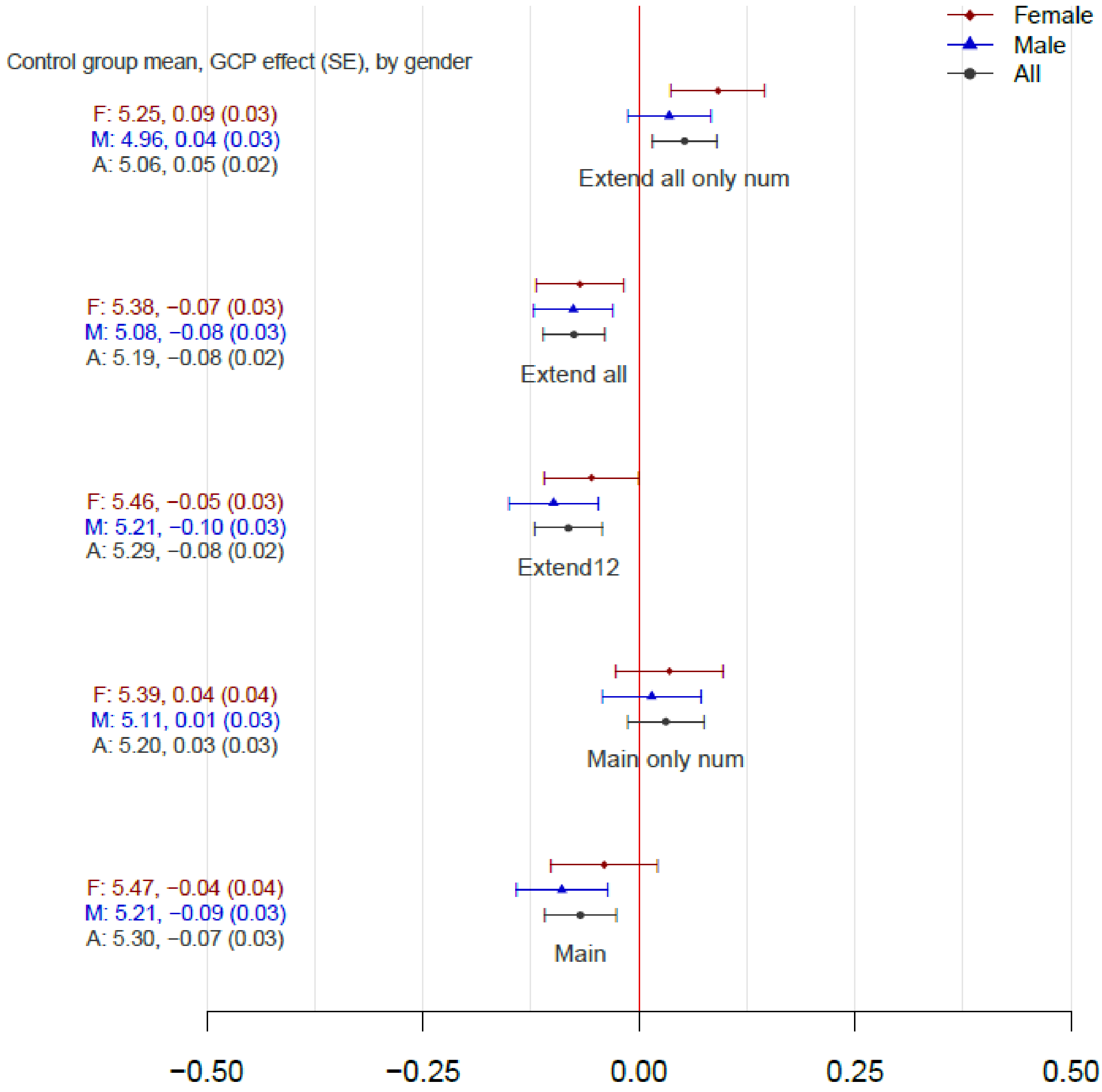
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut math score, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A8: GCP Effects on Bagrut Hebrew Score, By Sample



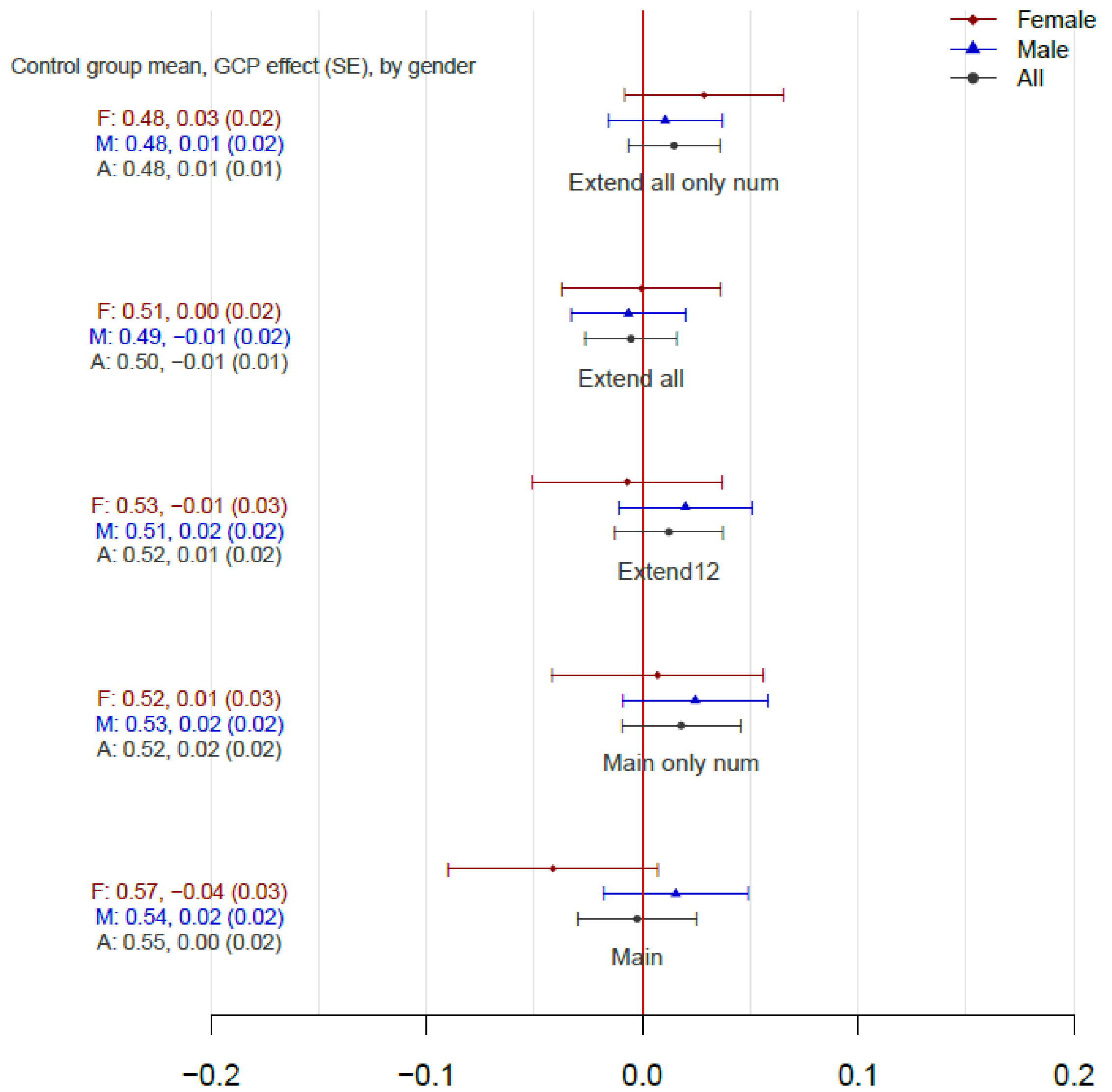
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut hebrew score, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

<u>Appendix Figure A9</u>: GCP Effects on Bagrut Bible Score, By Sample



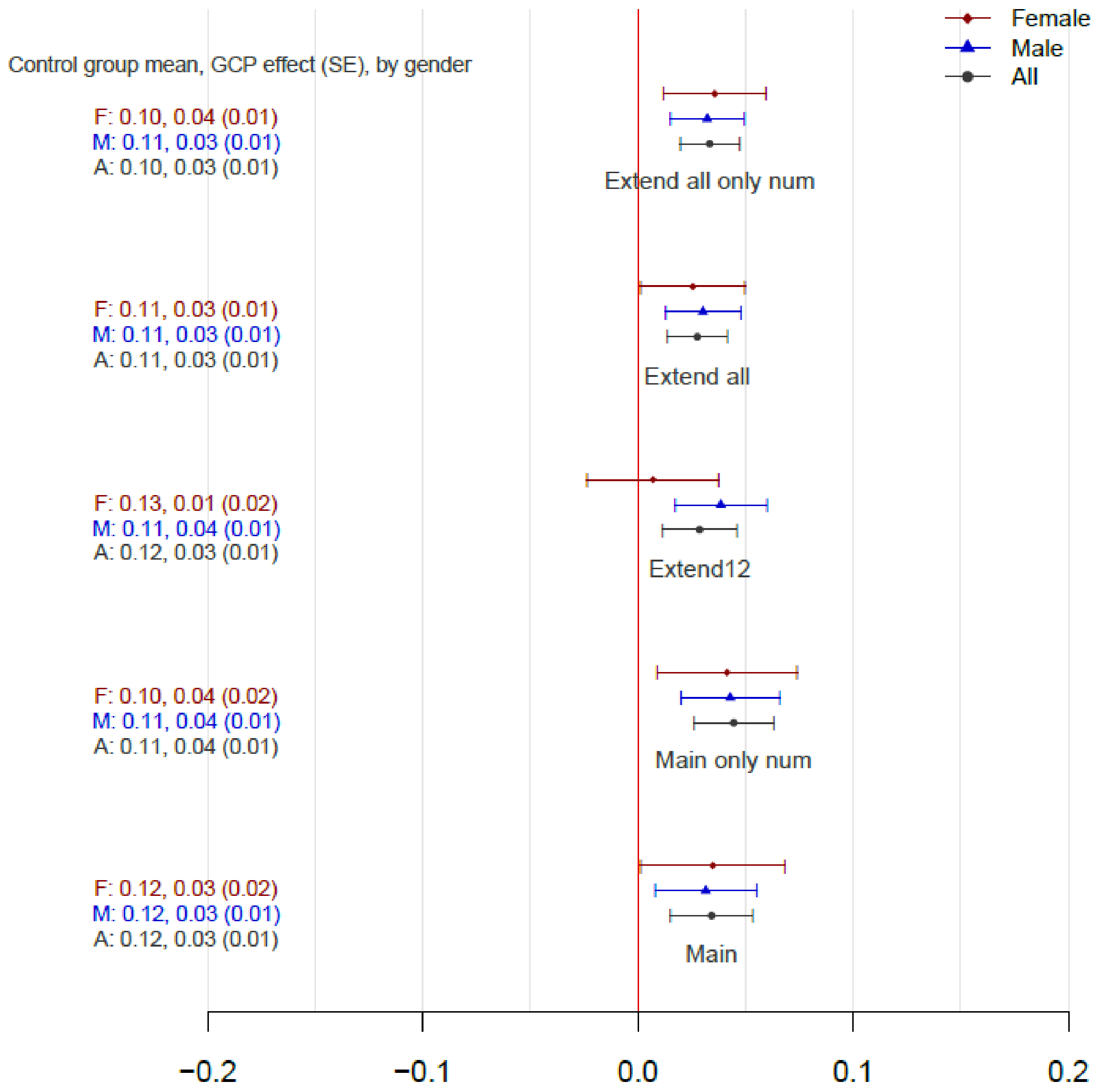
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut bible score, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

<u>Appendix Figure A10</u>: GCP Effects on MA Degree Attainment, By Sample



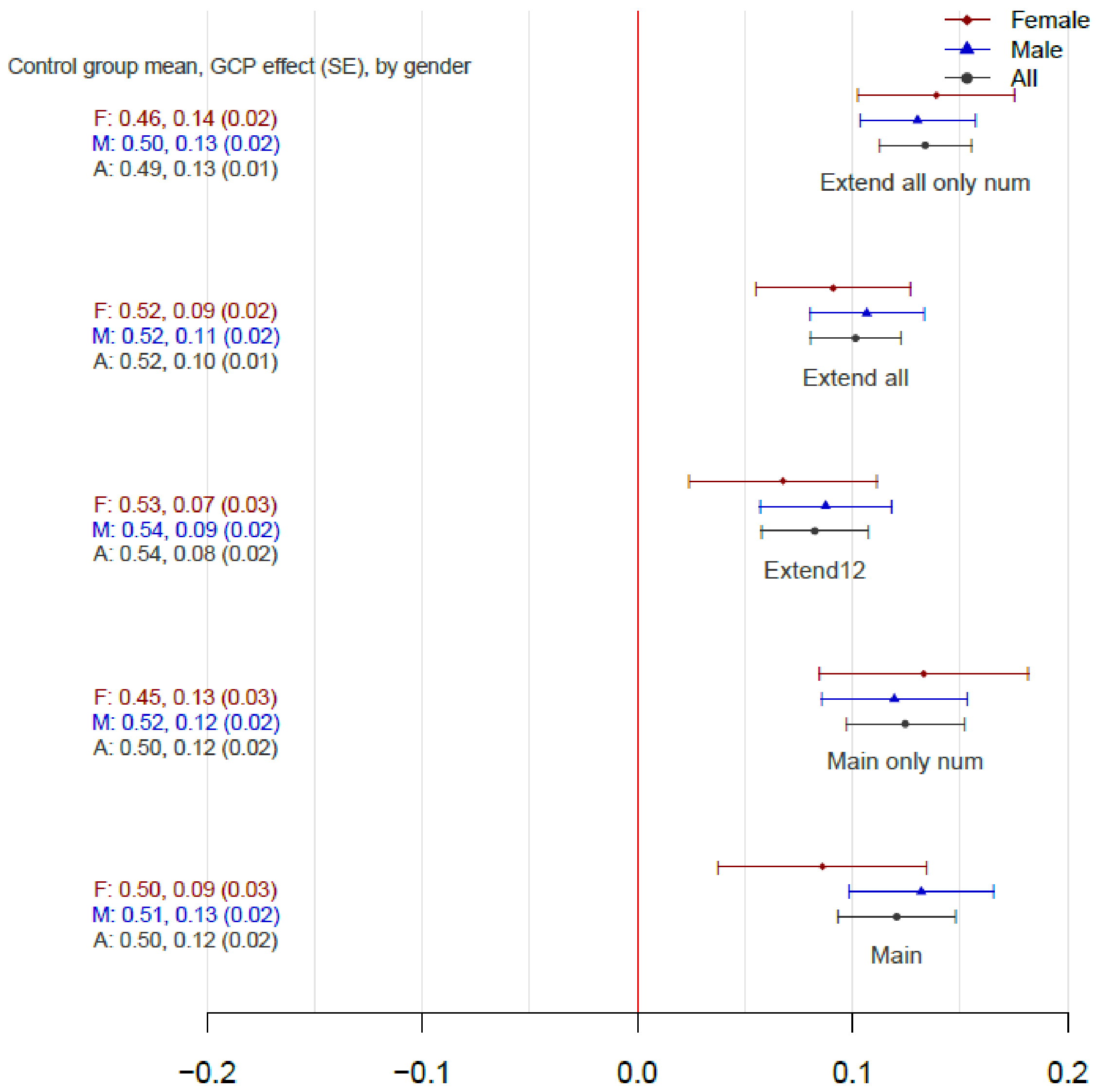
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on MA degree attainment, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A11: GCP Effects on PHD Degree Attainment, By Sample



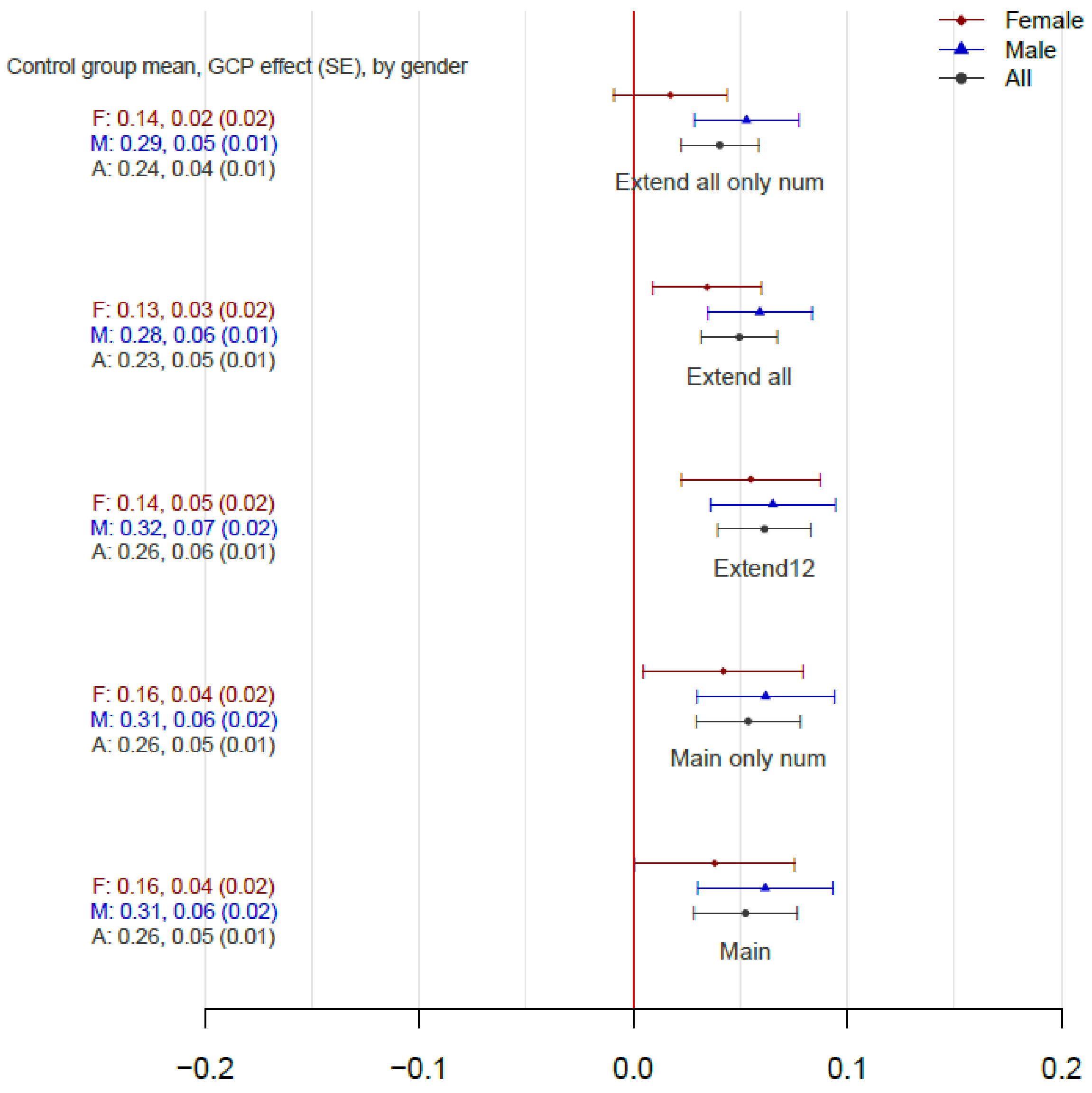
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on PHD degree attainment, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

<u>Appendix Figure A12</u>: GCP Effects on Double Major BA, By Sample



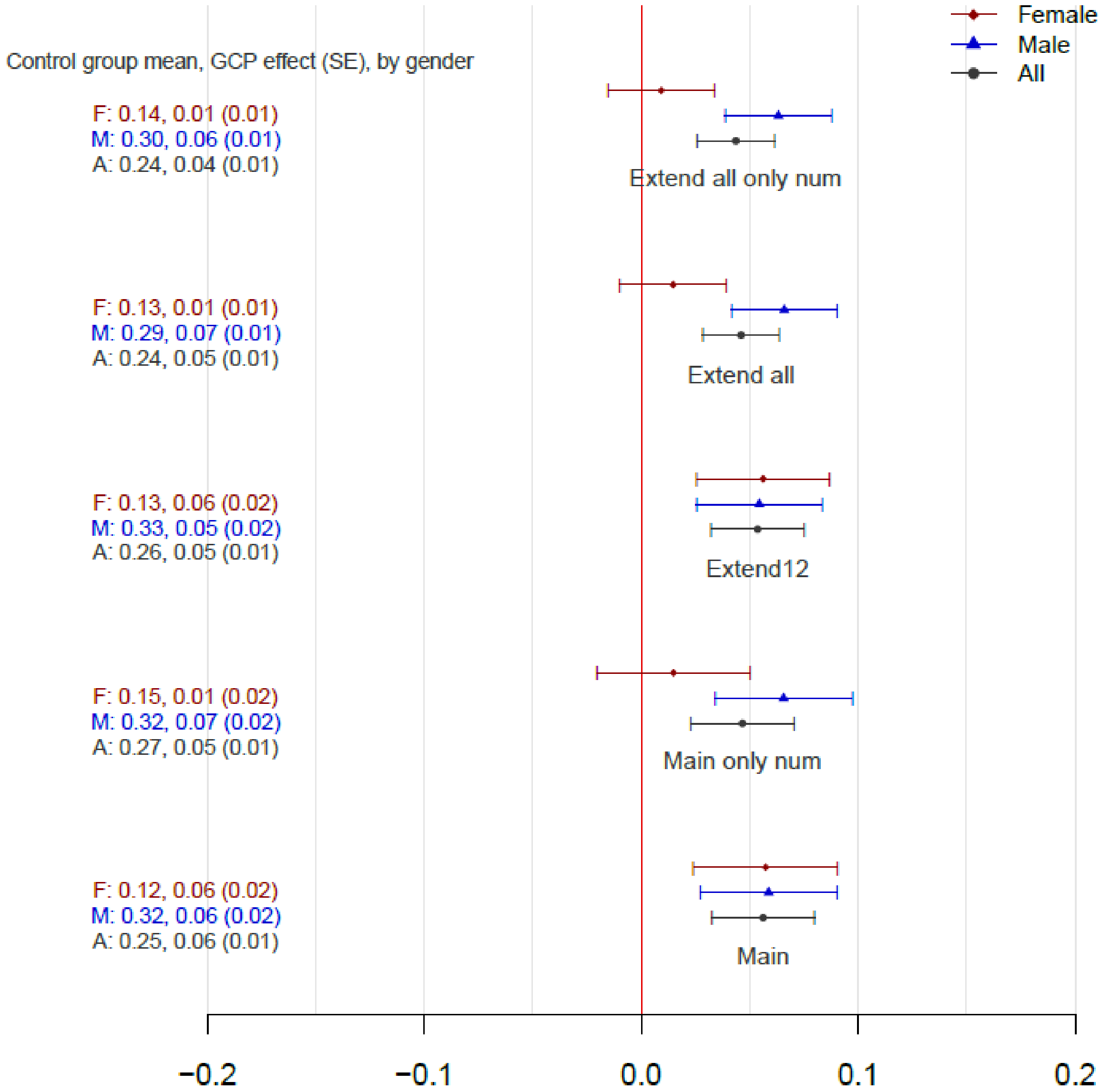
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on double major BA, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A13: GCP Effects on STEM Double Major BA, By Sample



