NBER WORKING PAPER SERIES

CHARITY BEGINS AT HOME (AND AT SCHOOL): EFFECTS OF RELIGION-BASED DISCRIMINATION IN EDUCATION

Victor Lavy
Edith Sand
Moses Shayo

Working Paper 24922
http://www.nber.org/papers/w24922

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2018

The expression 'Charity Begins at Home' is the most common translation to English of the 6th century Talmudic expression 'The Poor of Your Own Town Come First' (Bava Metzia 71a), which is commonly interpreted to imply that you should care for your own people before caring for others, or in psychology-economics jargon: you should show in-group bias. We thank Josh Angrist, James Fenske, Jonathan Guryan, Imran Rasul, and participants at seminars at Ben Gurion University, IFS London, Bank of Israel, IDC Herzliya, Northwestern University, University of Bonn, University of Zurich, University of Warwick, and CAGE Venice Applied Micro Economics Conference for useful comments and suggestions. We also thank the Israel's Ministry of Education and Dr. Haim Gat and Eliad Trefler for allowing restricted access to schooling data in the Ministry online protected research lab. Evgeni Rachkovski provided excellent research assistance. Lavy acknowledges financial support from the European Research Council through ERC Advanced Grant 323439. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Charity Begins at Home (and at School): Effects of Religion-Based Discrimination in Education
Victor Lavy, Edith Sand, and Moses Shayo
NBER Working Paper No. 24922
August 2018
JEL No. J24,J48,J71

ABSTRACT

Religions often preach preferential treatment of fellow believers. This paper examines whether one's religious status (secular or religious) leads one to discriminate against people with a different religious status; how this affects human capital formation; and whether this discrimination is affected by exposure to others with a different religious orientation. We develop a method of detecting individual religious status and apply it to study grading decisions on national matriculation exams in Israel's Jewish state education. Comparing grades given by religious versus secular examiners to religious versus secular students, we find evidence of in-group bias. This bias is almost entirely driven by male examiners. Exploiting bunching in the grade distribution, we are able to examine who drives this observed bias: the secular or the religious. In addition, we find that in some cases exposure at home and at work to others with different religious beliefs may attenuate the bias. These biases in grading have long-run implications since they affect students' eligibility for university admission and as a result their occupation and earnings in adulthood.

Victor Lavy
Department of Economics
University of Warwick
Coventry, CV4 7AL
United Kingdom
and Hebrew University of Jerusalem
and also NBER
v.lavy@warwick.ac.uk

Moses Shayo
Department of Economics
The Hebrew University of Jerusalem
Jerusalem 91905, Israel
mshayo@huji.ac.il

Edith Sand
Bank of Israel
P.O. Box 780
91007, Jerusalem
Israel
edith.sand@boi.org.il

A data appendix is available at http://www.nber.org/data-appendix/w24922

1. Introduction

Recent years have seen an explosion of interest in the economics of inter-group discrimination and prejudice (see Charles and Guryan 2013 and Bertrand and Duflo 2017 for reviews). This literature has largely focused on racial, ethnic, and gender-based discrimination. At the same time, the economics of religion has long focused on understanding the causes and effects of religious orientation and secularization (Barro and McCleary 2003; Gruber and Hungerman 2008; Iyer 2016). This paper inquires whether religious orientation itself can serve as a source of discrimination and examines its consequences in the context of human capital accumulation.

Even within a nominally religiously homogeneous society, people vary enormously in their level of religiousness. While in the US non-Christian religious groups represent less than 7% of the population, 24% of Americans are currently estimated to be religiously unaffiliated (Jones and Cox 2017; see also Hout, Fischer, and Chaves 2013). Furthermore, the share of the religiously unaffiliated has been growing and they tend to be overrepresented among younger cohorts.[1] Inter-group discrimination across levels of religious orientation is thus potentially widespread (and this potential may be increasing as polarization between religious and secular segments of the population grows). This is especially plausible as many religions openly preach preferential treatment of fellow believers, and sometimes harsh treatment of non-believers. However, due to data limitations, such discrimination goes largely unnoticed.

Beyond the challenges of identifying discrimination in a non-experimental setting, there is the issue that religious status is often hard for researchers to observe. We propose a method of inferring religious status based on a very meaningful and revealing choice: which school to send one's children to. This allows us to assign religious status not only to children but also to parents. Importantly, we are able to implement this method on large-scale administrative data involving professional decision makers making highly consequential decisions. Specifically, we study grading decisions in Israel's matriculation system: a centralized, country-wide system of exams that, to a significant extent, determines both a student's prospects for continuing to higher education as well as her field of study (and hence

[1] About 85% of the unaffiliated identify as secular (the majority), agnostic, or atheist. And of course even within the religiously affiliated Christian population, there is enormous diversity in denomination.

occupation). Throughout, our focus is on the Jewish population and hence we are able to isolate discrimination across levels of religiousness within a given religion.

We take advantage of six important features of this setting. First, the Israeli public school system is divided into religious and secular schools. Religious schools not only stress religious teachings but also observe various religious precepts (e.g., kosher food). Hence, virtually all religious families send their children to religious schools while the vast majority of secular families send their children to secular schools. This provides the basis of our classification of religious and secular individuals. Second, due to the centralized nature of the system, the exact same matriculation exams are taken by both religious and secular students.[2] Moreover, these exams are randomly assigned to be graded by professional examiners, such that each exam booklet is graded independently by two examiners.

Third, while the exams are anonymous, religious Jews add a special inscription at the top of the first page of every written document.[3] This in principle allows the examiner grading the exam to know whether the student is religious or not (and to potentially take that into account when grading the exam). Fourth, since the examiners are themselves teachers, we have information on their demographic characteristics. Crucially, we are able to link examiners to their children's schooling records and thereby to infer the examiners' religious status.

Fifth, we have detailed data on the grades given to each exam booklet, where the grades range from 0 to 100. Observing the entire distribution of grades allows us to exploit bunching at certain points in the distribution (e.g., a grade of 55 implies passing; 54 implies failing) in order to better understand the source of grading biases, beyond what can be learned from a difference-in-differences analysis. Finally, in addition to the grade in the state-run matriculation exam (known as the "external" grade), each student also receives an "internal" grade in each specific subject, given by her school before taking the state-run exam. This allows us to rule out spurious correlations between the grader's and the student's religious status on the one hand and the student's performance in a specific subject on the other hand.

Consider first the fundamental question: do religious and secular examiners discriminate in favor of students with the same religious status as themselves? We exploit the

[2] We do not include in the analysis exams that vary across religious and secular schools.
[3] The inscription is *BS"D*, an acronym for *Besiyata DiShmaya*, an Aramaic phrase meaning "with the help of heaven." Religious Jews write this inscription (or a variation thereof) at the top of the first page of every written document as a reminder to them that all things come from God.

random assignment of exam booklets to examiners to estimate a difference-in-differences model, allowing for systematic differences across religious status both in student ability and in examiner standards. Intuitively, we compare the mean difference in grades given to religious versus secular students by religious and secular examiners, controlling for student and subject fixed effects. Since in some subjects (e.g., math) the matriculation program includes several variants (usually varying by level of proficiency), each distinguished by a questionnaire number, we use instead a questionnaire fixed effect. Through the rest of the paper we use "questionnaire" to refer to a specific variant of the subject. Using data from over 3.5 million grades given in 112 questionnaires in the years 2010–2015, we find evidence of a small but significant tendency toward religion-based in-group bias. Importantly, the bias is driven almost entirely by male examiners: an exam grade is on average about 0.03 standard deviations higher when assigned to a male examiner of the same (rather than different) religious status as the student. This is a pattern we see in almost all of our results: female examiners exhibit little if any religion-based discrimination. The estimated bias is not driven by other student characteristics that might be correlated with student religious status. Remarkably, the bias is just as large in math and science as it is in non-STEM subjects.

The biases we uncover have significant effects on the probability of passing the exam (a prerequisite for obtaining a matriculation diploma). These effects are especially meaningful for students who come from a low-education background (i.e., both parents have 12 years of schooling or less): if the exam is assigned to two examiners with a different religious status than the student, the chances of passing are about one percentage point lower than if it is assigned to examiners with the same religious status.

A difference-in-differences analysis allows us to detect in-group bias. It does not, however, allow us to identify the source of this discriminating behavior: whether it is due to the secular or religious examiners (or both). The main difficulty is that we do not have a direct measure of the quality of the exam, and there may be systematic unobserved differences between exams written by secular and religious students. This limitation is common in studies of in-group bias in non-experimental data (e.g., Shayo and Zussman 2011; Anwar, Bayer, and Hjalmarsson 2012). Here, we propose a way to help address this limitation. Our approach is based on the existence of bunching of test scores at particular thresholds: the 55 grade required for passing and the perfect 100 grade. We can thus test whether the likelihood of just crossing the threshold is higher when the student is religious rather than secular. Importantly, we can

test this separately for religious and secular examiners. The results for the passing threshold are not conclusive: it appears that secular examiners are somewhat less likely—and religious examiners are somewhat more likely—to hike a religious student's grade from 54 to 55 or 56. Both estimates are very imprecise, however. The picture is much clearer for the 100 threshold. While male secular examiners are slightly less likely to hike grades in the 98–99 range to 100 when the student is religious, male religious examiners are between 6 and 10 percentage points *more* likely to do so when the student is religious.

Finally, we find evidence that, in line with inter-group contact theory, religious-status-based discrimination might also be affected by exposure to people from other groups: in our case, people with a different religious status. We examine several measures of exposure both at the community level (the neighborhoods where the examiners live) and at the workplace level (the schools where they teach). For male examiners—who are responsible for almost all of the observed bias—we find that in-group bias is significantly reduced when the neighborhood or school includes more of the other group. The analysis controls for examiner by neighborhood/school fixed effects, which helps address the concern that the results are driven by selection of examiners who, say, move to a different neighborhood. Nonetheless, changes in school or neighborhood composition are not random: while the results for male examiners are consistent with the contact hypothesis, they may not be causal.

The paper relates to four main strands of the literature. First, a vast literature studies racial and gender discrimination in various settings such as the labor market and several aspects of law enforcement (see Charles and Guryan 2013 and Bertrand and Duflo 2017 for reviews). We contribute to this literature in three important ways. (i) We study discrimination across a very salient but little-studied dimension, namely, religious status. (ii) We study discrimination in the school system, which can have long-term implications for professional development and lifetime earnings. (iii) We find that discrimination along religious lines is almost entirely driven by men.