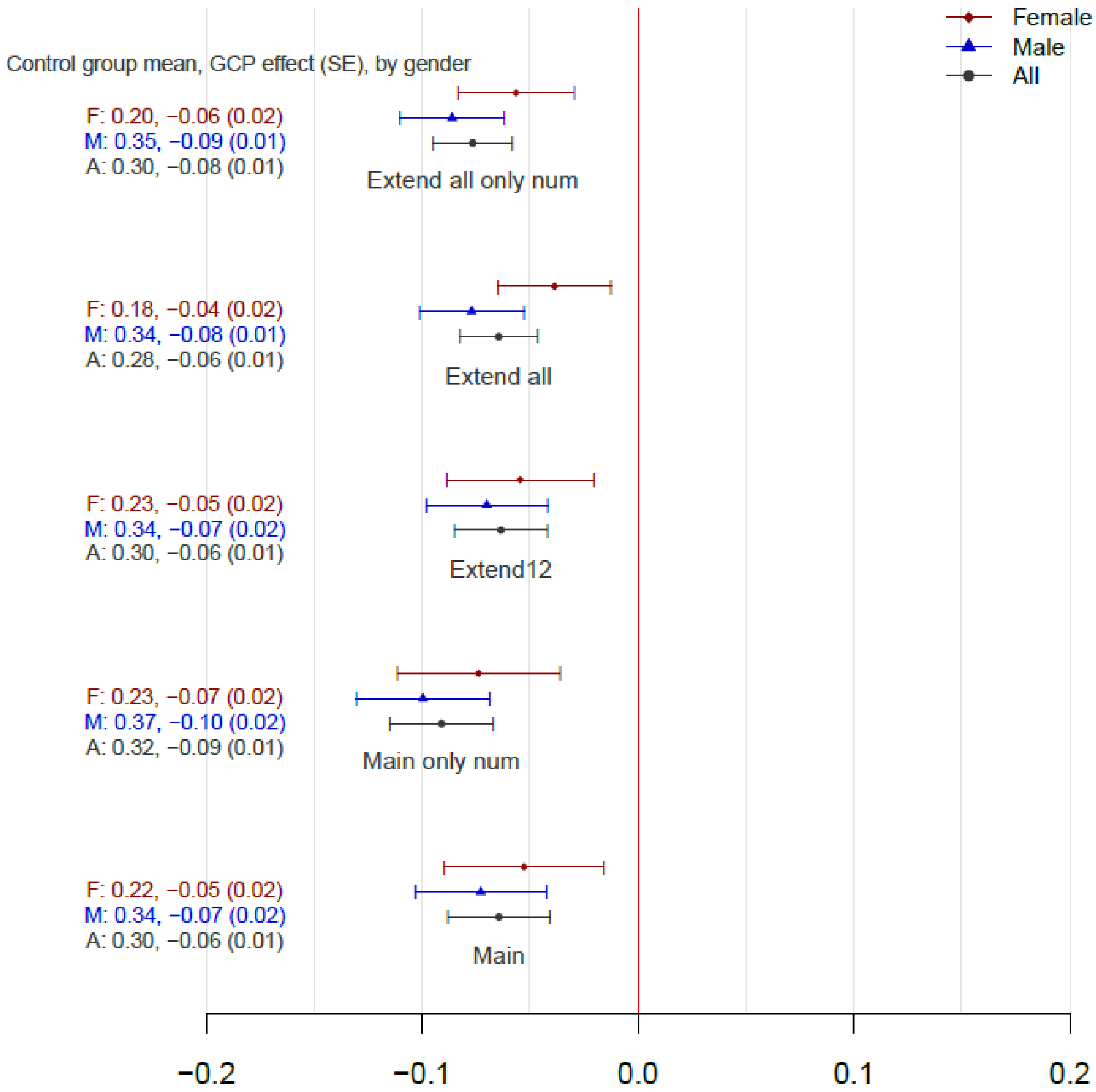
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on STEM double major BA, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

<u>Appendix Figure A14</u>: GCP Effects on BA in Math, Statistics, and Computer Sciences, By Sample



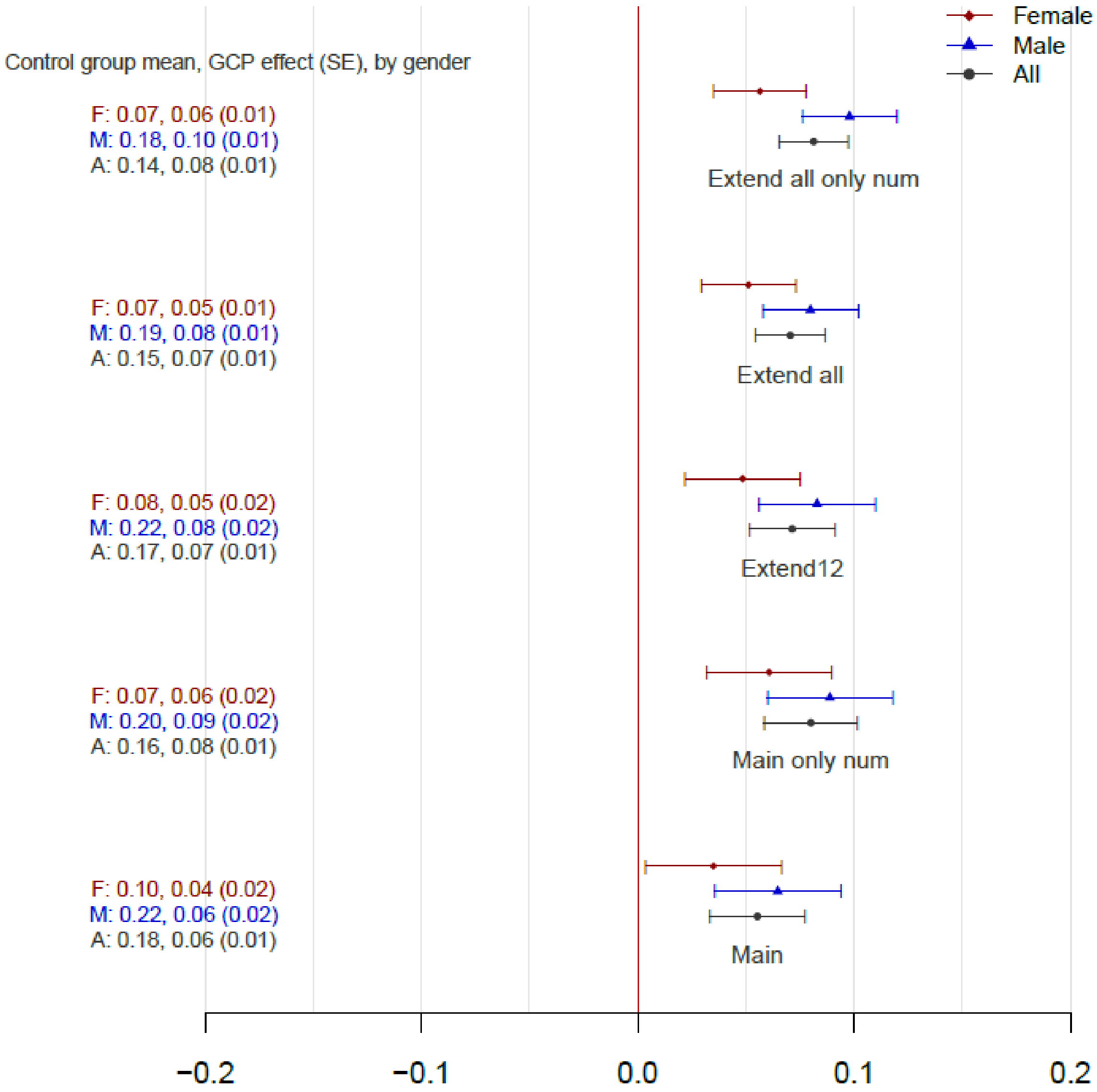
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Math, Statistics, and Computer Sciences, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A15: GCP Effects on BA in Engineering, By Sample



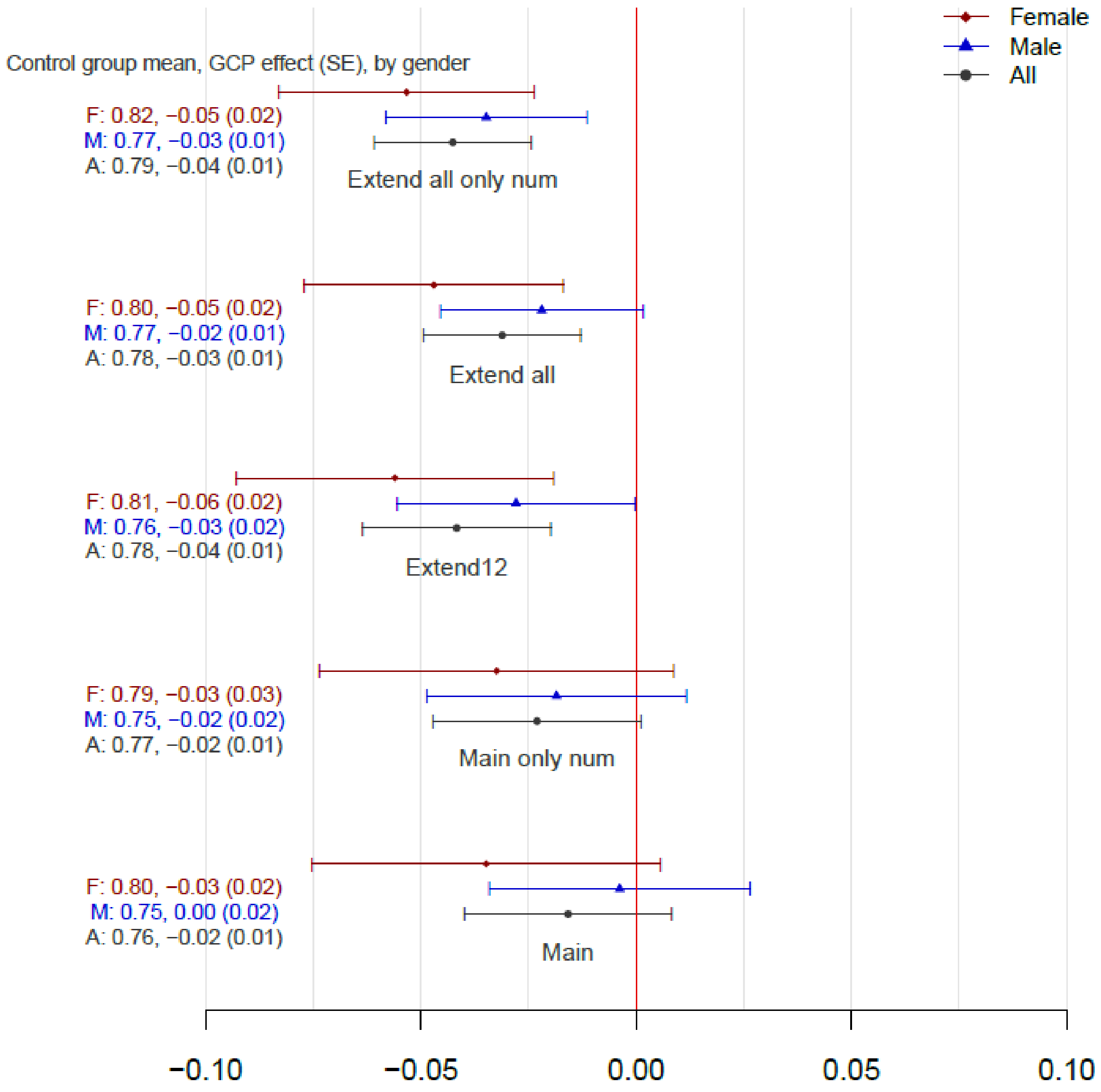
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Engineering, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A16: GCP Effects on BA in Physical Sciences, By Sample



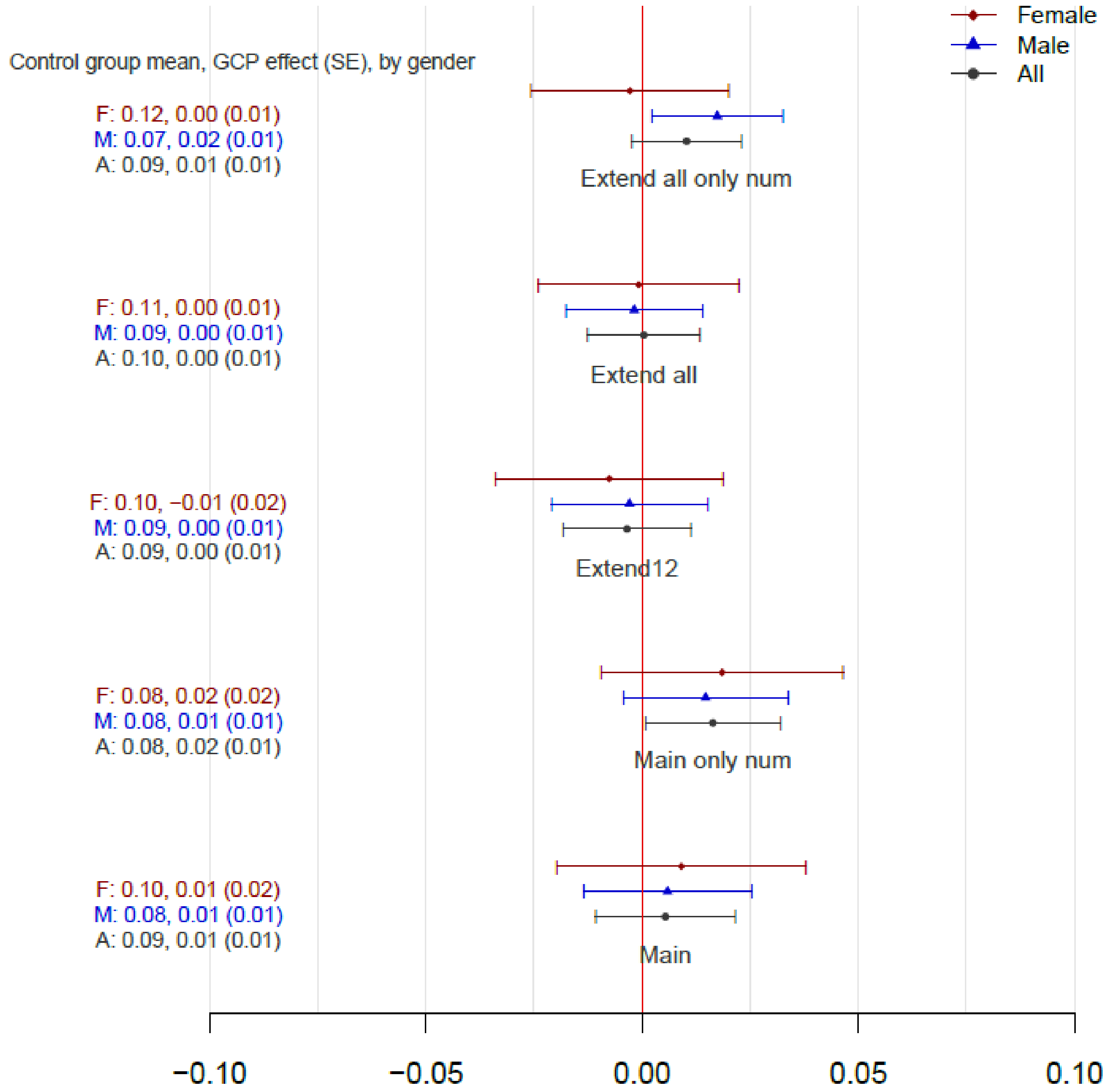
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in Physical Sciences, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A17: GCP Effects on Employement in 2018, By Sample



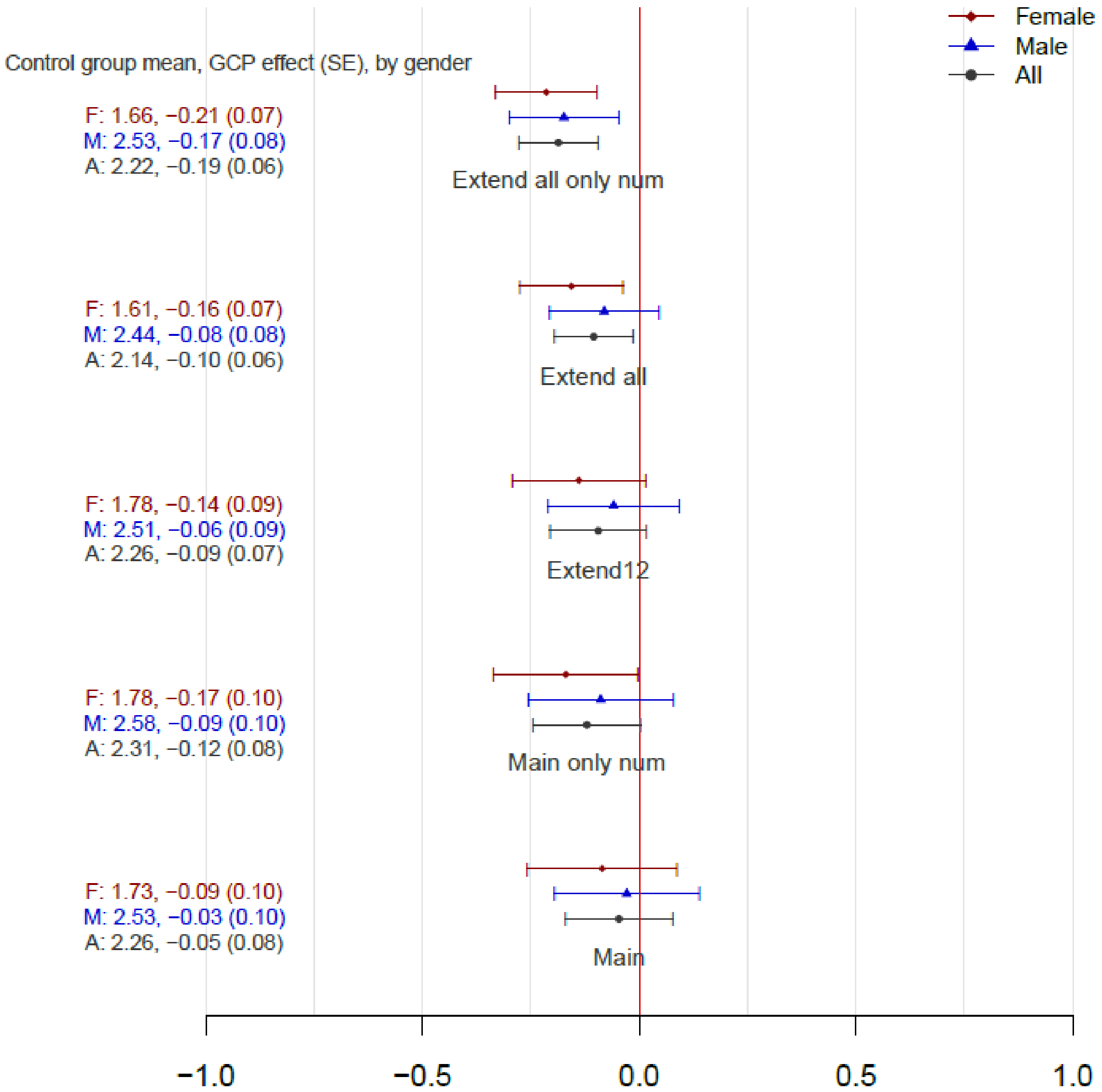
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employement in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A18: GCP Effects on Self Employement in 2018, By Sample



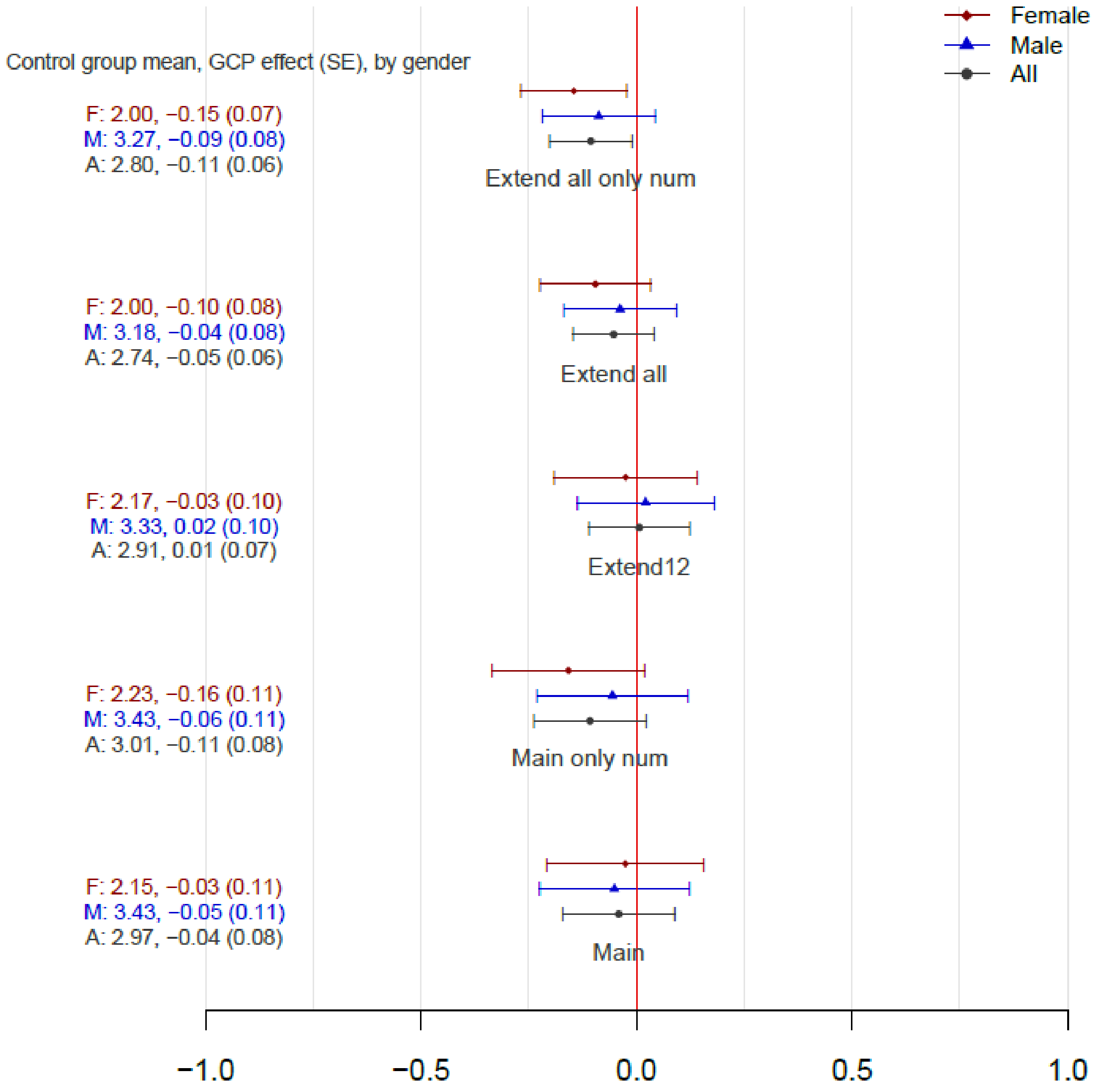
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on self employement in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A19: GCP Effects on Salaried Worker Income in 2018, By Sample