Second, we contribute to the literature on economics of education where measures of teachers' grading biases are used as a measure of discrimination. Lavy (2008), Björn, Höglin, and Johannesson (2011), Hanna and Linden (2012), Cornwell, Mustard, Van Parys (2013), Burges and Greaves (2013), Diamond and Persson (2016), Botelho, Madeira, and Rangel (2015), Lavy and Sand (2015), and Terrier (2016) use the systematic difference between non-blind and blind assessment across groups as a measure of such discrimination, as was used

originally by Blank (1991) and Goldin and Rouse (2000).[4] Particularly relevant to our study is the field experiment reported by Feld, Salamanca, and Hamermesh (2015), who vary whether or not a student's name is revealed to graders in Maastricht University, and find evidence of nationality-based favoritism by Dutch and German graders.

Third, the economics of religion has long studied the effects of religiousness and secularization at both the national and individual levels (Iyer 2016 provides a recent review). At the individual level, the literature focuses on such outcomes as income, education, and health-related behavior (Gruber and Hungerman 2008; Bryan, Choi, and Karlan 2018). Our analysis provides a useful complement: while religiousness may have positive (or negative) effects relative to secularism, the cleavage itself can have important implications as it can generate prejudice and discrimination, leading to bad allocations.

Finally, we contribute to the vast literature on social identity and in-group bias. Much of this literature is based on lab experiments (in social psychology these start with Tajfel et al. 1971 and hundreds of follow-ups; in economics see Eckel and Grossman 2005, Chen and Li 2009, Klor and Shayo 2010), but a growing number of studies document in-group bias in naturally occurring data (Price and Wolfers 2010; Shayo and Zussman 2011, 2017; Hjort 2014; Fisman et al. 2017; Bar and Zussman 2017; Sandberg 2018). Relatedly, we provide evidence on inter-group contact theory (Alport 1954 and hundreds of follow-ups), which has received increasing attention from economists in recent years (see Bertrand and Duflo 2017).

The rest of the paper is structured as follows. Section 2 presents the institutional background of the Israeli matriculation exam system. Section 3 describes the data and provides descriptive statistics. Section 4 presents the empirical framework and identification strategy. Section 5 reports the results. Section 6 offers a summary and some conclusions.

2. Institutional Background

The Israeli High-School and Matriculation Exam System

[4] Lavy (2008) finds that in high schools, male students are discriminated against in all subjects. Based on evidence from primary schools in the US, Cornwell et al. (2013) found that boys who perform equally well as girls are graded less favourably by their teachers and that this gap can be largely explained by these students' non-cognitive skills. Other papers using a similar methodology examine the existence of racial discrimination: Burgess and Greaves (2013) find that in English public schools, black Caribbean and black African students are under-assessed relative to their white peers while other minority groups (such as Indian, Chinese, and Asian) are over-assessed. Botelho et al. (2015) find that black students are discriminated against relative to their white classmates in Brazilian schools. Björn et al. (2011) report a similar attitude toward students from foreign backgrounds in Swedish high schools.

Israeli post-primary education consists of middle school (grades 7–9) and high school (grades 10–12). When entering high school (tenth grade), students choose whether to enroll in the academic track leading to a matriculation certificate (*bagrut* in Hebrew – to be explained below) or in the vocational track leading to a high-school diploma.

In this paper we focus on schools in the academic track where the language of instruction is Hebrew. The vast majority of students in these schools are Jewish.[5] Importantly, these schools can belong to two distinct sectors, according to religious status. "State schools" are secular and serve the secular Jewish population. "State-religious schools" serve mainly the religious Jewish population.[6] The latter are managed and supervised by an autonomous and independent administration system within the Ministry of Education. They observe religious practices (such as kosher food), and hence are practically the only state school alternative for the religious population. These schools also emphasize religious teachings and in some of the subjects follow a different curriculum. It should be stressed however that both secular and religious state schools are public schools, funded by the state.

The matriculation certificate is a prerequisite for university admission and receiving it is an economically important educational milestone. Students complete the matriculation process by passing a series of state exams administered in tenth, eleventh, and, in greater part, twelfth grade. Students choose to be tested at various levels of proficiency: questionnaires in each subject award one to five credit units per subject, depending on difficulty. A minimum of twenty credits is required to qualify for a matriculation certificate. All students are tested in a given questionnaire on the same day. Most exams are held in the summer (mid-May to early July), and only about 15% are held in winter (January–February). Some subjects are mandatory and, for many, the most basic level of study is three credit units. At least one elective is required at an advanced level (of four or five credit units). Since religious and secular schools share the same core curriculum, they also share over half the matriculation test questionnaires.

The final matriculation score in a given questionnaire is the mean of two intermediate scores: "internal" and "external." The first is based on a school-level (internal) exam, graded at the school by the student's own teacher, before the external exam takes place. The external exam is a state-level exam produced and supervised by the Ministry of Education. These state

[5] Schools with both Jewish and non-Jewish students exist mainly in municipalities with a minority Arab population and even in these schools the proportion of non-Jewish students is very small.
[6] Ultra-Orthodox Jews have their own school system which is not part of our analysis since ultra-Orthodox schools do not include an academic track leading to a matriculation certificate. There were slightly more than one thousand Jewish high schools in 2016 (excluding ultra-Orthodox schools), of which one third were religious schools.

exams are "external" to the school because they are written and scored by an independent agency.

Importantly, each external exam booklet is graded independently by two examiners, randomly assigned by a computer algorithm. These two examiners are expert teachers who have been instructing the subject of the exam for at least several years. In order to reduce the possibility that teachers will inappropriately inflate their students' scores, the protocol eliminates the possibility of examiners grading their own students' exams. In addition, the computerized process sends all exam booklets that were distributed in a specific classroom to the same two examiners together with the seating arrangement in the classroom in order to facilitate the detection of cheating on the exams. The final external score is the average of these two examiners' evaluations.

Revealing Religiousness

The external exam booklets do not reveal a student's identity to the grader: they only include the student's ID number and school code. Nonetheless, while the grading process is anonymous, religious Jews write a special inscription—*BS"D*—at the top of every page of every written document. Since virtually everyone in Israel is aware of this practice, the religious status of the student is in essence revealed to the examiners.[7]

To validate the assumption that students from religious schools write *BS"D* on their exam booklets, we were allowed to randomly sample 442 exam booklets. The sample contains 199 booklets from a 2-credit Hebrew questionnaire exam from 2015 (100 students from religious schools and 99 students from secular schools) and 243 exam booklets from a 3-credit mathematics questionnaire exam from 2014 (119 students from religious schools and 124 students from secular schools). In 83% of the cases the religious status of students' schools coincides with a religious *BS"D* notation (86% in math and 80% in Hebrew). The inconsistent cases are mostly due to students from religious schools who do not write *BS"D* (26% in math and 39% in Hebrew), while very few students from secular schools wrote *BS"D* (3% in math

[7] Figure 1 presents as an illustration of the BS"D (בס"ד) notation, the first pages of religious students' notebooks with the BS"D (בס"ד) notation at the top of each page. Note that the pages include Hebrew, math/science and English paragraphs.

and 2% in Hebrew).[8] As noted above, an examiner grades all the exam booklets that are distributed in a specific classroom and therefore if the majority of booklets from a given classroom bear the religious *BS"D* inscription, the examiner will likely assume that the few students in the room who did not write this inscription are also religious.

3. Data and Descriptive Statistics

The data used in this study includes all matriculation questionnaires taken in the summer session by Jewish students in both the religious and secular state education system in the school years 2010–2015.[9] Since we did not have information on the matriculation exams' language, we excluded Arab students who attended Arab schools and foreign-born students from the sample as their exam booklets were most likely not written in Hebrew. The basic database of matriculation test scores in each year was merged with the student database and the school database of the relevant year. Each matriculation test score record contains information on the test: student, school, and class identifiers, grade, questionnaire number, number of credit units, scores given by the first and second examiners, and school-level ("internal") exam score. Importantly, we also have data on both examiners' identifiers. Merging the matriculation exam record of each student with the student database of the same year added further information on students' characteristics (grades, class and school assignment and school zip code, gender, ethnicity based on parents' country of birth, number of siblings, and parents' education).[10] The religious orientation of students was determined according to their schools' religious orientation by merging the data with the school file (containing each school's location, religious orientation, and whether it is a gender-segregated school).

A crucial requirement for the analysis was obtaining information on examiners. The fact that examiners have to teach the subject of the exam in high school for several years before grading matriculation exams enables us to obtain information from teachers' files for the years

[8] Appendix Table A1 presents the coefficients of balancing tests for writing *BS'D*. The dependent variable in each regression is the characteristic of the student and the explanatory variable is a dummy for religious student who wrote *BS'D* (the regression includes questionnaire FE). The first column includes all students and the second column includes religious students only. Overall, the estimates indicate that writing *BS"D* is highly correlated with the religious status of students (first column) and that writing *BS"D* among religious students (second column) is more prevalent among female students, among students with more siblings, and among students with low parental education.

[9] We have data on questionnaires given in the summer session only. Matriculation questionnaires are jointly taken by both secular and religious sectors, if the proportion of religious students that take the questionnaire is in the range [0.1, 0.9].

[10] Parents' country of birth is in general defined by fathers' country of birth. In case of missing values or Israeli-born fathers it is defined by mothers' country of birth.

2000–2015. The information on each examiner (main field of instruction, main school assignment, gender, number of children, age, education and ethnicity, school assignment and school zip code) is obtained from the teacher database of the relevant year or earlier (in case the examiner did not teach in a certain year) and merged with the school database of the same year in order to add schools' religious orientation.

The examiners' religious status is defined according to the religious status of their children's school. Specifically, in order to define the religious orientation of the examiners, we constructed a new database that defined the religious orientation of each parent who had a child enrolled in high school during 1998–2016. This new parent database was obtained by merging students' files (which contain parents' identifiers) for the years 1998–2016 with the same year's school databases containing schools' religious orientation. Parents were defined as religious if at least one of their children attended a religious school. Since we have students' files for many years (1998–2016) we were able to determine the religious status of most of the examiners in our sample according to this definition (about 85% of the examiners and 87% of the graded exam booklets). According to a series of balancing tests (see Appendix Table A2), students who were assigned to examiners who had missing values for religious status did not differ significantly in their characteristics from the other students.

We developed several measures of examiners' exposure to different environments each year at school: the proportion of religious/secular peers at school, the proportion of same-subject religious/secular peers, and the proportion of same-gender religious/secular peers. These variables were constructed at the examiner level in each year by merging the information on examiners' peers at school from teachers' files in each year with the parents' files. The teacher database contains information on all teachers in each school, including their demographic information and main fields of study. Therefore, merging it with parents' files enables us to compute for each teacher in a given year the proportion of peers at school from a religious background, the proportion of peers at school from a religious background who teach the same subject, and the proportion of peers at school from a religious background who have the same gender.