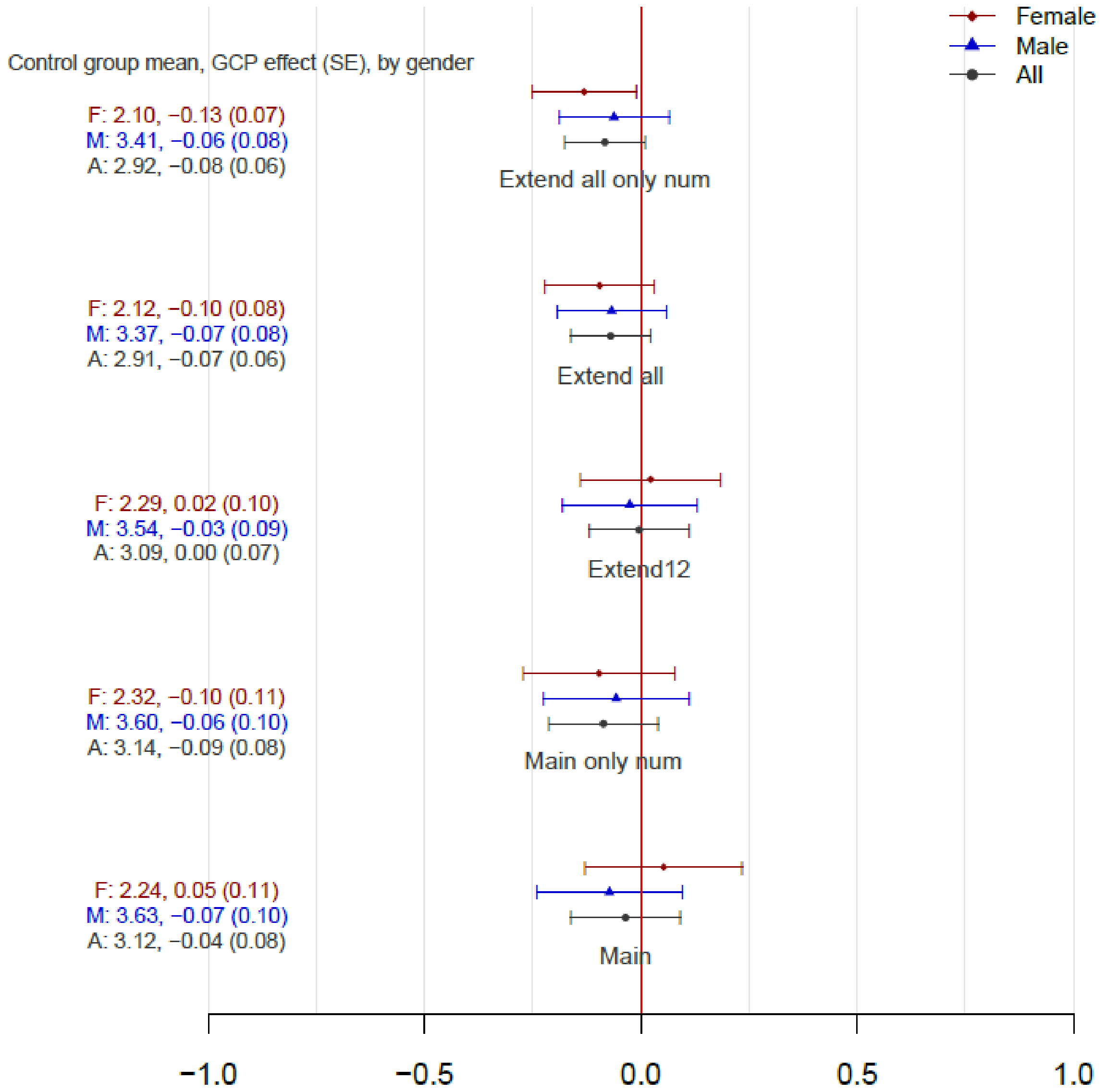
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Salaried Worker Income in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A20: GCP Effects on Salaried Worker Income (employed=1) in 2018, By Sample



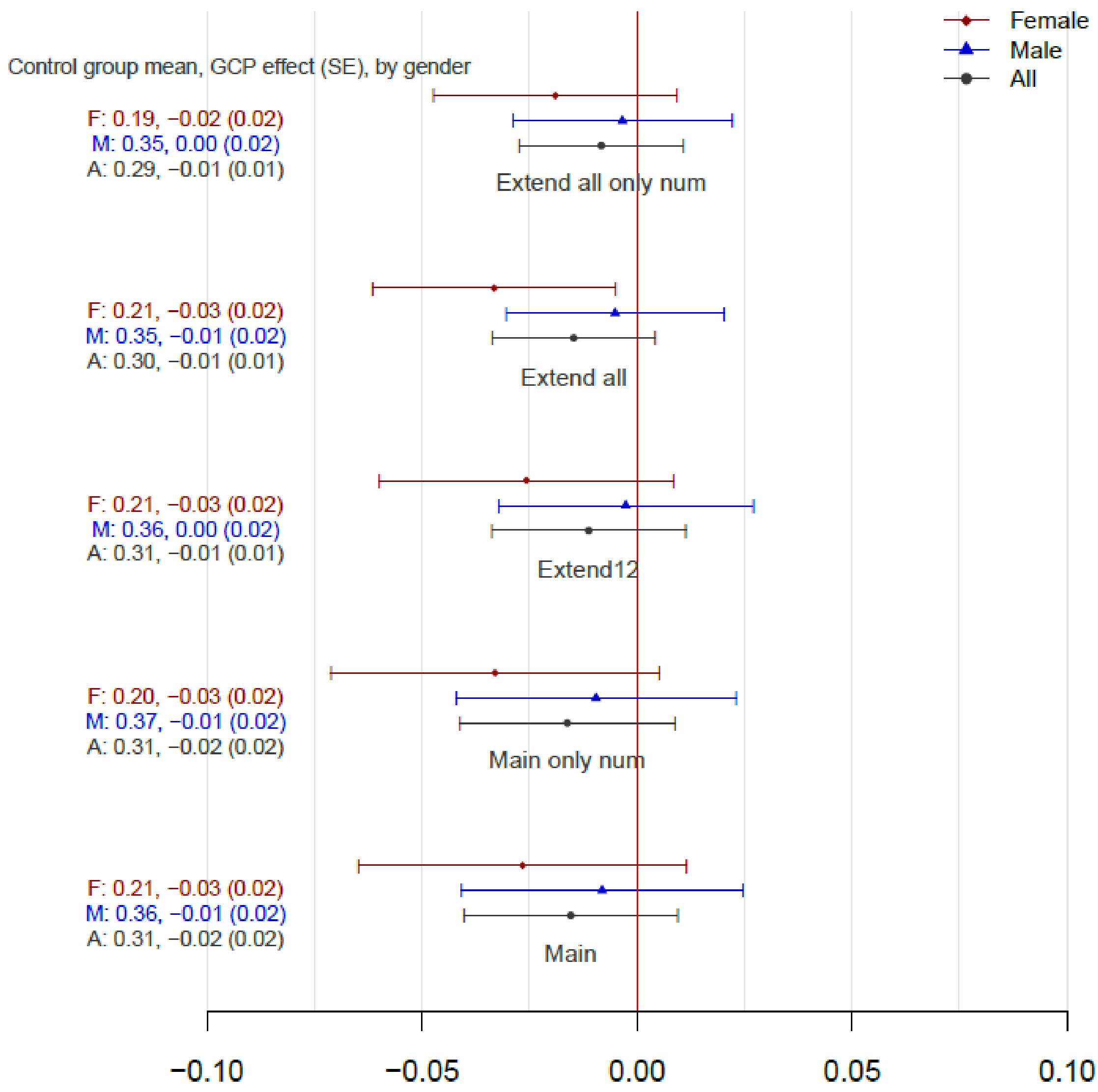
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Salaried Worker Income (employed=1) in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A21: GCP Effects on Income (employed=1) in 2018, By Sample



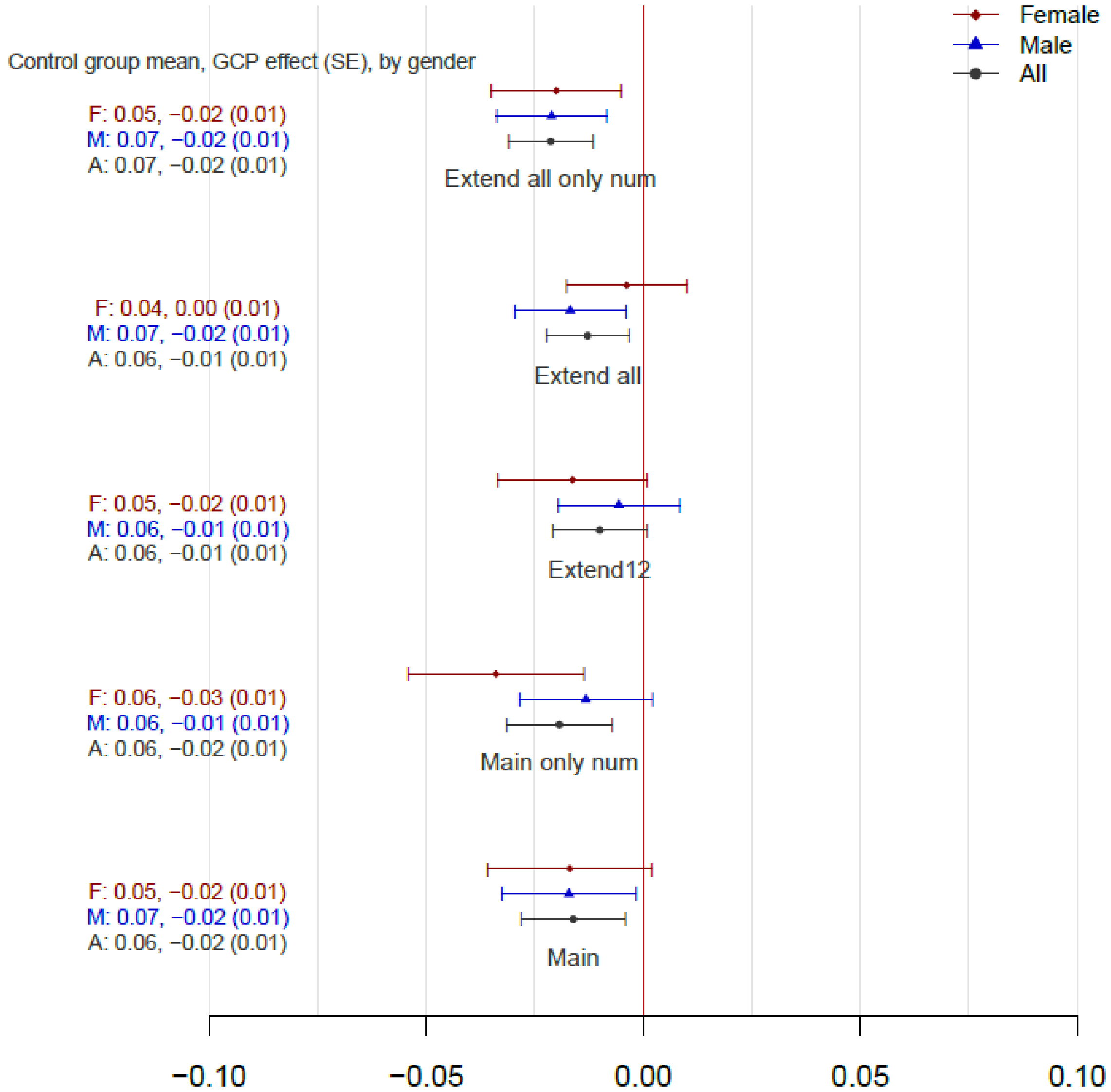
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Annual Income (employed=1) in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A22: GCP Effects on Employement in High-Tect Services Sector in 2018, By Sample



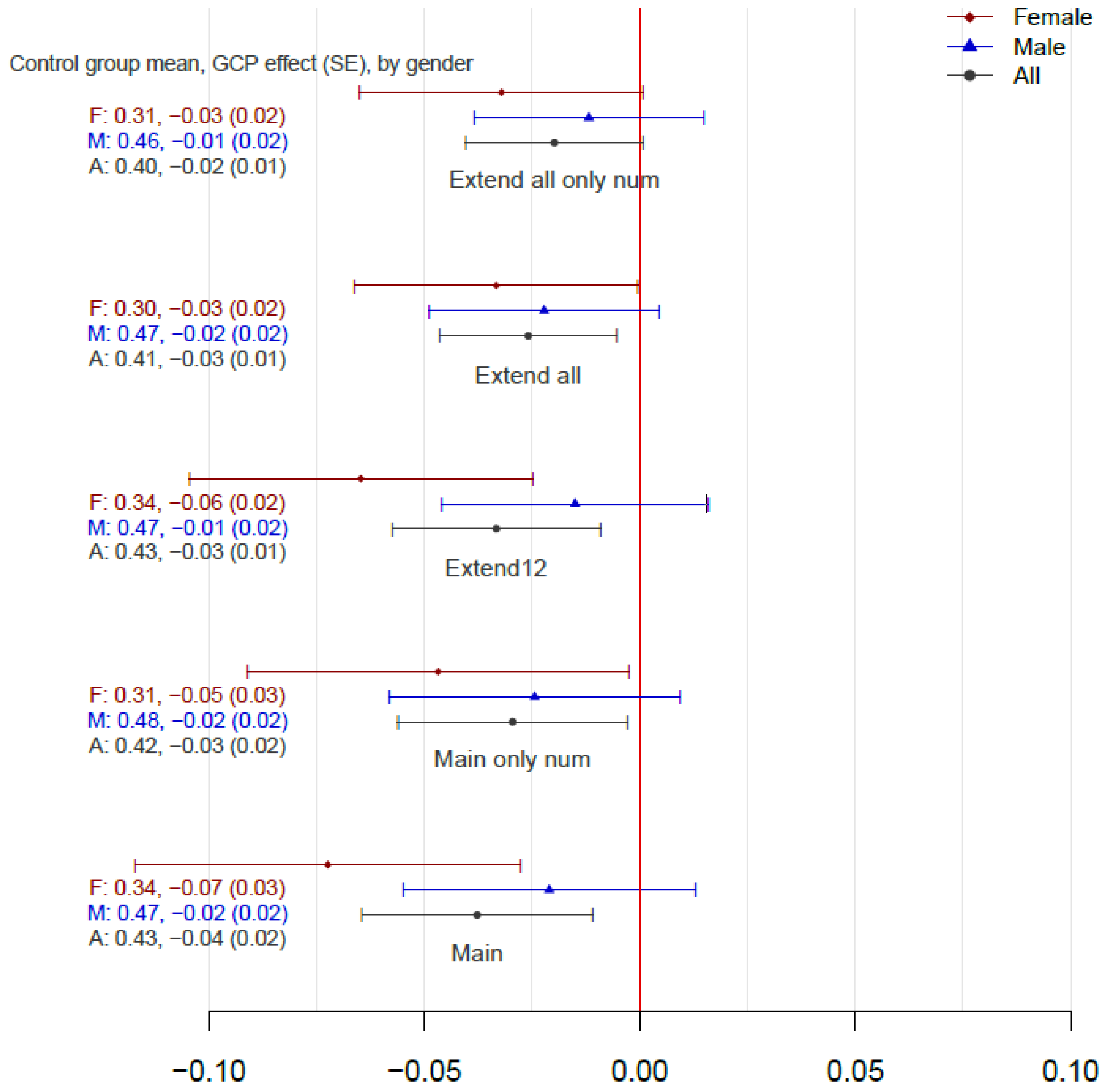
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employement in the High-Tech services sector in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A23: GCP Effects on Employement in High-Tect Manufcturing Sector in 2018, By Sample



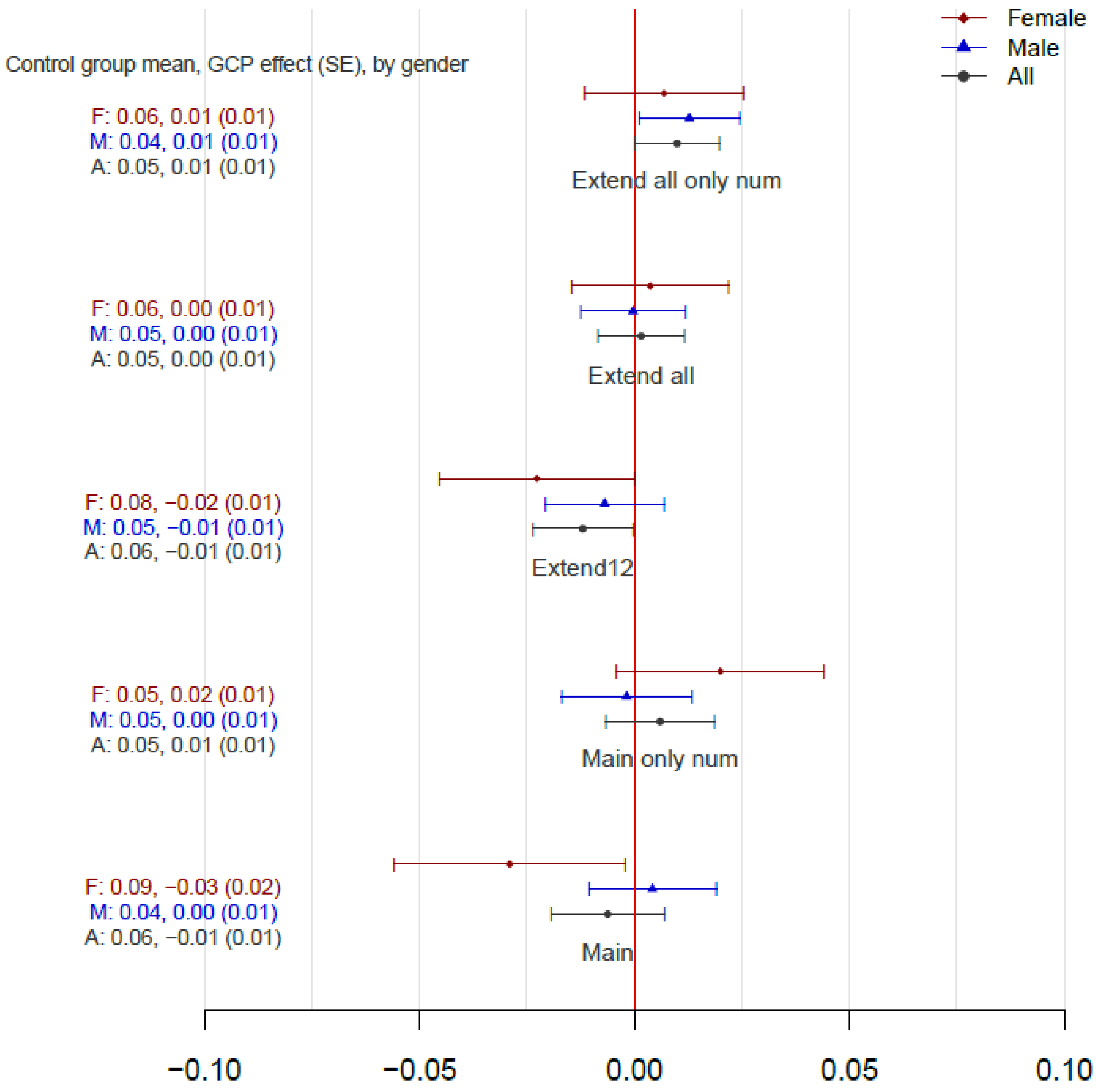
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employement in the High-Tech mnufacturing sector in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A24: GCP Effects on Employement in the Knowledge Economy in 2018, By Sample



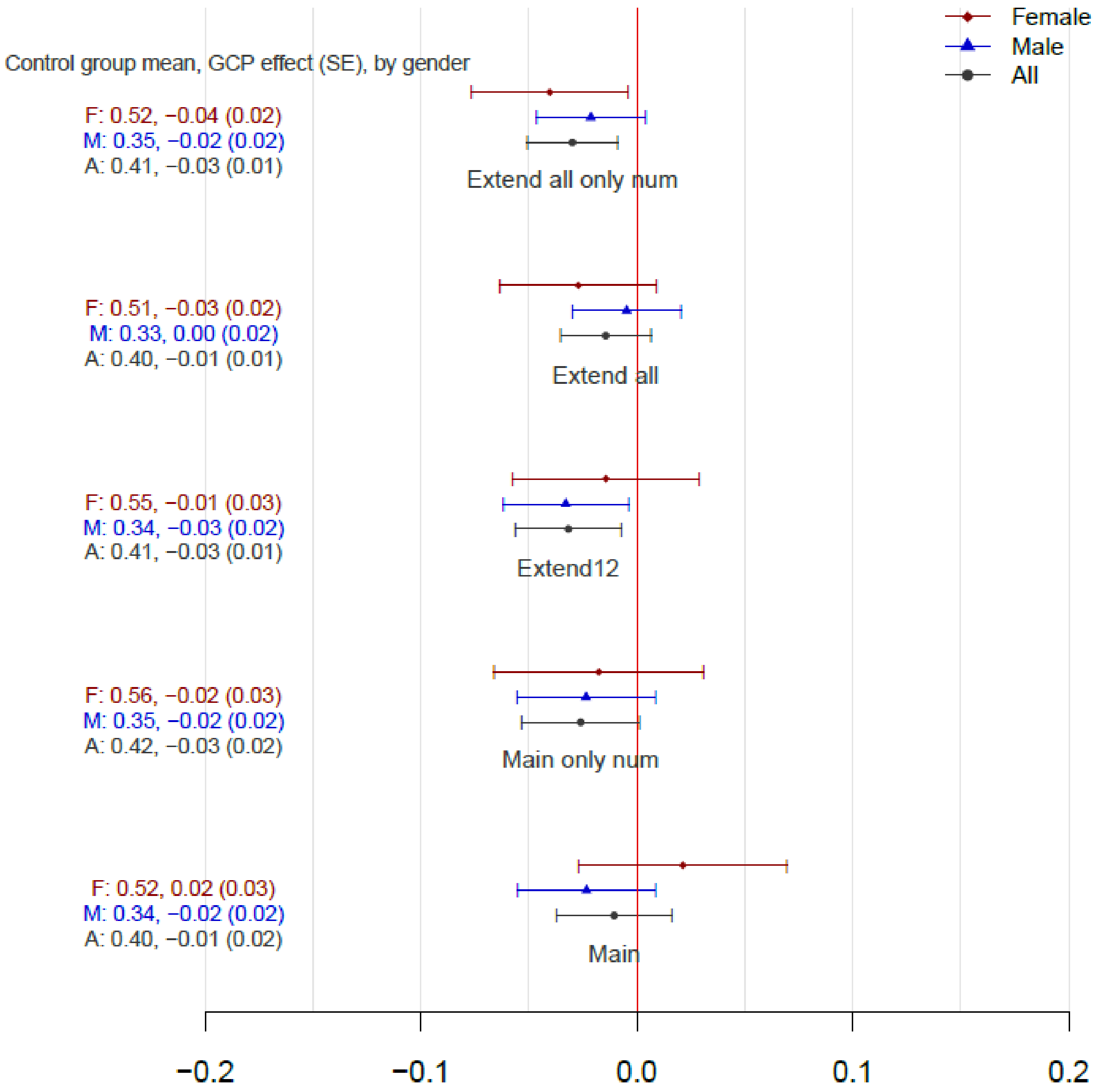
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employement in the knowledge economy in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A25: GCP Effects on Employement in the Academic Sector in 2018, By Sample



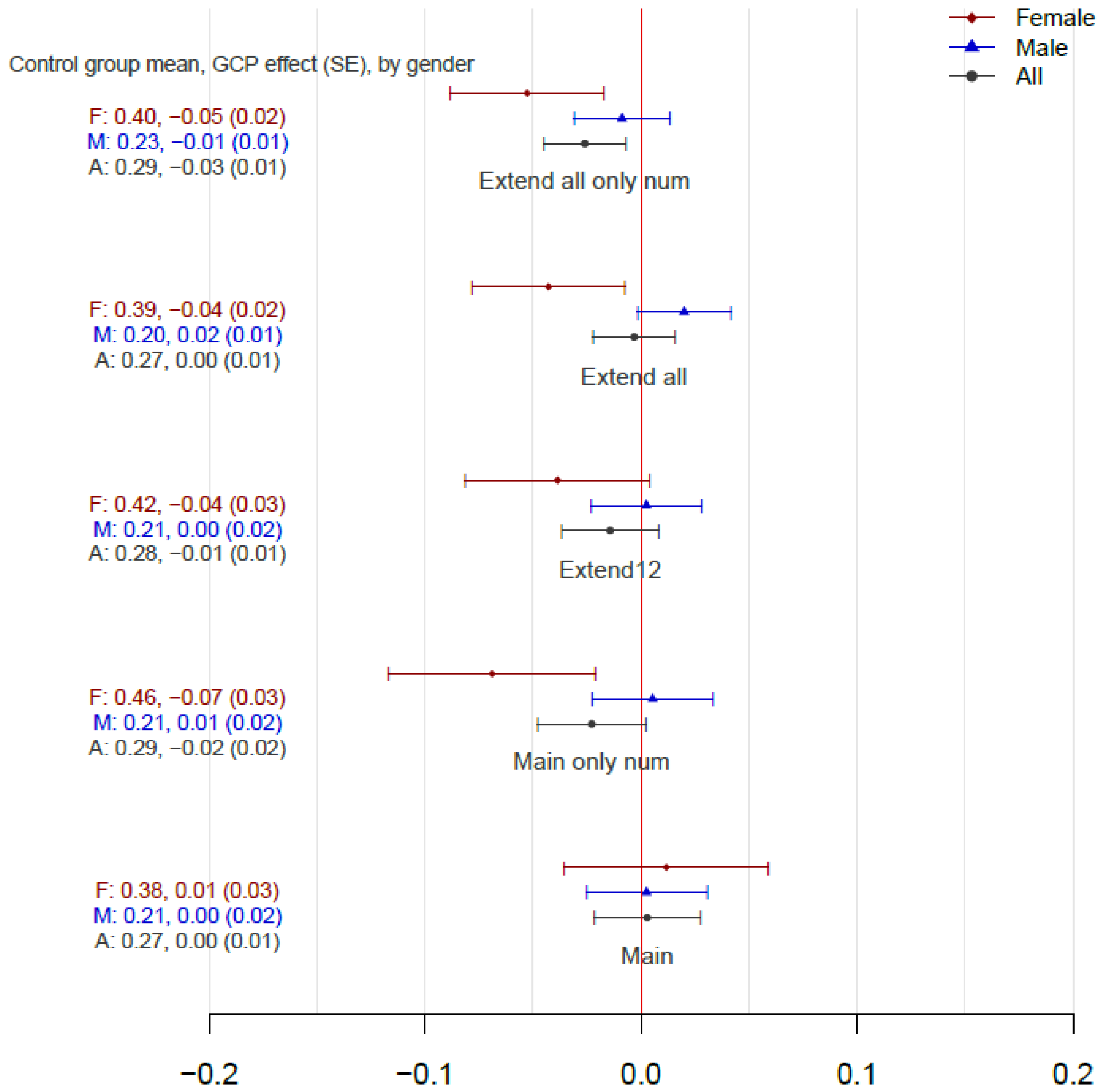
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on employement in the academic sector in 2018, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A26: GCP Effects on Marriage Before 30, By Sample



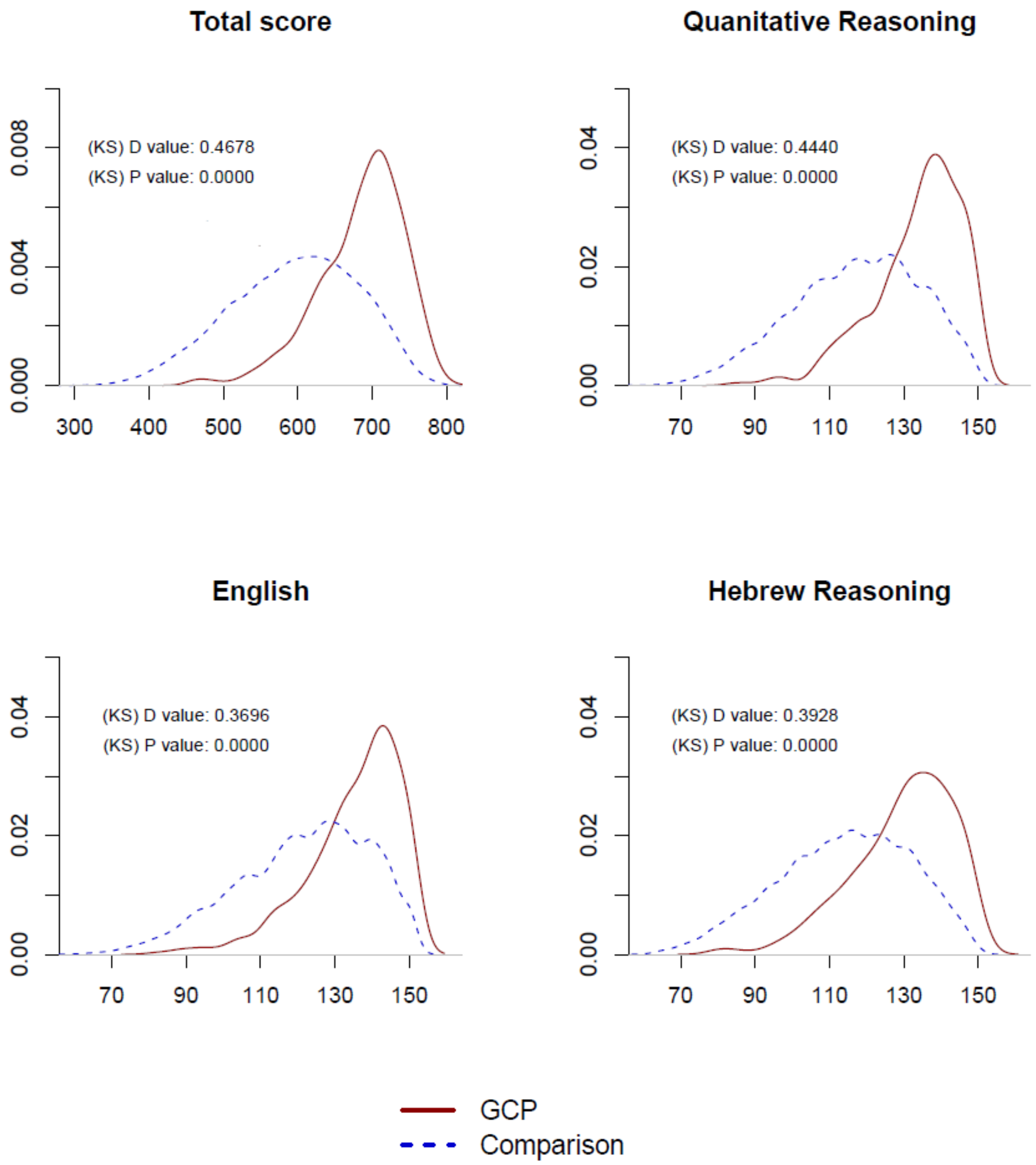
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on marriage before 30, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A27: GCP Effects on Having Children Before 30, By Sample



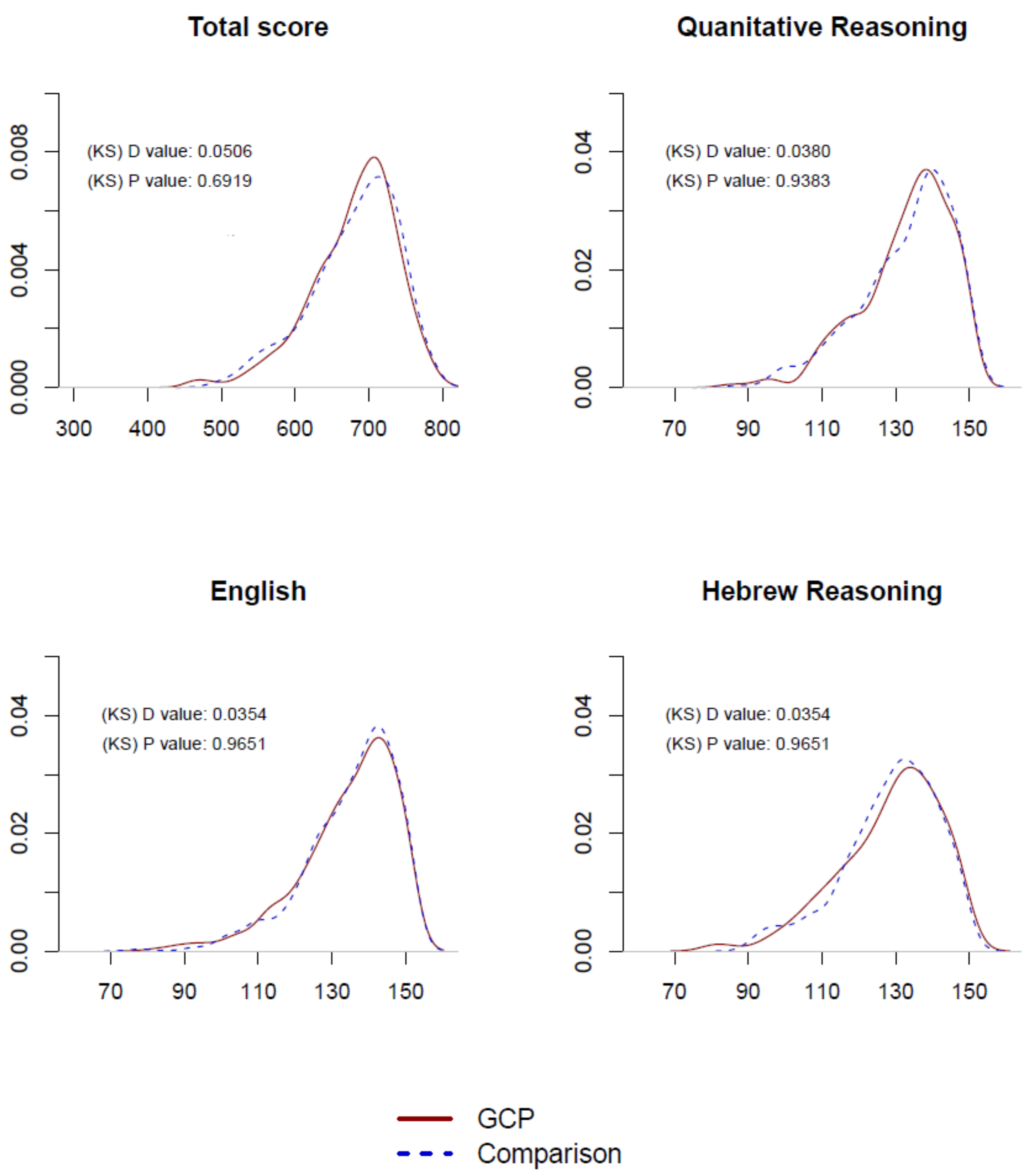
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on having children before 30, by sample. *Main* is the main sample, described in text. *Main only num* is identical to the main sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score). *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. *Extend all only num* is identical to Extend all sample, except that we omit the Hebrew and English UPET scores from the matching (include only the numeric score).

Appendix Figure A28: Psychometric Scores Distributions, GCP and Comparison Group Pull, Before Matching (2006-2010 Cohorts Sample)



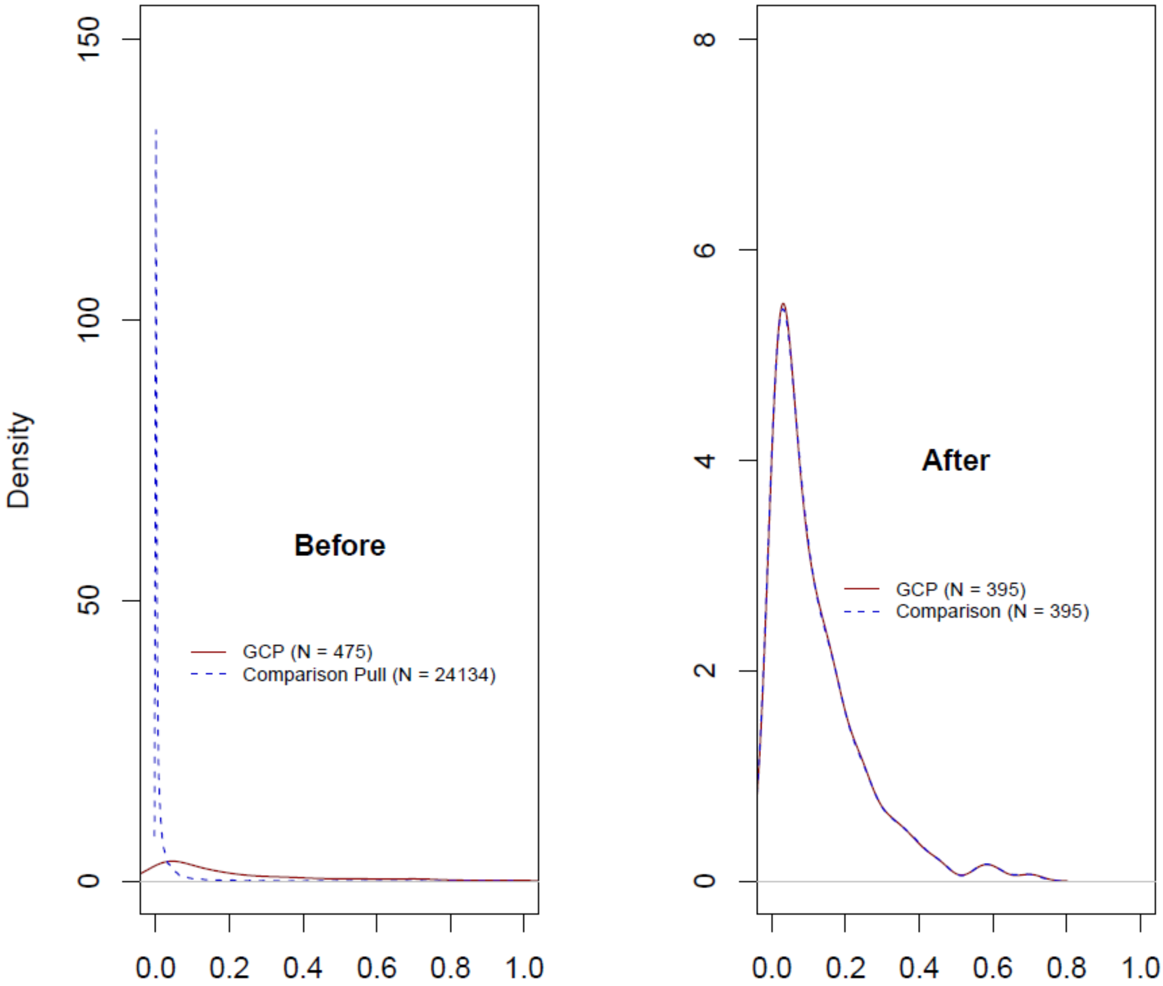
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group pull (before matching). The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions.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

Appendix Figure A29: Psychometric Scores Distributions, GCP and Comparison Group Pull, After Matching I (2006-2010 Cohorts Sample)



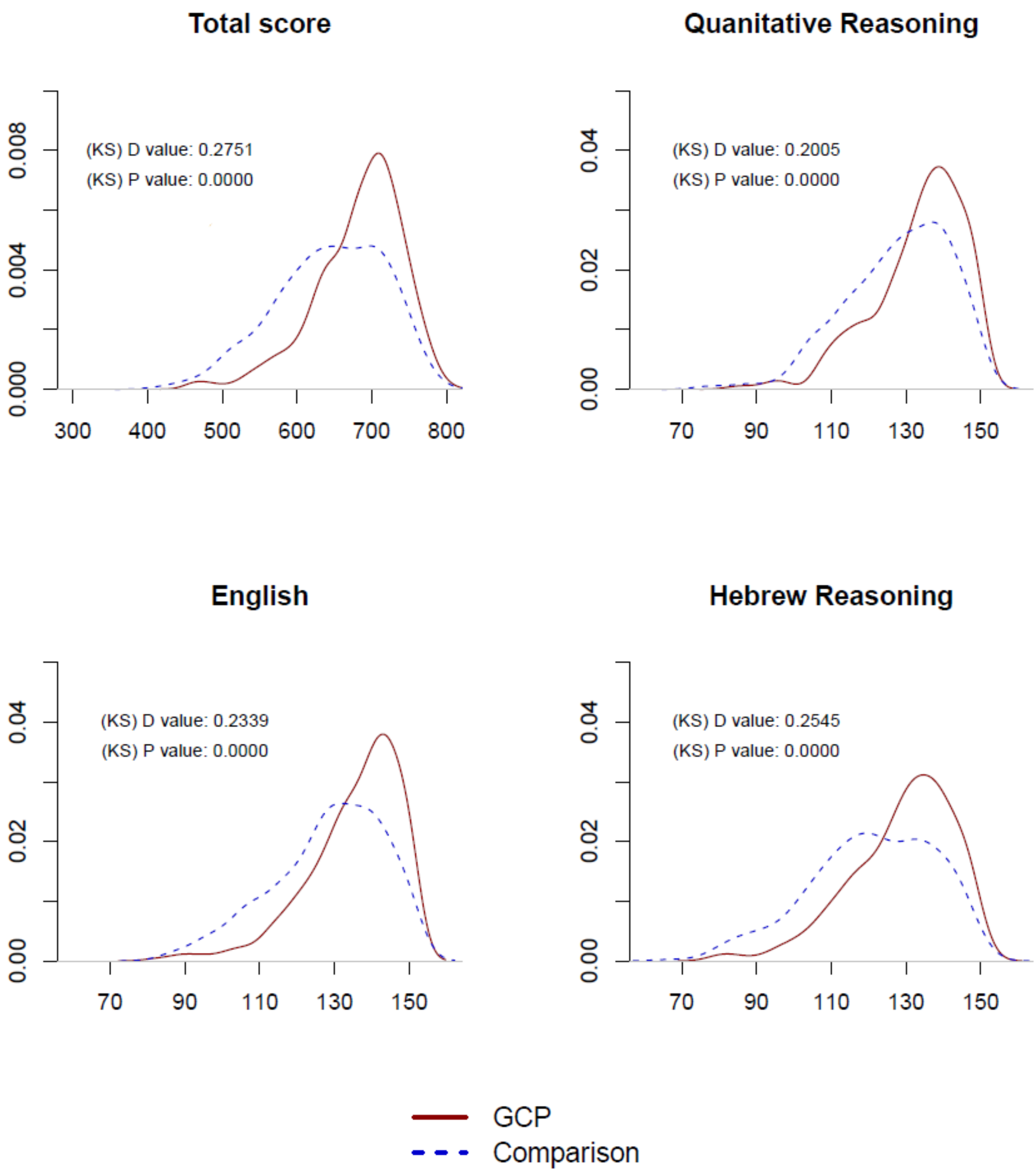
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group after the matching. The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions.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

Figure A30: Propensity Score Distributions, Before and After Matching I (2006-2010 Cohorts Sample)