Similarly, we also computed a geographical measure of examiners' exposure to a different religious environment each year in his/her neighborhood, using the proportion of religious/secular students within the examiners' zip code. Since both students' and teachers' files contain neighborhood zip codes, we were able to characterize for each teacher's zip code

in a given year the proportion of students who attended religious schools according to students' files, and merged it with teachers' files for the relevant year.

The final merged panel dataset consists of data for six years of matriculation exams between the years 2010–2015. The dataset includes matriculation test characteristics (student, school, class, both examiners identifiers, questionnaire number, number of credits, scores given by the first and second examiners, school-level ("internal") exam score), students' characteristics (grades, class, and school assignment and school zip code, gender, ethnicity, number of siblings, and parents' education), school characteristics (location, religious orientation, and whether it is a gender-segregated school), and the information on the two examiners of each exam booklet (main field of instruction, gender, age, education and ethnicity, main school's characteristics, and peers' and neighbors' religious orientation).

Descriptive Statistics

Table 1 presents descriptive statistics at the student level, for all students and by students' religious orientation. The total number of students who took at least one "summer session" matriculation exam in Hebrew during the years 2010–2015 is 423,002 students. One-quarter of these students came from religious schools. Comparing students' background by religious orientation reveals that the proportion of girls and the number of siblings are higher among religious students (the proportion of girls is 62% versus 51% and the average number of siblings is 2.25 versus 0.9). Other characteristics are similar for both sectors. Additional statistics on students' test scores by students' religious orientation are presented in Appendix Table A3. On average, secular students have external test scores similar to those of religious students (70.5 versus 70), as well as a similar probability of passing the exam.

Table 2 presents descriptive statistics of examiners, by gender and religious orientation. There are about 2.5 thousand examiners in our sample, most of whom are female examiners (82.7%) and one-third of whom are religious examiners (33.8%). Of the religious examiners, one third are Ultra-Orthodox (11.1%) and about 10% teach at schools located in segregated religious areas (religious settlements). Except for being a bit less educated than their secular counterparts (the proportion of secular examiners with an M.A. or a Ph.D. is 67% while the corresponding proportion of religious examiners is only 57%), secular and religious examiners have similar observed characteristics. Comparing examiners' characteristics by gender reveals that female examiners are less likely to teach science (44% versus 65%), are more likely to be

Ultra-Orthodox (12% versus 5.5%), are younger (51 years old versus 55), and are less educated than their male peers (the proportion of female examiners with an M.A. or a Ph.D. is 65% versus 69%).

Descriptive statistics on examinations are presented in Appendix Tables A4 and A5. The data are based on around 4 million exam booklets, from one thousand schools. Since we have 2.5 thousand examiners, the mean number of booklets graded by each examiner is 1650 (std.=1443), and the mean number of booklets per school graded by each examiner is 12.3 (std.=5.33). The mean number of exam scripts per school graded by each examiner in each year is 9.37 (std.=7.44). This is due to the fact that all exam booklets that are distributed in a specific classroom are graded by the same examiner and the maximum number of students who are allowed to be examined in the same classroom is 20. Since all booklets that are distributed in a specific classroom are graded by two examiners, the mean number of exam booklets per school graded by the same two examiners in each year is almost the same (8.78, std.=6.453). In addition, the mean number of booklets taken by each student is 4.88 (std.=2.77) and the total number of questionnaires is 112.[11]

Appendix Table A5 presents summary statistics on the number of exams taken by religious and secular students who were graded by religious and secular examiners. The proportion of secular booklets graded by religious examiners out of the total of graded secular booklets (28.7%) is a bit lower than the proportion of religious booklets graded by religious examiners out of the total of graded religious booklets (30.5%) because some subjects are studied more extensively than others within a sector. This is the reason why the assignment of booklets to examiners is random only within a given questionnaire (as will be shown in Section 4).

In addition, summary statistics on the examiners' exposure to different religious environmental are presented in Appendix Table A6. Overall, the means and medians of secular examiners' exposure measures are lower than those of religious examiners and the variances are higher for all different types of exposure. This might be due to the fact that the proportion of secular individuals in the total population is much higher and more religious individuals live and work in secular areas than the other way around. The fact that there is a relatively high variation in examiners' exposure measures, which is also prevalent in the within-examiners'

[11] This is due to the fact that the sample includes only matriculation exams that are taken by students from both the secular and religious sectors. The mean total number of exams taken by each student is twice the mean number of exams taken by each student in the sample.

exposure measures (as reflected in the percentage of the variance of the dummies for high exposure measures that results from within-examiners' variance), enables us to test whether examiners' in-group biases are affected by their exposure to others with different religious beliefs, comparing the same questionnaire and the same examiners over time (and even the same examiners by questionnaire and by zip code/school).