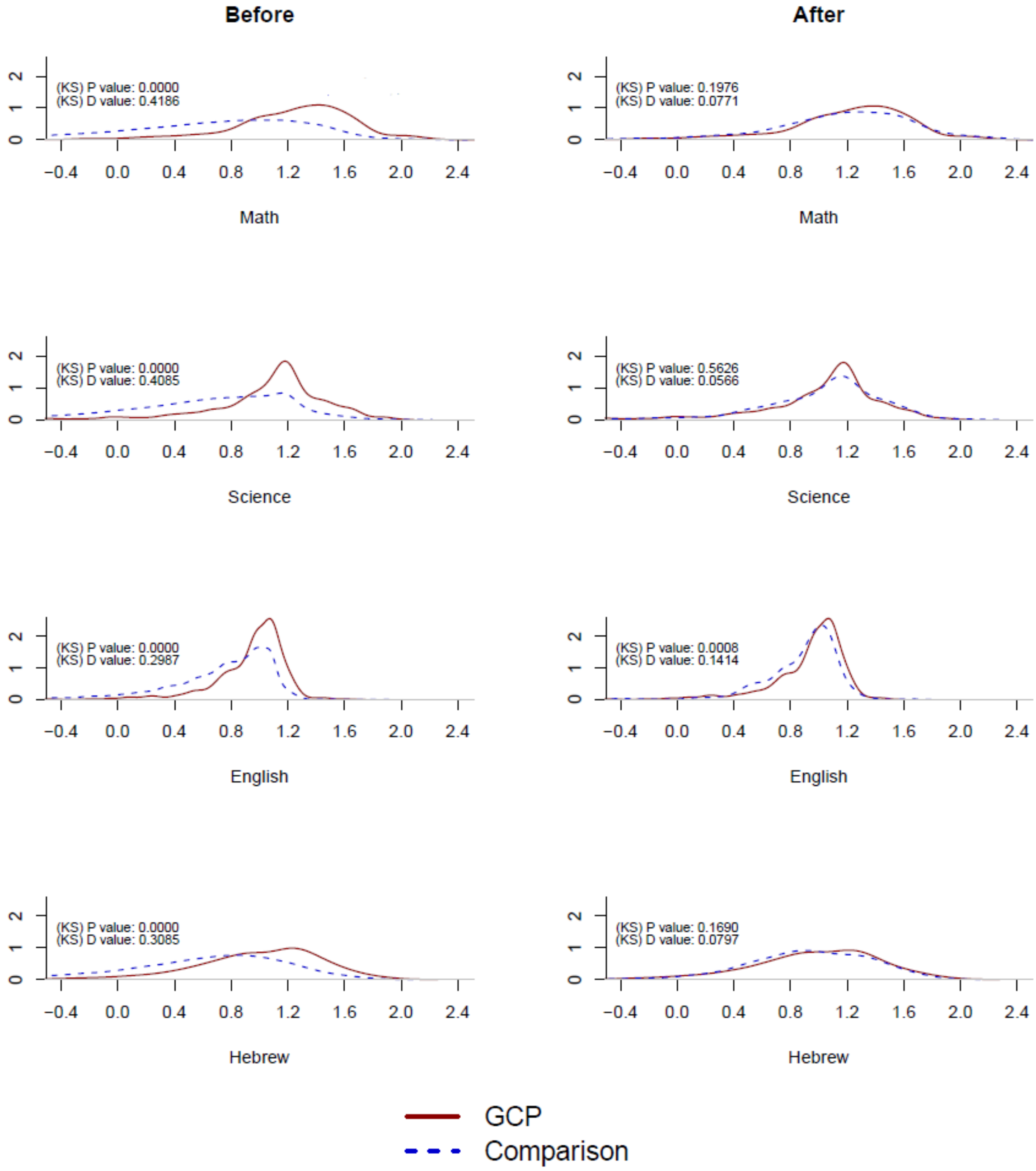
Notes: This figure plots the distribution of the propensity score, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching,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

Appendix Figure A31: Psychometric Scores Distributions, GCP and Comparison Group Pull, After Matching II (2006-2010 Cohorts Sample)



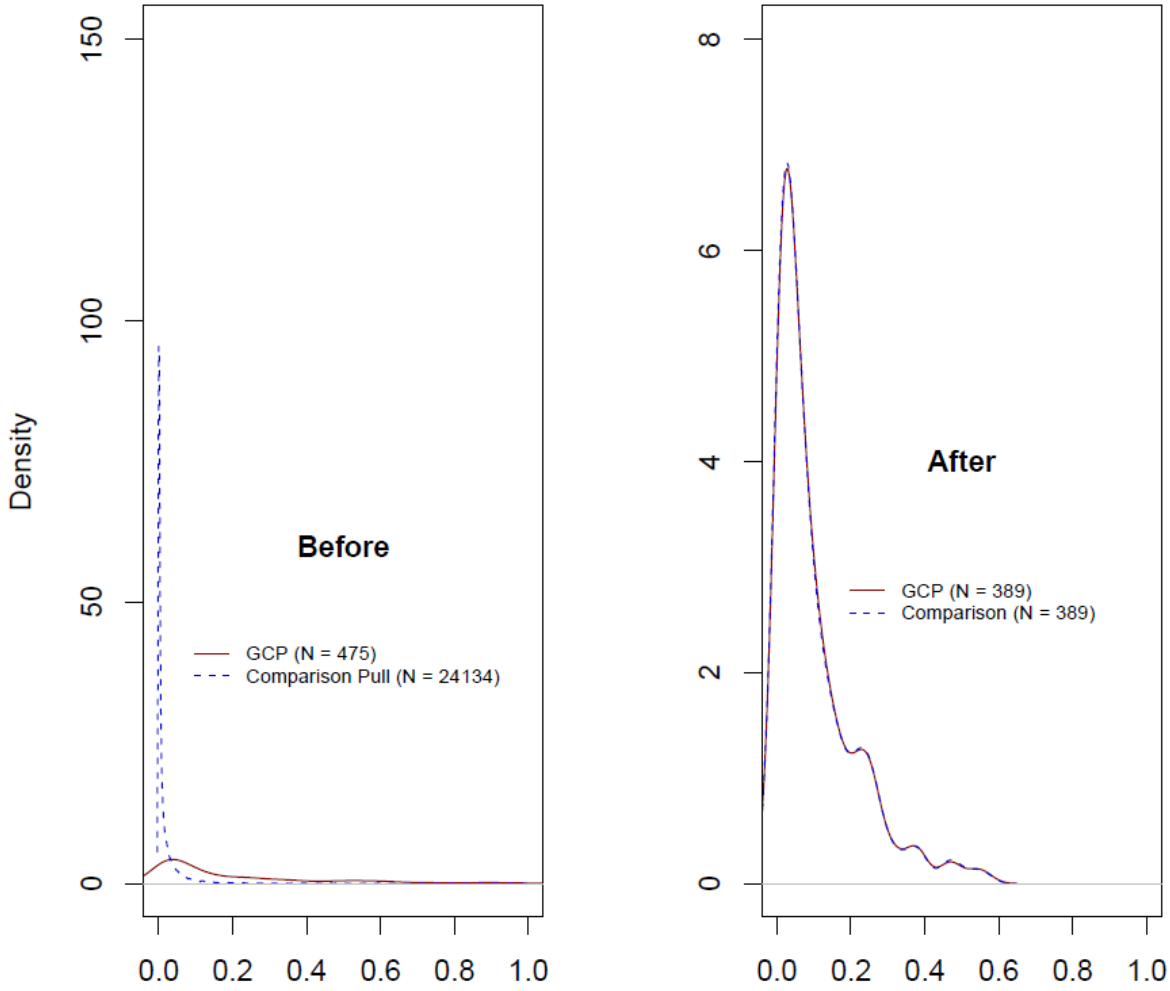
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group after the matching. The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions.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

Figure A32: Pre-treatment Middle-school Test Scores, Before and After Matching II



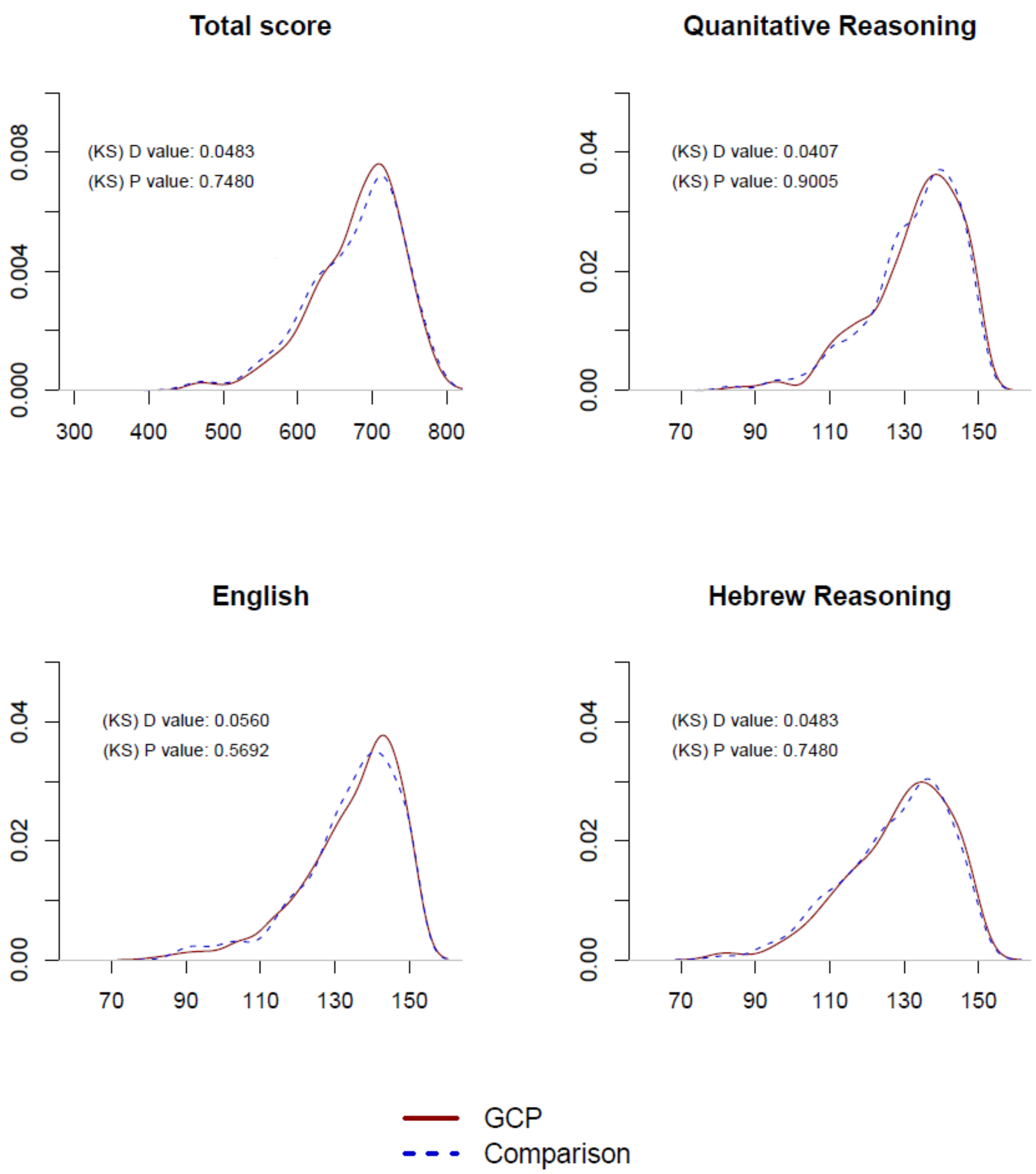
Notes: This figure plots the distribution of the Pre-treatment Middle-school Test (Metzav) test scores, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graphs on the left show the distributions before the matching, and the graphs on the right shows the distributions after the matching (version II, includes these tests in the logit regression). 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, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Figure A33: Propensity Score Distributions, Before and After Matching II (2006-2010 Cohorts Sample)



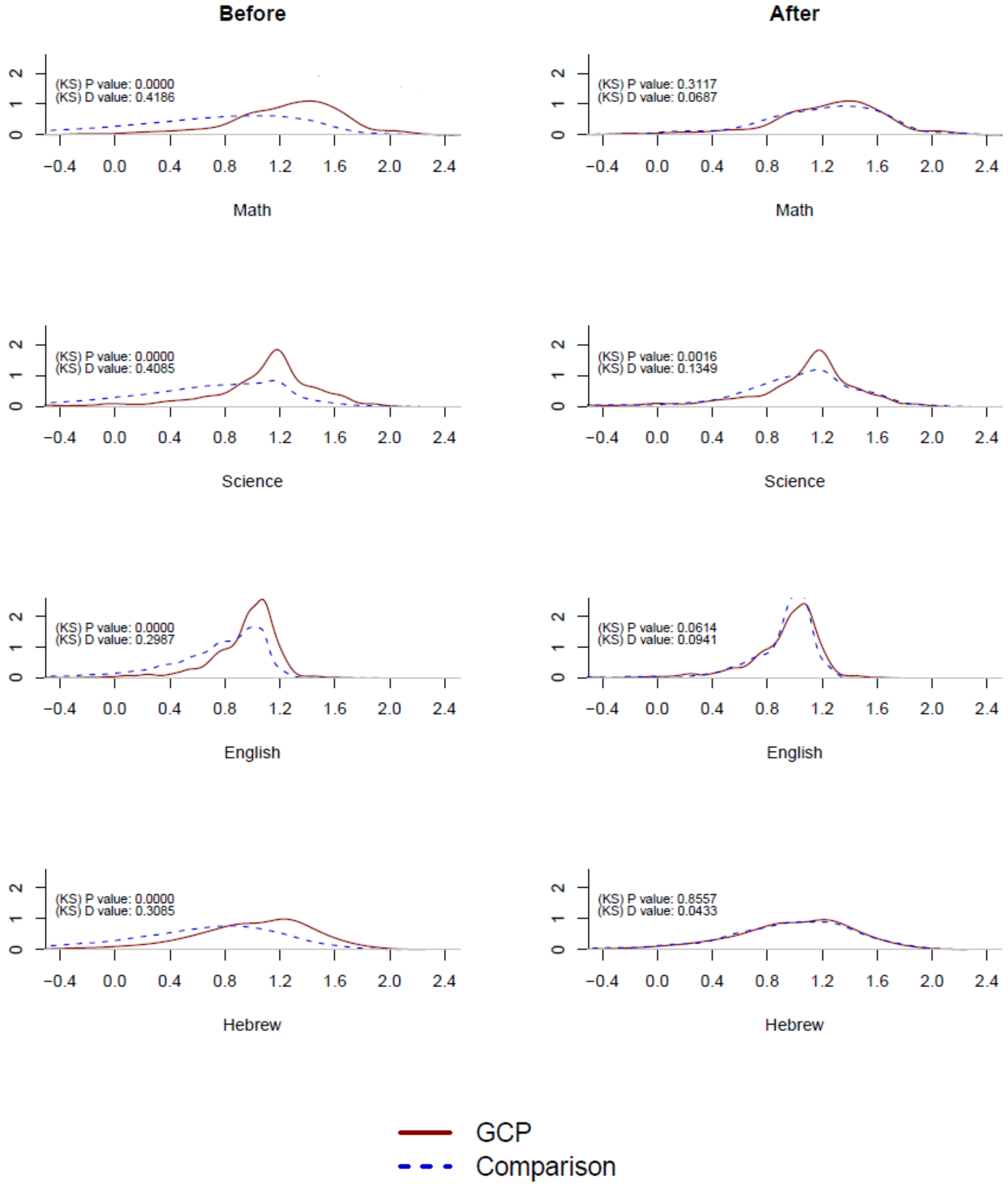
Notes: This figure plots the distribution of the propensity score, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching,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

<u>Appendix Figure A34</u>: Psychometric Scores Distributions, GCP and Comparison Group Pull, After Matching III (2006-2010 Cohorts Sample)



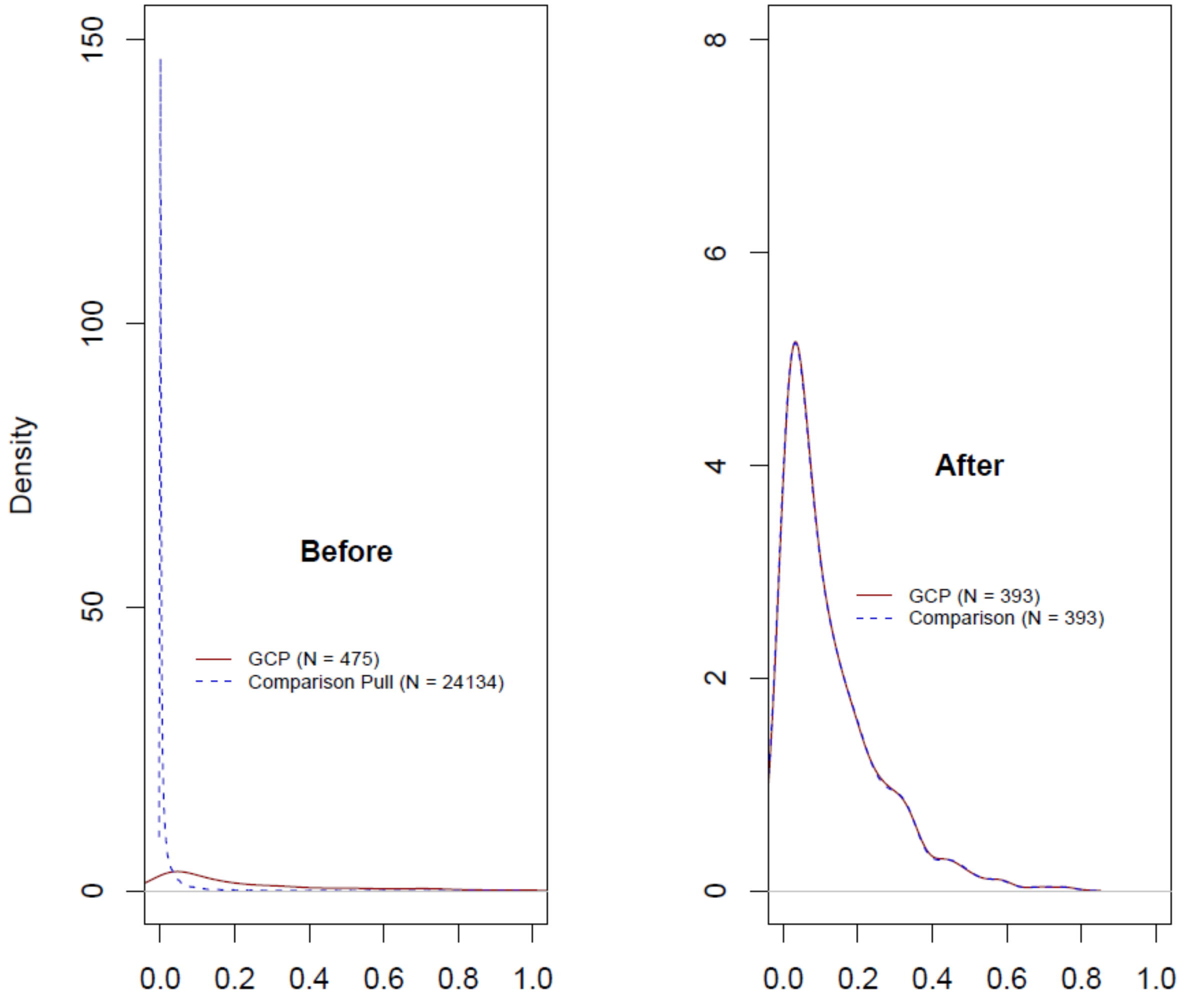
Notes: This figure plots the distribution of psychometric scores, by group- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group after the matching. The graphs also show the Kolmogorov–Smirnov test for the equality of the probability distributions.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

Figure A35: Pre-treatment Middle-school Test Scores, Before and After Matching III



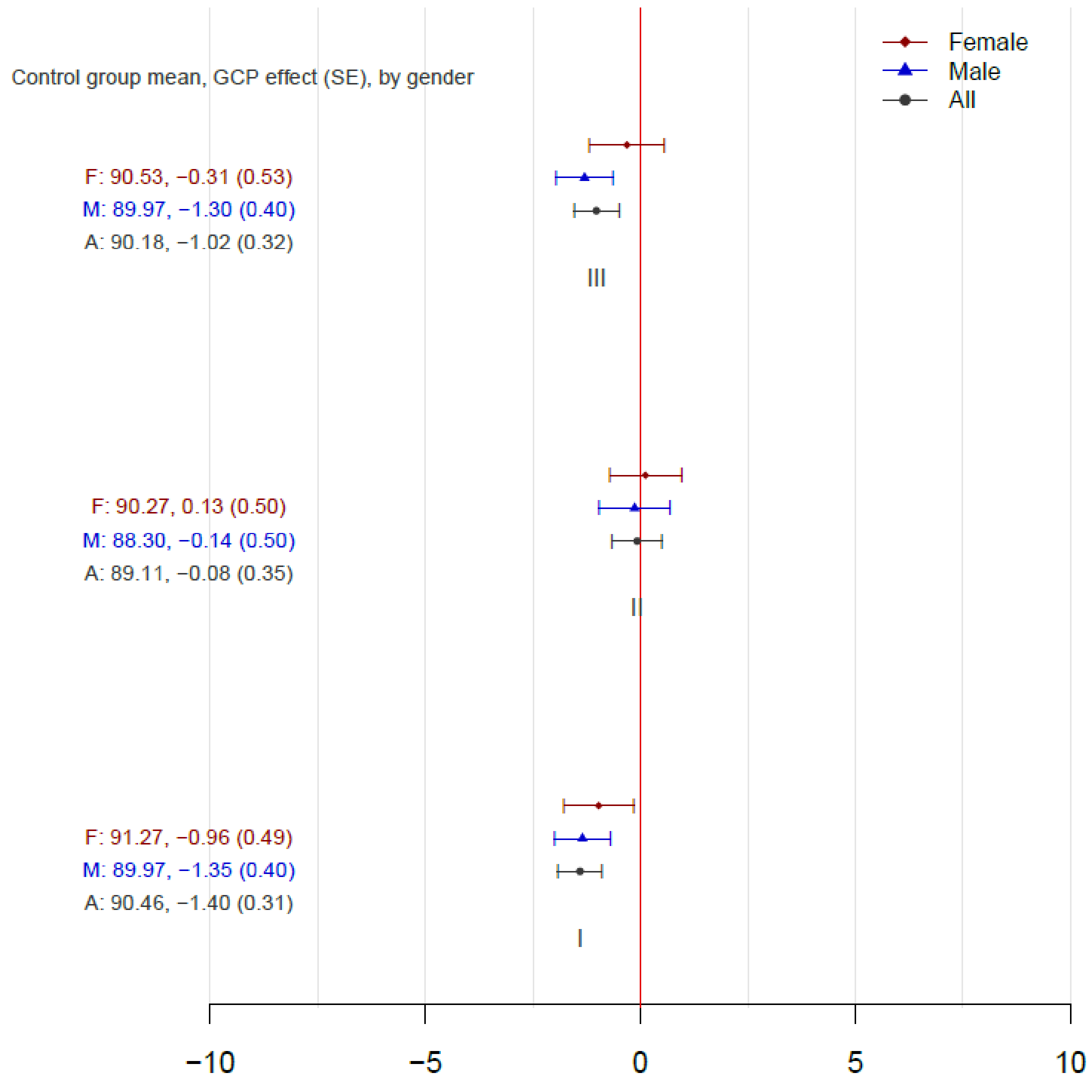
Notes: This figure plots the distribution of the Pre-treatment Middle-school Test (Metzav) test scores, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graphs on the left show the distributions before the matching, and the graphs on the right shows the distributions after the matching (version III, includes these tests in the logit regression). 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, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Figure A36: Propensity Score Distributions, Before and After Matching III (2006-2010 Cohorts Sample)



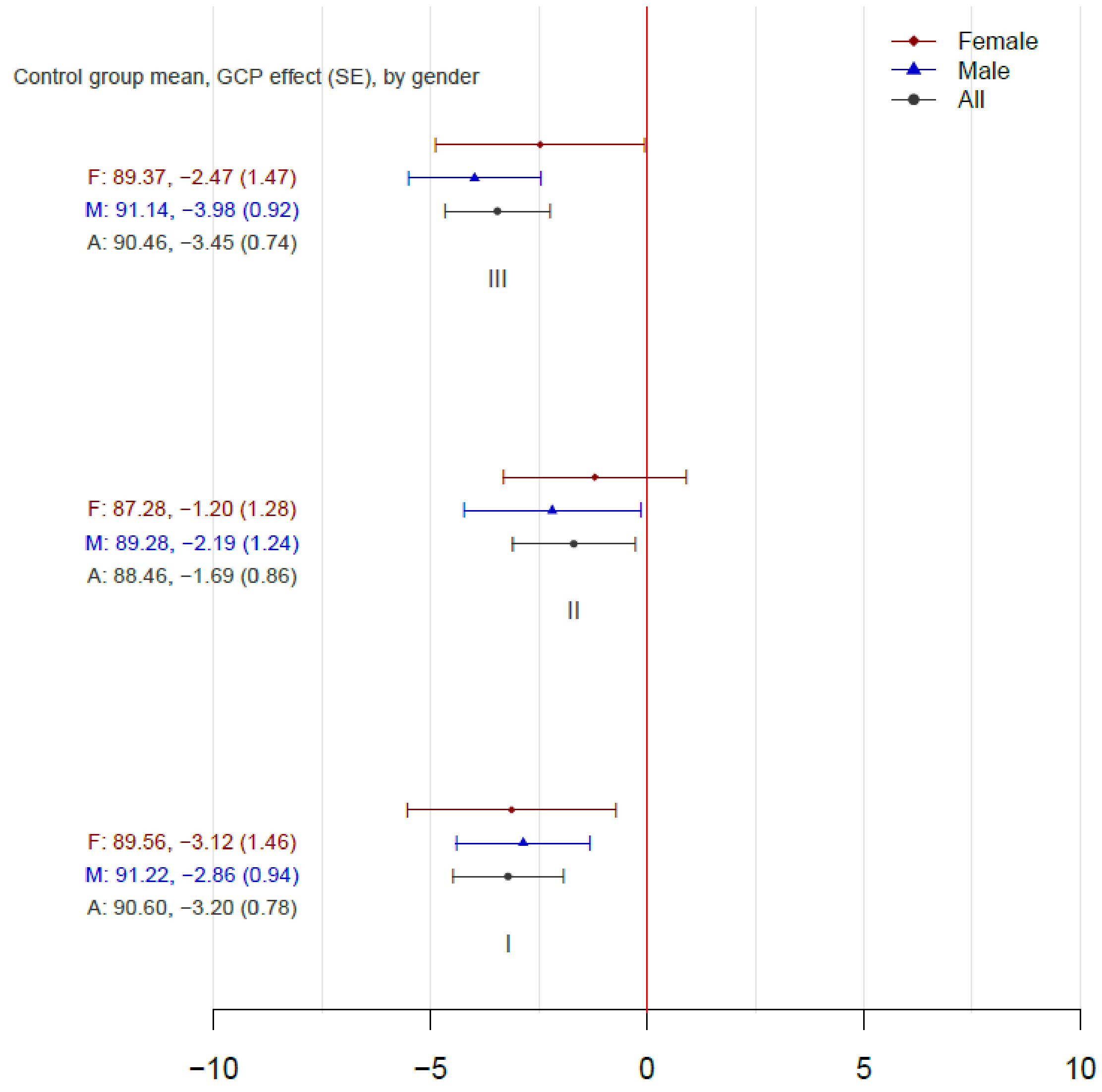
Notes: This figure plots the distribution of the propensity score, by groups- the red solid line represents the sample of GCP students, and the blue dashed line represents the comparison group (includes non-GCP students from other cities). The graph on the left shows the distributions before the matching,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

Appendix Figure A37: GCP Effects on Bagrut Mean Compsite Score, By Matching (2006-2010 Cohorts Sample)



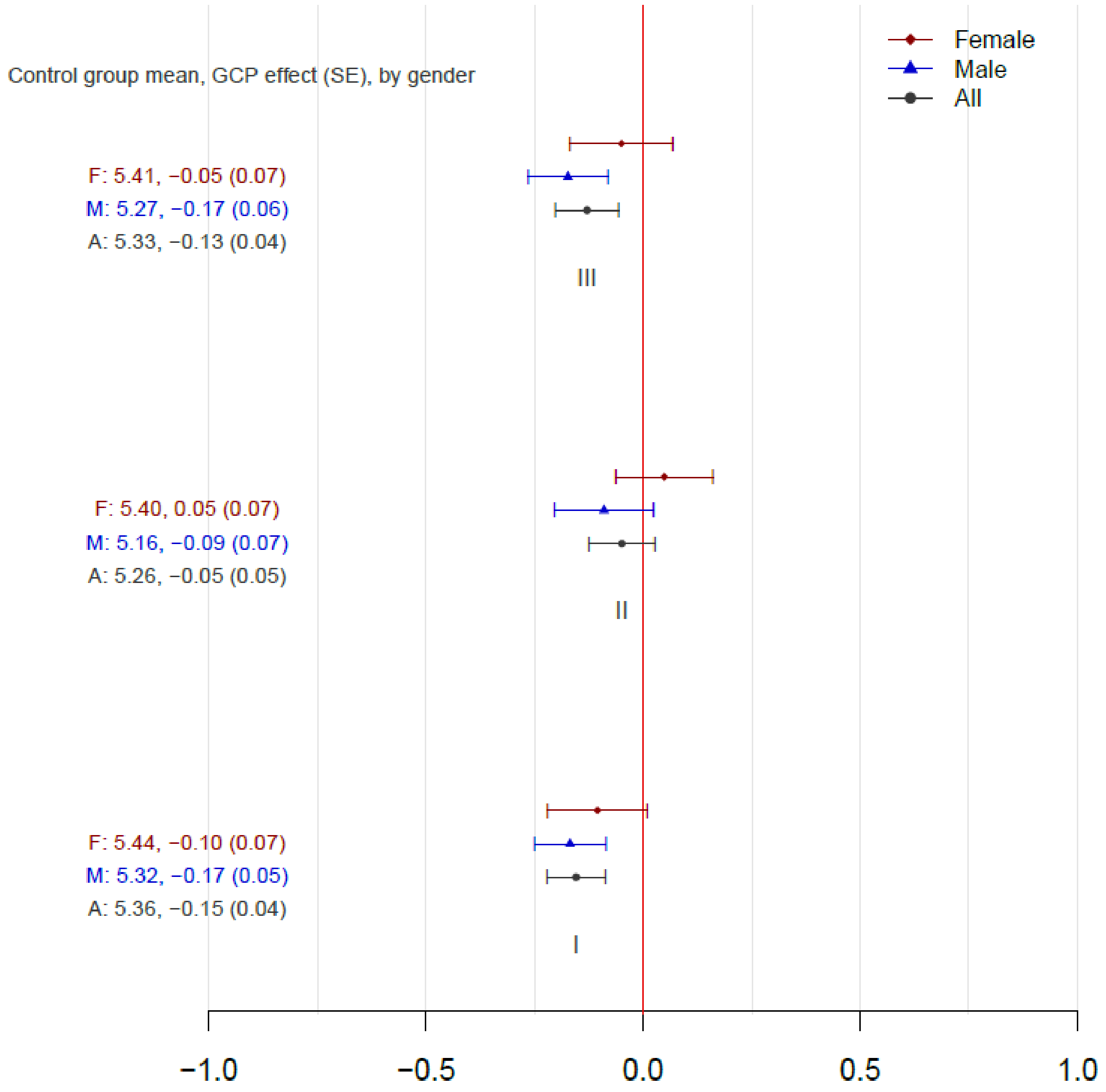
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut mean composite score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A38: GCP Effects on Bagrut Math Score, By Matching (2006-2010 Cohorts Sample)



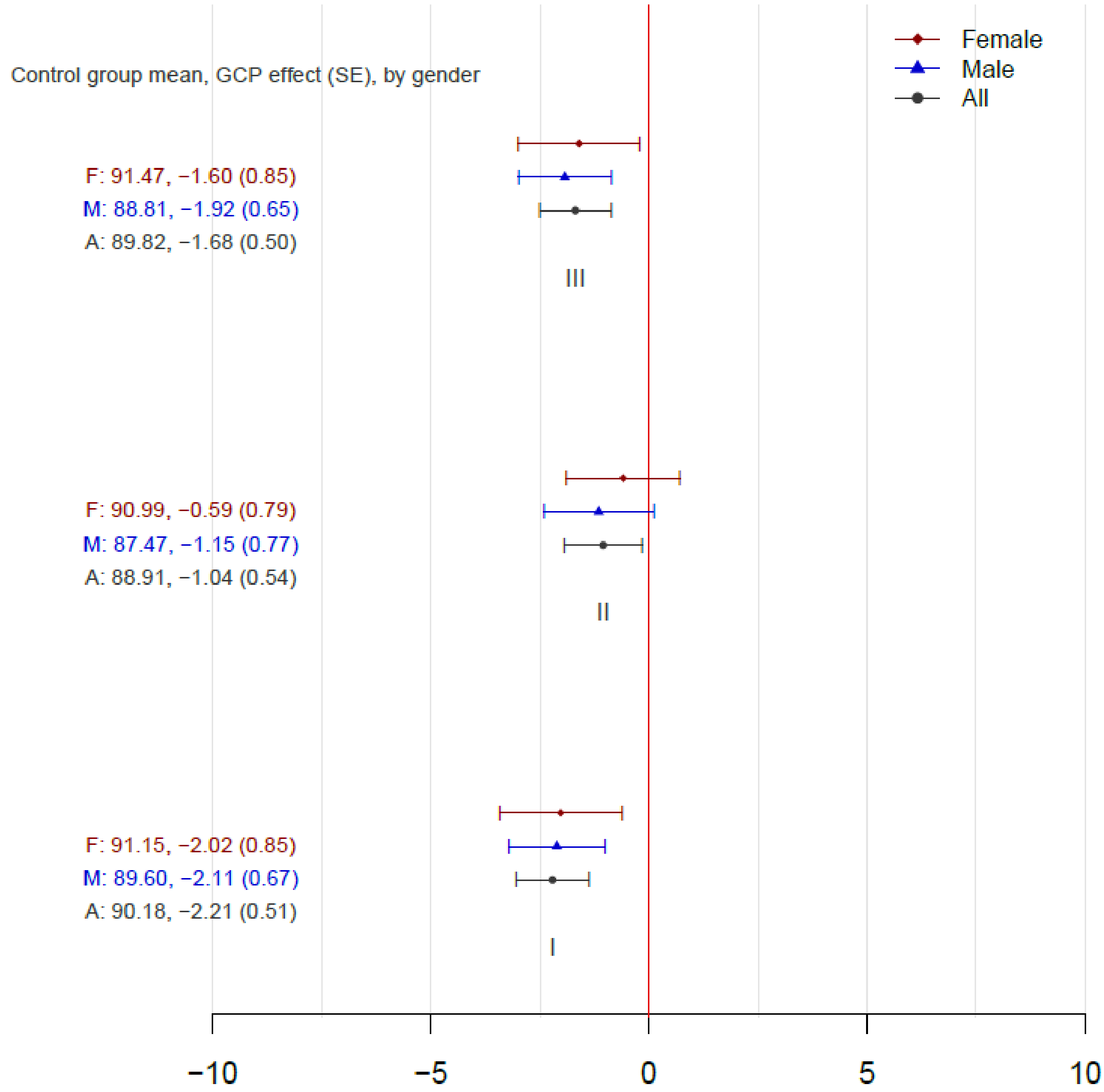
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut math score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A39: GCP Effects on Bagrut Hebrew Score, By Matching (2006-2010 Cohorts Sample)



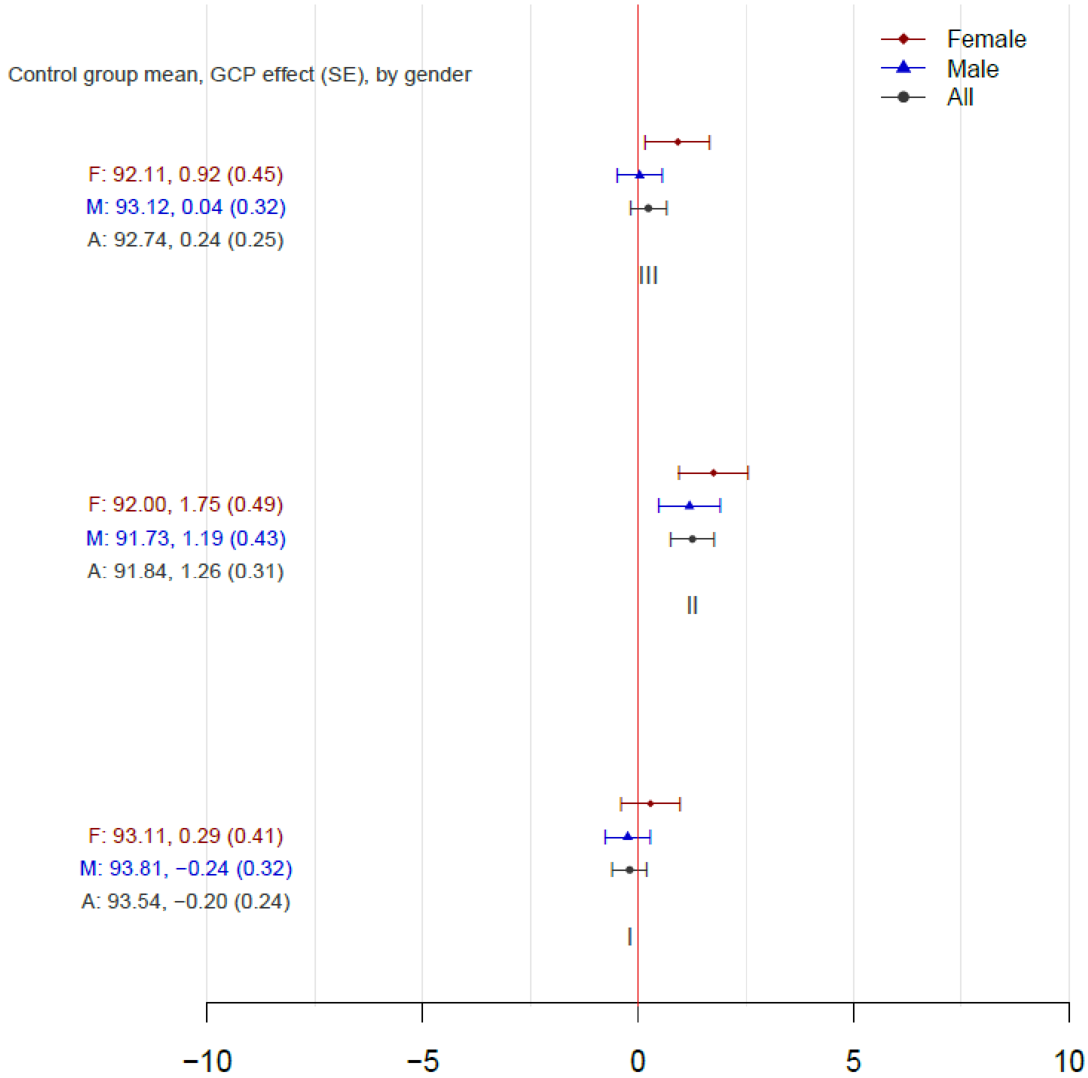
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut hebrew score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A40: GCP Effects on Bagrut Bible Score, By Matching (2006-2010 Cohorts Sample)



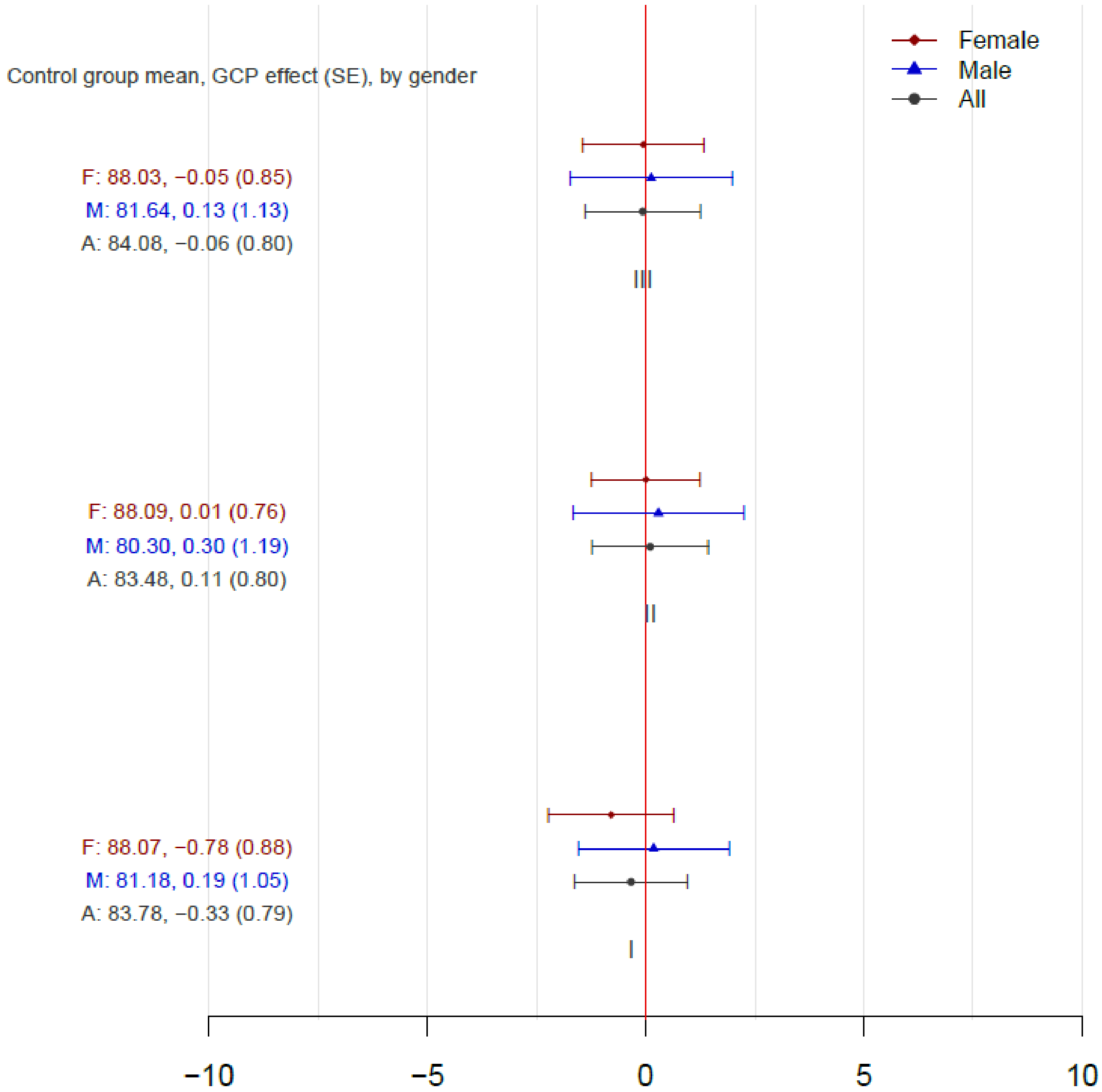
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut bible score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

<u>Appendix Figure A41</u>: GCP Effects on Bagrut English Score, By Matching (2006-2010 Cohorts Sample)



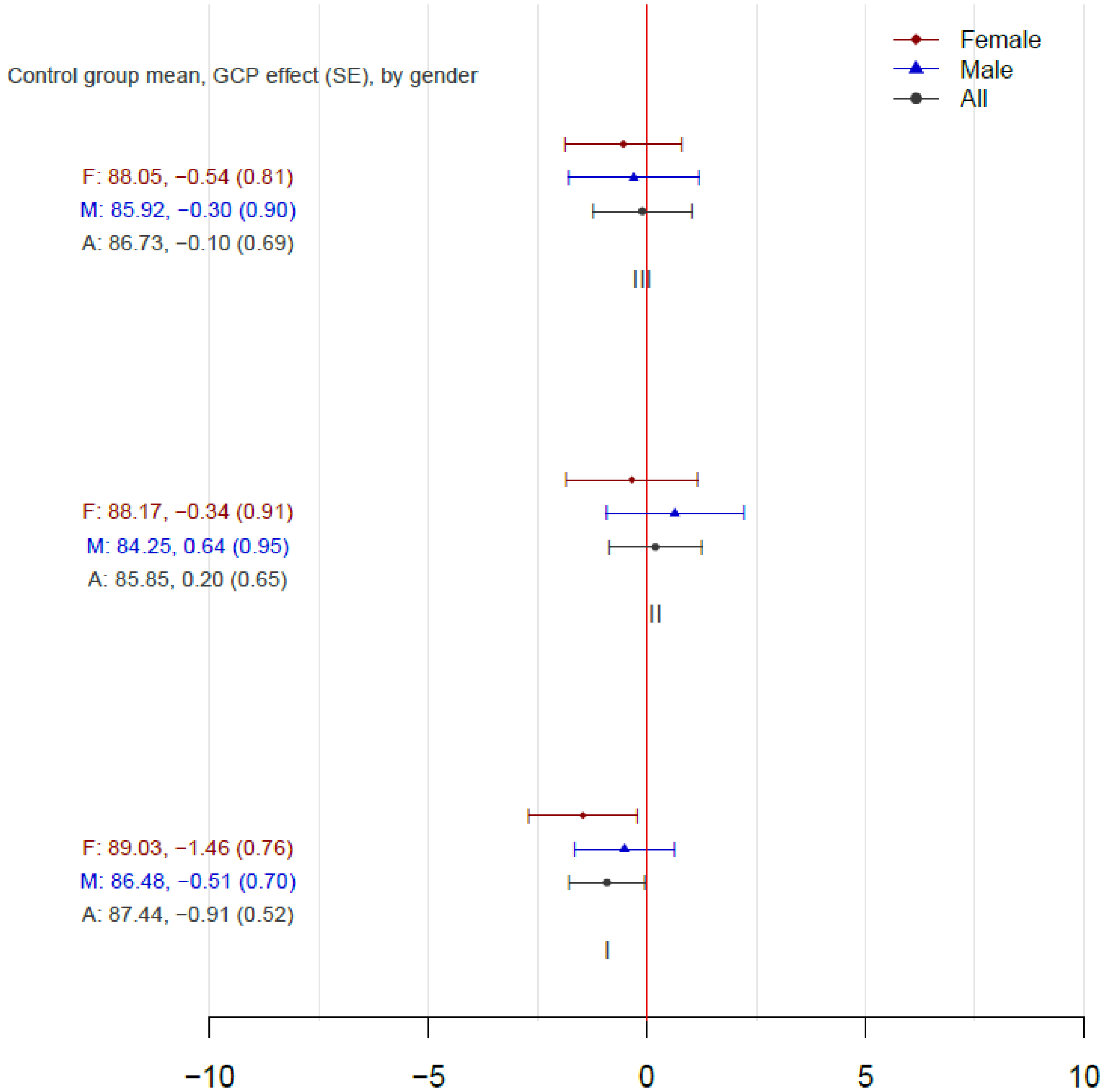
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut english score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A42: GCP Effects on Bagrut Literature Score, By Matching (2006-2010 Cohorts Sample)