4. Identification and Estimation

The main goal of the paper is to estimate religion-based in-group bias. In order to identify this in-group bias, we rely on the random assignment of students' exam booklets to examiners within a given questionnaire. We conduct a series of balancing tests in order to examine this identifying assumption. Specifically, we test whether booklets assigned to religious examiners were different from booklets assigned to secular examiners within a given questionnaire, in terms of students' characteristics and religious orientation. Table 3 presents the results of these balancing tests for all examiners (column 1), and separately for male (column 2) and female examiners (column 3). Each estimate is derived from a separate regression where the explanatory variable is the dummy for religious examiner and the dependent variables are students' characteristics (students' religious status, gender, number of siblings, father's years of education, mother's years of education, and five ethnicity indicators: parents born in Asia/Africa, Europe/America, former Soviet Union, Ethiopia, or Israel). Additionally, each regression includes questionnaire and year fixed effects as control variables. Except for one case, none of the estimated effects in Table 3 are significantly different from zero, indicating that characteristics of students whose exam booklets are assigned to religious examiners are not systematically different from those of students whose booklets are assigned to secular examiners, within a given questionnaire. These balancing tests confirm that the computer algorithm that assigns exam booklets of a given questionnaire to examiners is indeed random with respect to examiners' religious status.

We exploit the fact that students' exam booklets are randomly assigned to examiners within a given questionnaire, in order to test whether examiners grade student exam booklets differently depending on students' and examiners' religious profiles. We consider the following benchmark difference-in-differences specification:

$$(1) \qquad y_{bijqt} = \alpha_0 + \alpha_1 ReligStudent_i + \alpha_2 ReligExaminer_j$$

$$+\alpha_3 ReligStudent * ReligExaminer_{ij} + \beta_i + \gamma_q + \delta_t + \varepsilon_{bijqt}$$

y_{bijqt} is the outcome (e.g., test score) of exam booklet b, written by student i and assigned to examiner j, in questionnaire q, in year t. $ReligStudent_i$ and $ReligExaminer_j$ are indicator variables for religious student and religious examiner. The baseline specification includes questionnaire (γ_q) and year (δ_t) fixed effects. We further include student fixed effects (β_i). ε_{bijqt} is an error term clustered within examiner.[12]

Equation (1 allows for two possible differences across religious groups that do not necessarily indicate religious bias. First, it is possible that exams written by religious students have different unobserved characteristics (including, but not limited to, different quality) than those written by secular students. Thus, α_1 may be nonzero even in the absence of religious bias. Second, it is possible that religious and secular examiners have different grading standards (e.g., religious examiners may be more lenient). In other words, α_2 may be nonzero even in the absence of religious bias. Examiner religious bias is captured by α_3. This coefficient reflects a difference-in-differences: by how much religious examiners are more generous than secular examiners when grading an exam written by a religious student rather than a secular one.

5. Results

We start by estimating in-group bias using several alternative specifications. Table 4 shows baseline results. The unit of observation is an exam booklet graded by a particular examiner and the dependent variable is the (normalized) score given by that examiner. The number of observations is twice the number of exam booklets, since each booklet is graded by two different examiners.

Before estimating equation (1), columns 1 and 2 estimate, separately for religious and secular examiners, the difference in grades given to religious versus secular students. The regressions control for questionnaire and year fixed effects. Note that both religious and secular examiners give lower grades to religious students, with the difference being larger among secular examiners. Column 3 in Table 4 estimates equation (1) but for comparability shows a specification that again only includes questionnaire and year fixed effects. Religious students'

[12] Notice that while the examiner is the relevant treatment and we allow for clustering at this level, the clustering problem is not very central in our setting since the main explanatory variable – $ReligStudent * ReligExaminer$ – varies within the treatment group. Nonetheless, we allow for clustering at the examiner level to address possible within-examiner correlations (we note that the uncorrected standard errors are much smaller, for example, the uncorrected standard error in the baseline specification is 0.002 instead of the clustered standard error of 0.006).

test scores are lower by 0.05 of a standard deviation and religious examiners are marginally more generous, their mean test score being higher by 0.019 of a standard deviation. The in-group bias estimate is reported in the third row and equals 0.011, which is the difference between the two religious student indicators' estimates in the first two columns. Thus, test scores are on average higher by 1% of a standard deviation when the exam booklet is assigned to an examiner of the same religion as the student. This estimate is small and also imprecisely estimated.

In column 4 of Table 4 we add student fixed effects. The religious student indicator drops because of perfect collinearity. The religious examiner coefficient declines almost by half but the in-group bias estimate remains unchanged and becomes statistically significant. This specification will be our preferred estimated equation throughout the paper. Nonetheless, in column 5 we present a specification that includes booklet fixed effects (note that the student and year fixed effects drop out). The estimated in-group bias in this specification is positive, somewhat smaller, and much more precise: its standard error is the lowest of all specifications. This last specification captures within-booklet differences in test scores given by examiners of a different religious orientation than both types of students. Since we further on stratify the sample to different subgroups (mostly male and female examiners' subsamples) with fewer exam booklets appearing twice in each subgroup, we do not address this more demanding estimation strategy in the subsequent analysis. Finally, note that the evidence of in-group bias in Table 4 does not allow us to tell whether the source of the bias is religious examiners, secular examiners, or both. This is because we do not know what the unbiased grades of secular and religious students would be. We propose a method of addressing this issue in Section 5.3 below.