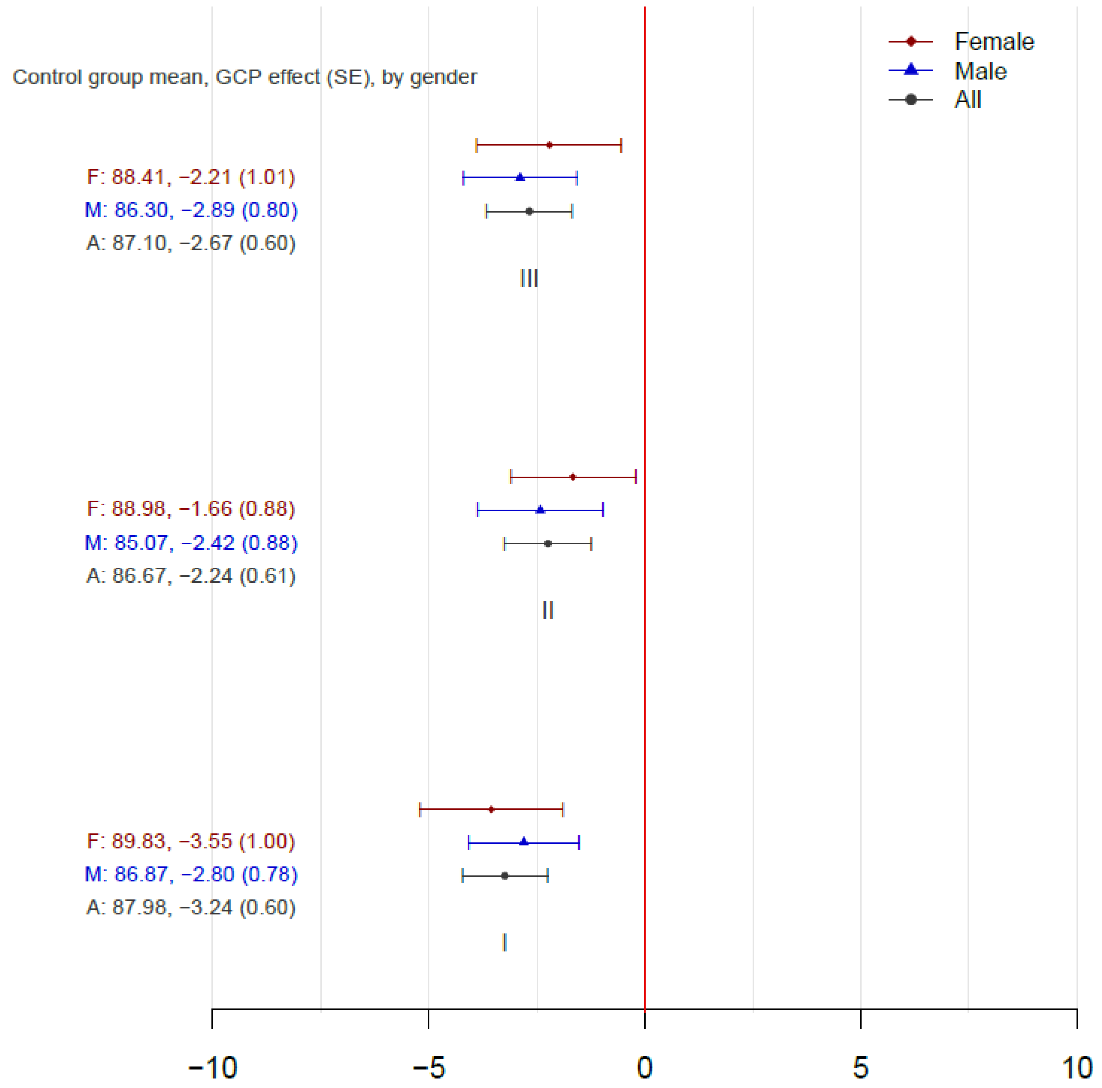
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut literature score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

<u>Appendix Figure A43</u>: GCP Effects on Bagrut History Score, By Matching (2006-2010 Cohorts Sample)



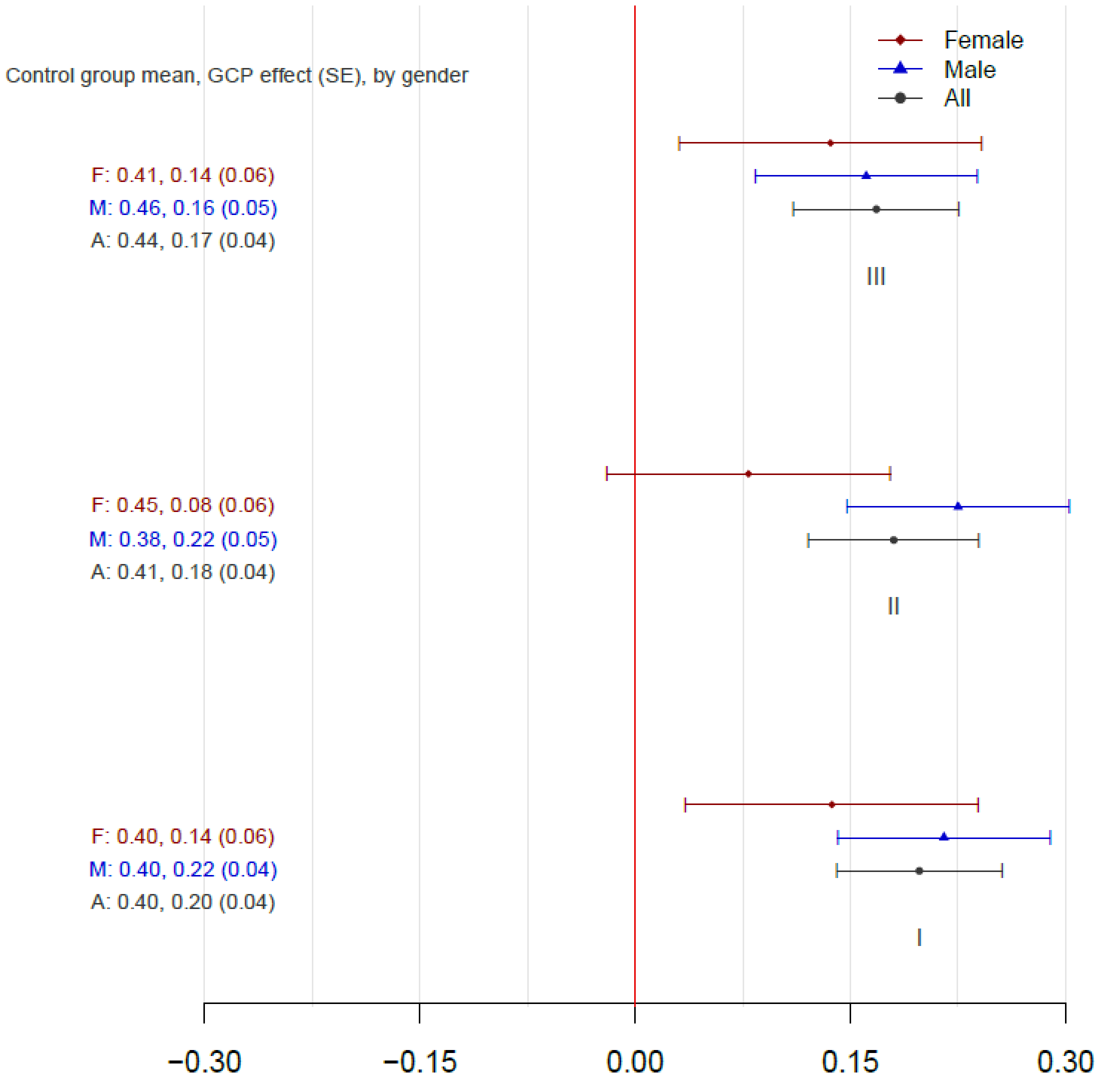
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut history score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A44: GCP Effects on Bagrut Civic Studies Score, By Matching (2006-2010 Cohorts Sample)



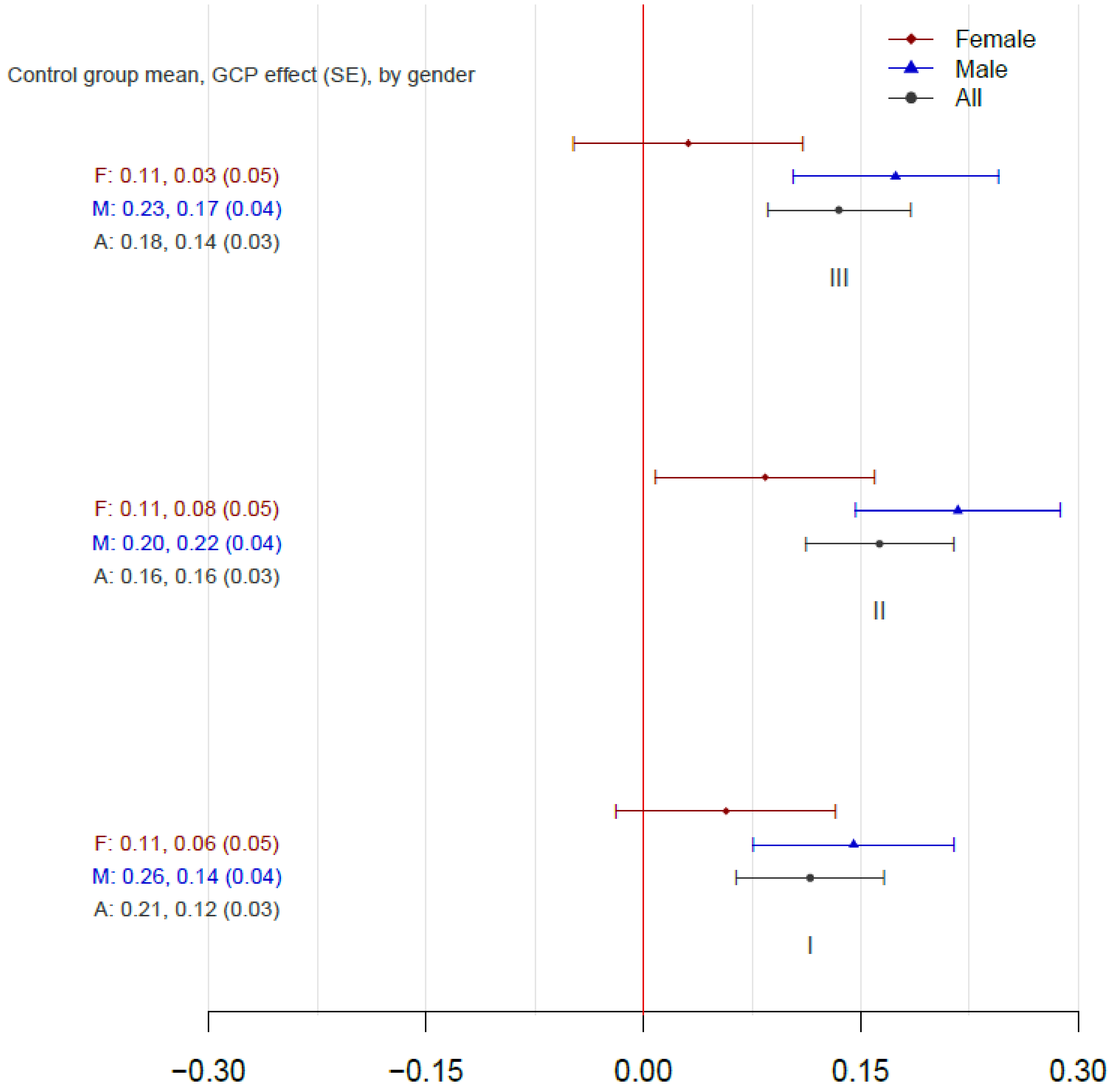
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on Bagrut civic studies score, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

<u>Appendix Figure A45</u>: GCP Effects on Double Major BA, By Matching (2006-2010 Cohorts Sample)



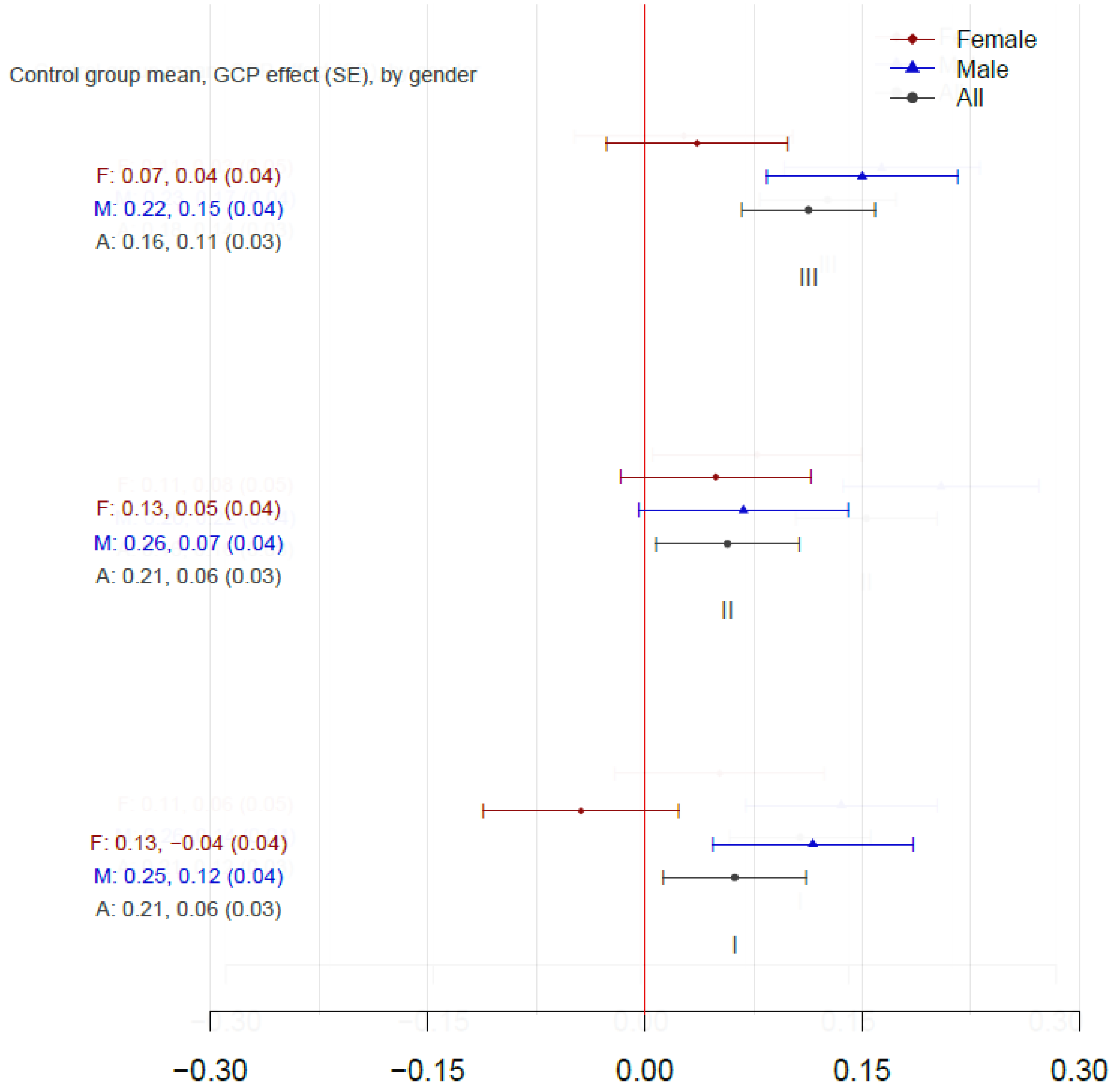
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on double major BA, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A46: GCP Effects on Double Major in STEM BA, By Matching (2006-2010 Cohorts Sample)



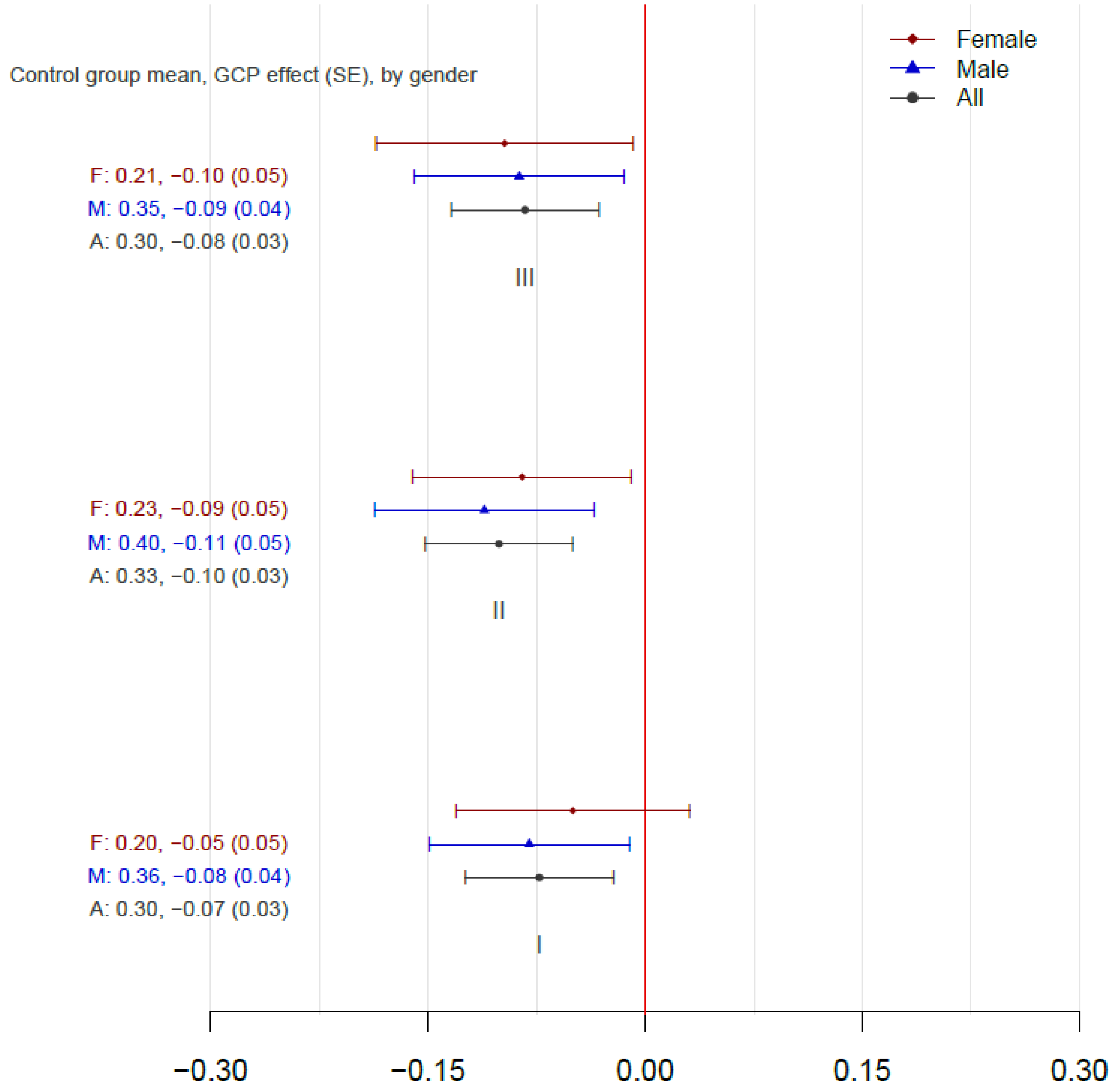
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on double major in STEM BA, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

<u>Appendix Figure A47</u>: GCP Effects on BA in Math, Statistics, and Computer Sciences, By Matching (2006-2010 Cohorts Sample)



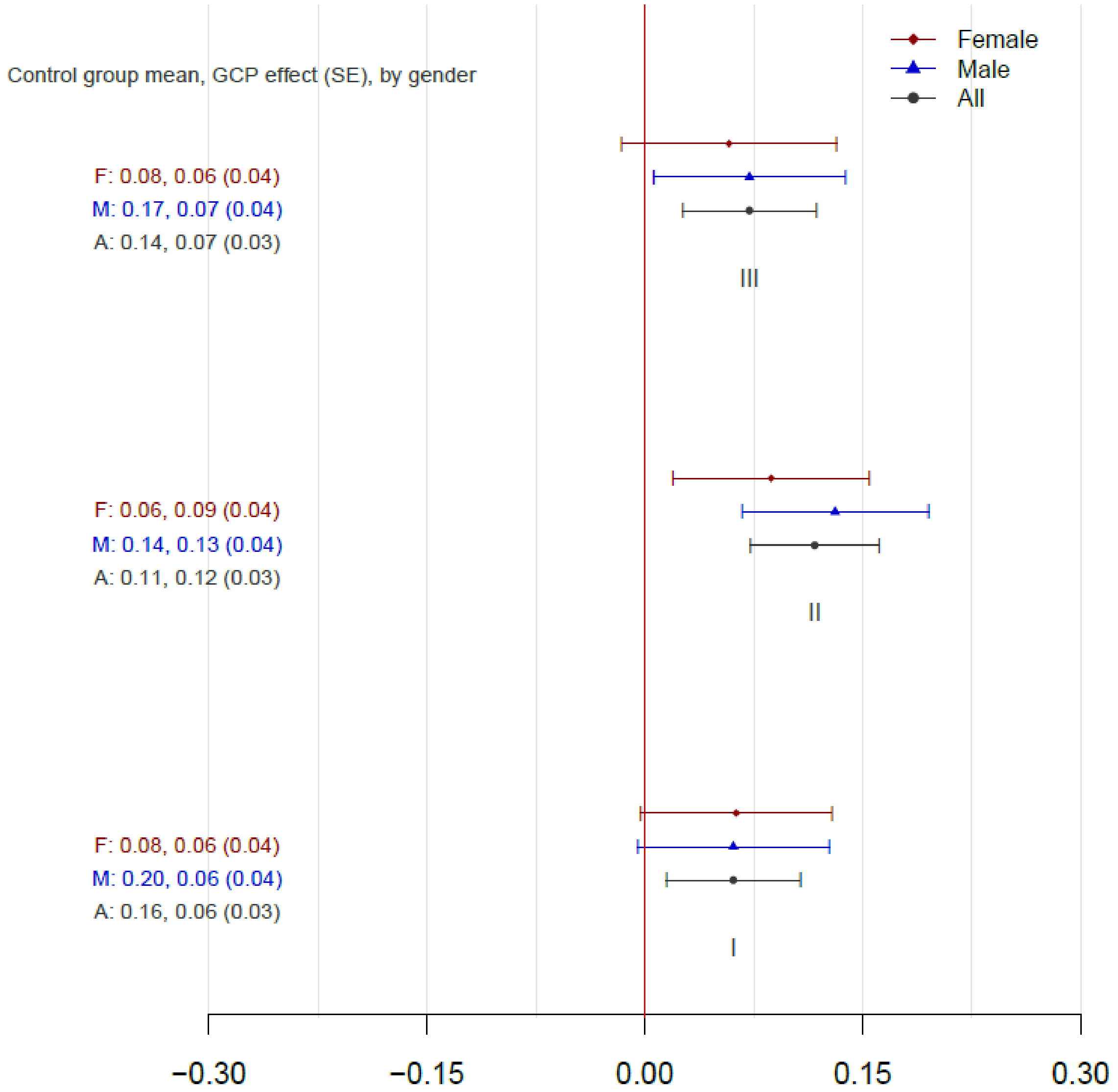
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in math, statistics and computer sciences, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A48: GCP Effects on BA in Engineering, By Matching (2006-2010 Cohorts Sample)



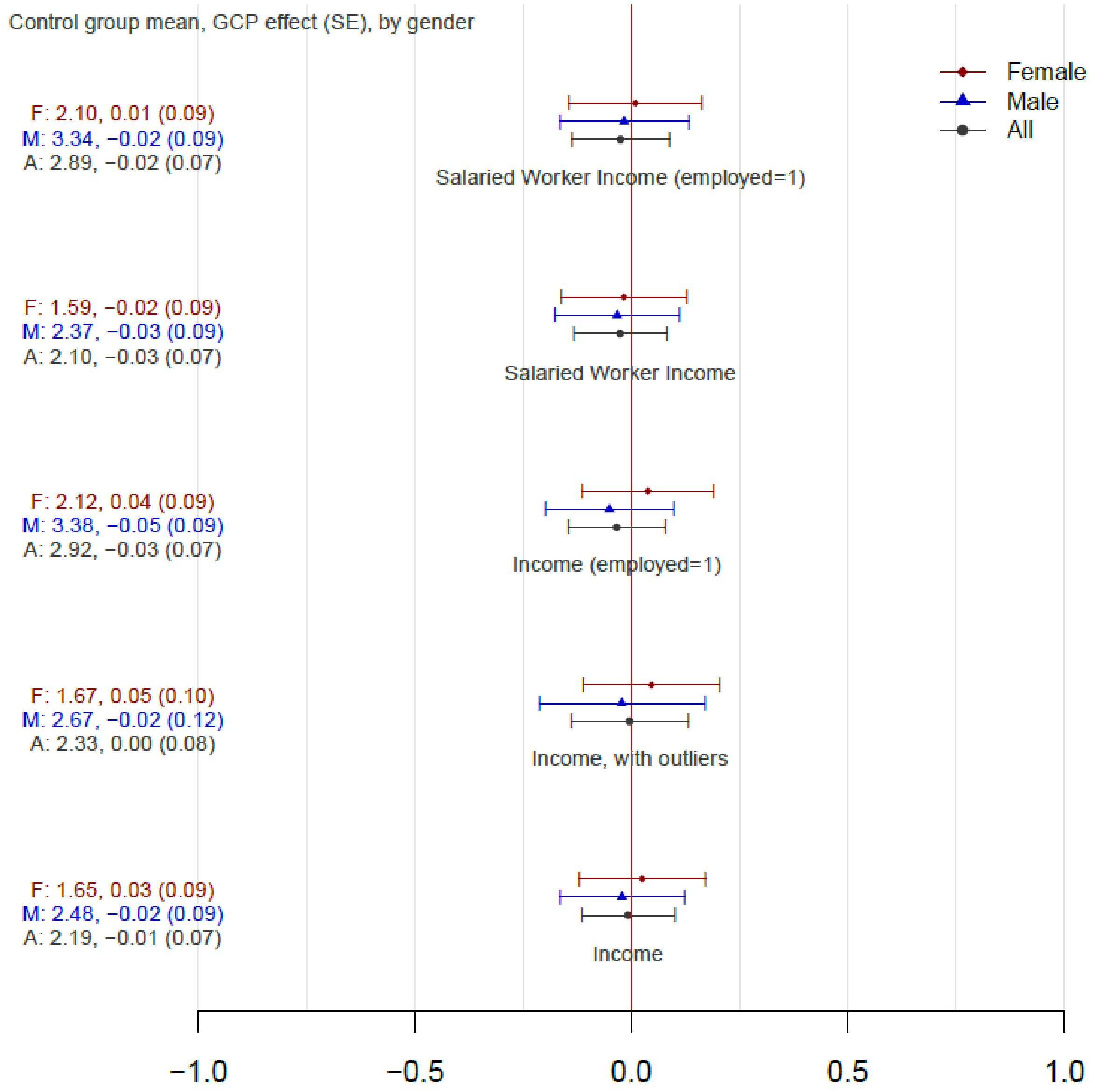
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in engineering, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A49: GCP Effects on BA in Physical Sciences, By Matching (2006-2010 Cohorts Sample)



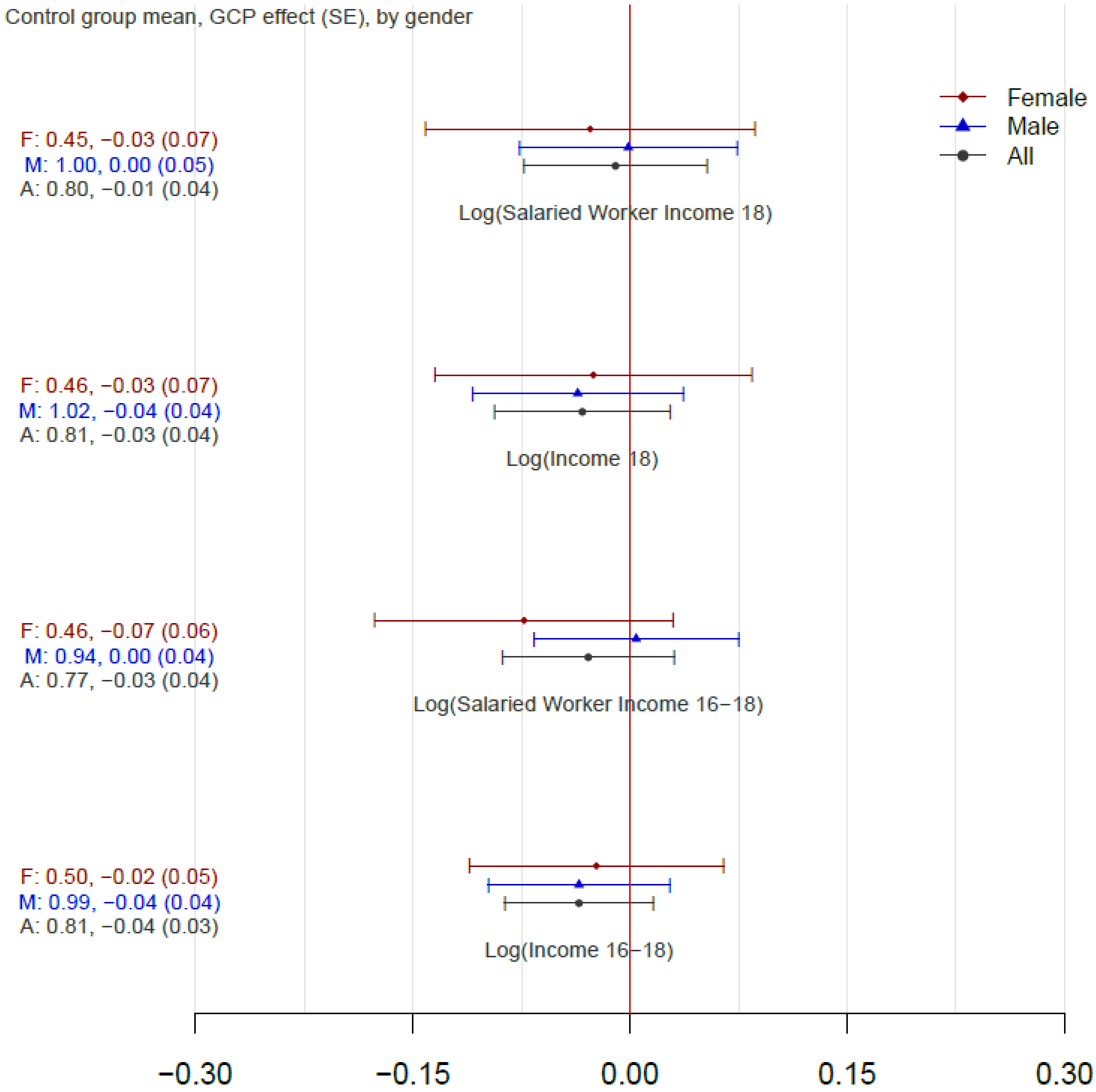
Notes: This figure plots point estimates and 90% confidence intervals for the effects of GCP on BA in physical sciences, by matching verssion: I refers to the matching that is identical to our main matching described in text, II includes the metzav 8th grade test in the matching specification and excludes the UPET scores, and III includes both metzav and UPET scores. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test at any age.

Appendix Figure A50: GCP Effects on Average Annual Income, in 2016-2018



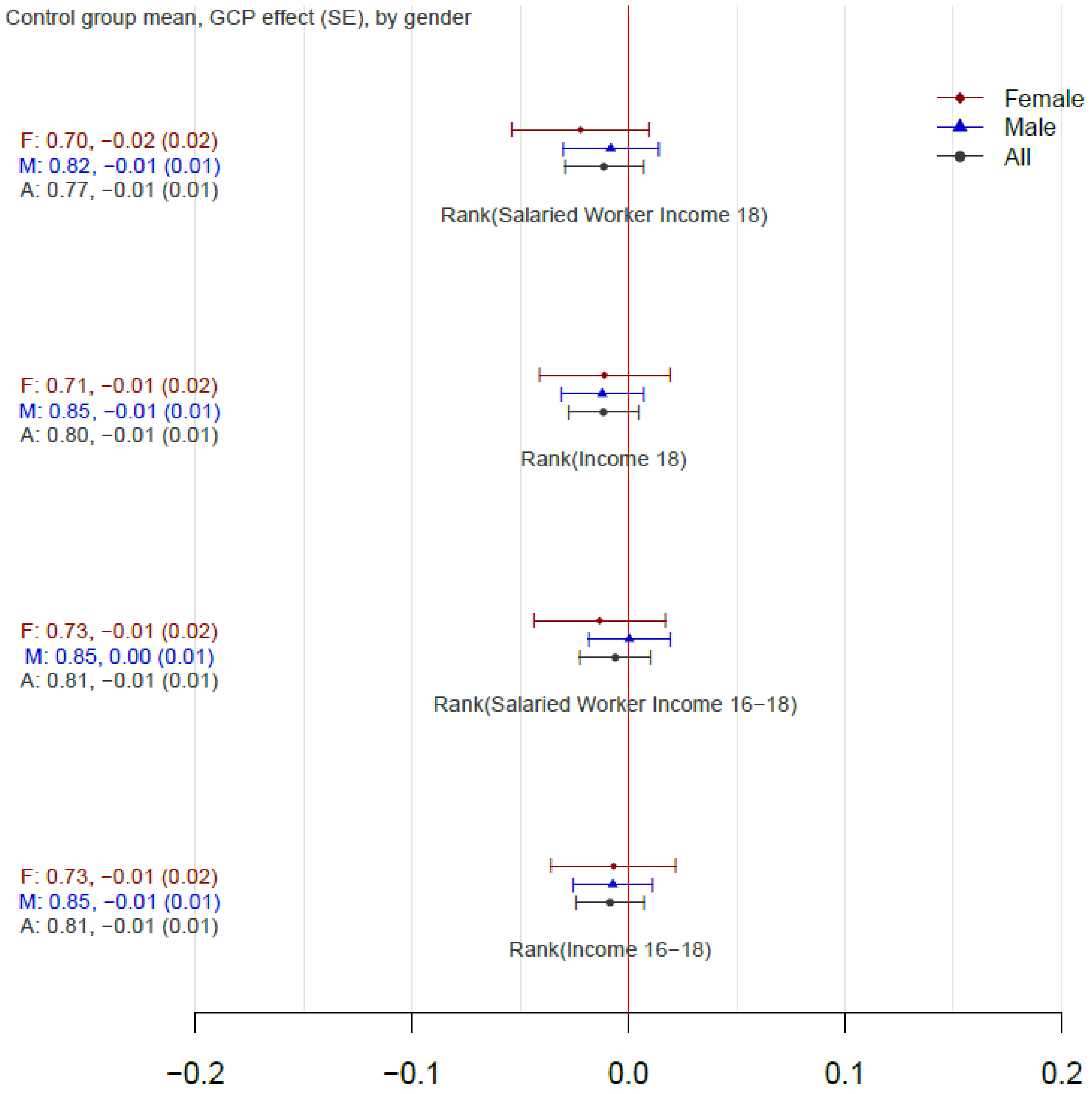
Notes: This figure plots the point estimates and 90% confidence intervals for the effects of GCP on average annual income, in 2016-2018. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males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

Appendix Figure A51: 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 2016-2018. Red dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The sample includes students from the cohorts of high-school graduetes in 1992-2005, who took the psychometric test during their 10-11th grade.

Appendix Figure A52: 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 dots and lines represnts the sample of females, blue dots and lines represents the sample of males, and dark grey dots and lines represents the full sample (males and females). The sample includes students from the cohorts of high-school graduetes in 1992-2005, who took the psychometric test during their 10-11th grade.

Appendix Table A1: Geographical Mobility

	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)
Moved between 6th and 10th grade	0.16	0.15	0.00	0.94
Moved between 9th and 10th grade	0.06	0.06	0.00	0.97
Moved between 6th and 9th grade	0.14	0.12	0.02	0.51
Number of Observations	613	613		

Notes: This table presents comparison of 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, Column (3) and (4) show the difference between the means and the corresponded p-value. The sample includes students from the cohorts of high-school graduetes in 2005-2013, who took the psychometric test during their 10-11th grade. * represents statistical significance at the 90% level.

Appendix Table A2: Alternative Matching Methods

Panel A. Girls

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Replacement	0.000	-0.044	0.100*	0.028	0.042*
N: 1,189 (626 GCP)	(0.022)	(0.052)	(0.031)	(0.021)	(0.021)
Caliper = 0.05	0.004	-0.020	0.100*	0.015	0.023
N: 2,268	(0.022)	(0.052)	(0.029)	(0.020)	(0.021)
Caliper = 0.2	0.005	-0.042	0.085*	0.018	0.039*
N: 2,362	(0.021)	(0.048)	(0.028)	(0.020)	(0.020)

Panel B. Boys

Outcome:	Mean Bagrut		BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Replacement	-0.048*	-0.105*	0.086*	0.024	0.038*
N: 2,219 (1,208 GCP)	(0.018)	(0.034)	(0.022)	(0.016)	(0.021)
Caliper = 0.05	-0.035*	-0.068*	0.115*	0.040*	0.053*
N: 1,186	(0.017)	(0.035)	(0.021)	(0.014)	(0.019)
Caliper = 0.2	-0.021	-0.051	0.102*	0.046*	0.051*
N: 1,242	(0.017)	(0.035)	(0.020)	(0.014)	(0.019)

*Notes: This table presents the main results of this paper in alternative matching method. The upper panel shows results for the girls in our main sample described in the text, and the lower panel shows the results for the boys in this sample. Replacement refers to PSM identical to the one described in the text, with the only change that 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 A3: Standard Errorrs Calculations, Girls Sample

Panel A.

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	0.0083	-0.0264	0.1053	0.0313	0.0670
SE	0.0294	0.0577	0.0284	0.0199	0.0202
SE Abadie Imbens	0.0247	0.0540	0.0277	0.0192	0.0202

Panel B.

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	0.0028	-0.0453	0.0860	0.0348	0.0573
SE	0.0220	0.0502	0.0289	0.0200	0.0198
SE Bootstrap	0.0230	0.0536	0.0293	0.0205	0.0205

Notes: This table presents standard errors calculations. The upper panel compare the standard standard errors that we use in our descriptive tables (1-5) with the correction offered by Abadie and Imbens (2008). The lower panel compare the standard standard errors that we use in our main figures (5-11) with bootstrapped standard errors. The sample is our main girls sample.

Appendix Table A4: Standard Errorrs Calculations, Boys Sample

Panel A.

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	-0.0326	-0.0727	0.1361	0.0284	0.0577
SE	0.0214	0.0404	0.0203	0.0142	0.0192
SE Abadie Imbens	0.0188	0.0379	0.0205	0.0138	0.0195

Panel B.

Outcome:	Mean Bagrut	Math Bagrut	BA Double Major	PHD	Math, CS, Stats
	(1)	(2)	(3)	(4)	(5)
Estimated Effect	-0.0293	-0.0720	0.1319	0.0315	0.0587
SE	0.0172	0.0354	0.0202	0.0141	0.0190
SE Bootstrap	0.0175	0.0368	0.0207	0.0137	0.0184

Notes: This table presents standard errors calculations. The upper panel compare the standard standard errors that we use in our descriptive tables (1-5) with the correction offered by Abadie and Imbens (2008). The lower panel compare the standard standard errors that we use in our main figures (5-11) with bootstrapped standard errors. The sample is our main boys sample.

Appendix Table A5: 8th Grade Metzav Test Scores

	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)
Math	1.20	1.13	0.07	0.06*
Science	1.06	0.94	0.12	<0.01*
English	0.90	0.89	0.01	0.60
Hebrew	0.97	0.89	0.08	0.03*
Number of Observations	395	395		

Notes: This table presents comparison of the scores in 8th grade tests, between treatment and comparison group. Column (1) shows the means among the treatment group, Column (2) shows the means among the comparison group, Column (3) and (4) show the difference between the means and the corresponded p-value. The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010 who took the psychometric test during 10th or 11th grade. * represents statistical significance at the 90% level.

Appendix Table A6: Correlation between Psychometric Score and 8th Grade Metzav Test Scores

	Numeric	Hebrew	English	Total
	(1)	(2)	(3)	(4)
A. GCP Participents				
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 psychometric score and 8th grade metzav test scores.The sample includes students who participated in the 8th grade Metzav tests, that is about half of the students in cohorts of high-school graduetes in 2006-2010. Panel A include only GCP participants, and panel B includes only students from cities with no GCP. Columns represent Psychometric scores, and rows represent metzav scores.

Appendix Table A7: Peers Quality

	GCP	Control	difference	p-value
	(1)	(2)	(3)	(4)
Father education	14.67	13.84	0.83	0.00*
Mother education	14.46	13.73	0.73	0.00*
Number of Students in class	26.68	44.39	-17.71	0.00*
Share Eligibile for Bagrut	0.97	0.87	0.10	0.00*
Psychometric Score (Total)	694.56	600.00	94.55	0.00*
Psychometric Score (quantitative)	135.57	119.50	16.07	0.00*
Psychometric Score (verbal)	131.39	114.22	17.17	0.00*
Psychometric Score (english)	136.68	120.19	16.49	0.00*
Share 5 Credits Bagrut in English	0.98	0.73	0.25	0.00*
Share 5 Credits Bagrut in Math	0.84	0.42	0.42	0.00*
Share 5 Credits Bagrut in Physics	0.51	0.31	0.20	0.00*
Bagrut Credit Points	29.34	25.87	3.47	0.00*
PHD	0.14	0.05	0.09	0.00*
Medicine	0.07	0.02	0.05	0.00*
Income 2018 (employed=1)	3.40	2.35	1.05	0.00*
Number of Observations	1,769	1,769		

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, Column (3) and (4) show the difference between the means and the corresponded p-value. All columns represents the averege level of peers (students in the same class in high scool) with respect to different outcome variables. * represents statistical significance at the 90% level.

Table A8: Treatment Estimated Effect on Girls, by Sample

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)
Main, exact cohort	-0.02	-0.01	-0.16	-0.12	-0.06	-0.05*	-0.01	-0.06*	0.00	0.02	0.04
N: 1,060	(0.03)	(0.02)	(0.12)	(0.13)	(0.13)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)
Main 1992-2000	-0.03	-0.01	-0.09	-0.05	-0.03	-0.02	0.00	-0.01	0.01	-0.01	0.00
N: 880	(0.03)	(0.02)	(0.13)	(0.15)	(0.15)	(0.03)	(0.01)	(0.03)	(0.02)	(0.04)	(0.03)
Extend all, exact cohort	-0.05*	0.00	-0.22*	-0.13	-0.10	-0.03	-0.03*	-0.04*	0.01	-0.01	-0.03
N: 1,932	(0.02)	(0.01)	(0.07)	(0.11)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)
Extend all, 1992-2000	-0.04*	-0.03	-0.09	0.00	0.00	0.01	0.01	-0.01	-0.01	-0.04*	-0.03
N: 1,438	(0.02)	(0.02)	(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 represent different sample, and each column represent different outcome variable. * represents statistical significance at the 90% level.*

Table A9: Treatment Estimated Effect on Boys, by Sample

	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)
Main, exact cohort	0.00	0.00	-0.11	-0.10	-0.18*	0.00	-0.03*	-0.03	0.00	-0.01	0.02
N: 2,094	(0.02)	(0.01)	(0.11)	(0.08)	(0.11)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Main, 1992-2000	-0.02	0.02	-0.06	-0.05	-0.03	-0.02	-0.01	-0.03	0.00	-0.02	0.01
N: 1,706	(0.02)	(0.01)	(0.13)	(0.11)	(0.13)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Extend all, exact cohort	-0.01	-0.01	-0.03	-0.04	-0.07	-0.02	0.01	0.00	0.01	0.02	0.01
N: 3,454	(0.02)	(0.01)	(0.10)	(0.13)	(0.10)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Extend all, 1992-2000	-0.02	0.01	-0.13*	-0.14	-0.14*	0.01	-0.01	0.00	0.01	-0.02	-0.01
N: 2,552	(0.01)	(0.01)	(0.08)	(0.11)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)

*Notes: This table presents the treatment effect estimated on boys, by sample. Each row represent different sample, and each column represent different outcome variable. * represents statistical significance at the 90% level.*

Appendix Table A10: STEM Studies and the Knowledge Economy

	N	Share Working in the Knowledge Economy
	(1)	(2)
No STEM	30084	0.20
Biological or Physical sciences	5943	0.43
Math, CS, Stats	8025	0.64
Engineering	15008	0.57
Math, CS, Stats and Engineering	1227	0.69
Number of Observations	497	497

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 graduetes in 2005-2013, who took the psychometric test during their 10-11th grade.