In Table 5 we present estimates based on stratified samples by gender of the examiner. Both male and female religious examiners give on average a higher test score than secular examiners (first row), though only the religious female estimate is significant. The striking results emerge in the second row: the in-group bias of male examiners is 0.030 (se=0.015), three times larger than the average effect shown in Table 4 and it is significantly different from zero. The female in-group bias is much smaller and not significantly different from zero. The same pattern shows up when we also include examiner by questionnaire fixed effects (columns

2 and 4). The male in-group bias estimate in this specification is four times larger than the female estimate.[13]

One major concern related to the interpretation of α_3 is that it might capture differential treatment by religious and secular examiners of some other student characteristic, rather than her religious status. For example, examiners might somehow be able to infer a student's ethnic background from her handwriting or style, and religious examiners might be more generous toward some ethnic group than secular examiners. If religious status is correlated with ethnicity, α_3 may pick up on this tendency rather than religion-based in-group bias. In Table 6 we present a robustness check for our suggested interpretation. The table includes nine columns, each presenting a different regression. All regressions are based on the full specification of equation (1), which includes year, questionnaire, and student fixed effects. In addition to the interaction $ReligStudent * ReligExaminer$, each regression also includes an interaction of the dummy for religious examiner with one of eight student characteristics. The eight characteristics (in order of the columns of the table) are indicators for male, mother's education, father's education, number of siblings, and ethnicity indicators (according to parents' country of birth): Israel, Europe/America, Asia/Africa, former Soviet Union, and Ethiopia). In column 9 we present results from a regression where we include all eight of these interaction terms jointly in the regression. The coefficients on these interactions are reported in the third row.

Two results stand out. First, and most remarkably, the in-group bias estimate is stable and virtually unaffected by the inclusion of the interaction of the religious examiner indicator and each of the student characteristics. Across all eight columns the in-group bias estimate is 0.010 or 0.011 and it is significantly different from zero. When all characteristics-interaction terms are jointly included (column 9), the estimate is 0.012 and significantly different from zero. The second meaningful result in the table is that 6 of the 8 additional interaction terms are not statistically significant. These results suggest that the in-group bias based on the

[13] Appendix Table A7 presents estimations of in-group biases based on raw test scores instead of standardized scores. The specifications are the same as in Table 5: for all examiners (columns 1 and 2) and for male examiners (columns 3 and 4) and for female examiners (columns 5 and 6). The magnitude and significance of the estimated in-group bias align with the results in Tables 4 and 5.

religious status of the student and the examiner does not capture omitted interaction bias of an examiner who is favorable toward any of the student characteristics.[14]

A second concern regarding the interpretation of our findings is that religious and secular examiners may grade a given exam booklet differently because they differentially like a particular feature in it, for example, the student's writing style, the student's way of reasoning, or perhaps because they agree with the views the student expresses in the exam. Naturally this is more likely when the student and the examiner share the same religious orientation. In other words, it could be that what we identify as in-group bias reflects a coincidence of taste and style shared by the student and the examiner and not religion-based discrimination by examiners. To address this concern we present in Table 7 evidence based on dividing the sample to STEM and non-STEM subjects. The latter include social studies, literature, and other humanities subjects where the examiner might be more prone to bias grades because of writing style or expressed views.[15] It can also be argued that in STEM subjects there is less scope for biased grading because the correct answer is more definitive. Panel A of Table 7 presents the estimates based on the STEM subsample and panel B presents the estimates based on all other subjects. Estimates are based on the full specification of year, questionnaire, and student fixed effects. In-group bias estimates from the two subsamples are clearly very similar, 0.012 in the STEM sample and 0.010 in the non-STEM sample, though only the former is statistically significant. The stratification of the sample by examiners' gender reveals similar patterns by gender: the estimated in-group biases of male STEM and non-STEM examiners are of comparable magnitude, though only the first is significant, which might be due to the fact that the number of male examiners in STEM subjects is almost twice that in non-STEM subjects. The estimated in-group biases of female examiners in both subgroups are much lower, and only the estimated in-group biases of female examiners in STEM subjects are significantly different from zero.

[14] Appendix Table A8 presents results from additional sensitivity tests by including two measures of religiousness at the school level: a dummy that indicates whether religious schools are gender-segregated religious schools or not, and the percentage of questionnaires per school that religious and secular students take in separate schools. These additional sensitivity test results indicate that the level of religious status of students who go to religious schools does not affect the estimated in-group bias, which provides further evidence for the commonness of writing *BS"D* by students from religious schools.

[15] Appendix Table A9 presents the estimated in-group biases by four core subjects of instruction: literature, social studies, English, and math. Each column presents estimates from a separated regression that includes a dummy for the relevant subject of instruction and its interactions with the variables of the main specification. The results indicate that in-group biases among social studies examiners are significantly higher than in-group biases in other subjects.

5.1 Implications of In-Group Bias for Final Matriculation Outcomes

In-group bias in grading behavior may have longer-term implications on students if its effect adds up to meaningful effects on the final matriculation outcomes. In particular, the composite matriculation score and probability of getting a matriculation diploma are important as the latter is a prerequisite for admission to universities and the former is a major factor in admission to selective and highly demanded fields of study such as medical school, computer engineering, etc.[16]

The composite matriculation score is an average (weighted by credit units) of all subject-specific final scores. Recall that the final score in each subject is an average of the external exam score and the internal score. The *external* score is the average of the grades given by the two examiners, which we have been analyzing thus far. The *internal* score is based on a school exam that is graded by the student's own teacher and is filed prior to the external exam taking place. Therefore, while the internal score represents an evaluation of the student in the particular subject, it should not be affected by the religious status of the examiners who are assigned to grade the external exam, and as such it can serve as a useful placebo outcome.

Table 8 presents evidence of the impact of in-group bias on the average external score and on the internal scores in each subject (the placebo outcome), as well as on the average final grade. Both internal and external scores are normalized. The last three columns estimate the effects on the probability of passing the exam. Since an external score is the average score of the two examiners, we use as a treatment measure the proportion of religious examiners for each exam booklet (which can be either zero, 0.5, or 1) times the indicator of religious student. The number of observations in these regressions is the number of exam booklets (rather than twice the number of exam booklets). All regressions include year, questionnaire, and student fixed effects. The standard errors are clustered at the student level.

Column 1 presents the estimated effect on the average external score. The in-group bias estimate, reported in the second row, is positive (0.020) and significant. When the treatment

[16] Ebenstein, Lavy, and Roth (2016) report that random transitory disturbances that affect cognitive performance during matriculation exams have permanent consequences. Exploiting variation across multiple exams taken by the same student, the study finds that transitory exposure to air pollution is associated with a significant decline in both student's performance on the exams and postsecondary educational attainment and earnings. For example, an additional point in the average matriculation score is worth between 45 and 66 shekels in monthly earnings. Relative to the average wage in their sample, this implies that each additional point is worth roughly a full one percent of monthly salary. These estimates imply that even modest declines in scores can have significant consequences on adult income. This conclusion is relevant as well to the findings we present in this paper, especially when noting the results below on the effects on students from a disadvantaged background.

indicator is equal to 0.5 (one of the two examiners is of the same religious orientation as the student), the in-group bias effect is equal to 0.01, identical to the respective estimate that we report in Table 4 (which corresponds to a treatment effect of *one* examiner having the same religious orientation as the student). By contrast, the placebo treatment effect on the internal grade (column 2) is an order of magnitude smaller, negative (-0.002), and not significantly different from zero. This result therefore confirms the absence of in-group bias, as expected for this outcome, and supports the validity of the natural experiment difference-in-differences estimate of the in-group bias that we report in Table 4.

Column 3 reports the impact of in-group bias on the final grade, which is an average of the external and internal scores. This estimate is 0.010 (se=0.004), close to the average of the estimates reported in columns 1 and 2.

Columns 4–6 in Table 8 present in-group bias effects on the likelihood of passing the exam. In column 4 the estimated effect of the treatment variable is 0.005, meaning that when the examiners and the student share the same religious status, the probability of passing a matriculation exam increases by half a percentage point and this effect is statistically significant (the mean probability of passing an exam in the sample is 89%). Column 5 presents an estimate based on a sample of students from low-education families. The estimate is 0.009, and the mean probability of passing an exam in this group is 83%. In contrast, column 6 reports the estimate for a sample of students from high-education families and it is practically zero. This is as expected since students in this sample have a much lower likelihood of being at the margin of failing or passing a matriculation exam. These estimates therefore imply that in-group bias can have distributional consequences, increasing the education gap between high and low socioeconomic status students, which later in life is likely to be reflected in higher income gaps.

5.2 In-Group Bias: Evidence from Test Score Bunching

The analysis so far has shown that, on average, an exam receives a higher grade when assigned to an examiner of the same level of religiousness as the student, and that this in-group bias is mainly driven by the male examiners. This section looks more closely at the grade distribution. Figures 2 and 3 present the distributions of religious and secular students' test scores for exam booklets graded by secular and by religious examiners, respectively. For both, we observe substantially larger mass at two points in the distribution: at 55, the passing score in a matriculation exam, and at 100, the highest score possible in these exams. This bunching can

be viewed as evidence that examiners systematically adjust grades to be just enough to pass the exam or, for the best students, to get a perfect score. In this section we examine whether there exists in-group bias in the likelihood of making such adjustments. In the next section we will use these patterns to identify who is responsible for the bias: the religious or the secular examiners.

As in our baseline regressions, we continue to allow religious examiners to systematically display more (or less) of this bunching behavior. We also allow religious students to systematically receive more (or less) of these upward adjustments. This may be due to a general bias for or against one of the groups, but in the case of the bunching at 100, it might in principle also be due to one group having a higher proportion of students who write outstanding exams that get censored at 100. However, as we will see below, religious students have the same likelihood as secular students to score 100 rather than any score in the range 90-99. Note that from the baseline regressions (columns 1–3 of Table 4), we cannot infer that religious students receive unjustified lower grades, as they may be systematically weaker. However, being more (or less) likely to receive an upward adjustment, especially at the passing threshold, might indicate general discrimination against one group, beyond any preference for one's own group.

Focusing first on the passing grade threshold, examiners may push up a grade within a close range of the passing grade and not necessarily from 54 to 55. In Table 9 we estimate a variant of equation (1) where the dependent variable is the probability of passing the exam (getting a grade higher than or equal to 55). We estimate these regressions using four different subsamples according to test scores: the estimates presented in column 1 are based on a sample that includes all exam booklets with test scores between 50 and 60; in column 2 the range is 54–60; in column 3 it is 54–57; and in column 4 it is 54–56. The estimates in each column are obtained from a separate regression that includes questionnaire fixed effects.[17] Columns 1-4 include all examiners. We find little consistent evidence of general discrimination in favor (or against) religious students (first row). The in-group bias estimates, however, are consistently positive. They are statistically significant only in the second column, partly due to the loss of precision arising from smaller sample sizes as we move to tighter ranges. These estimates imply

[17] We do not include student fixed effects in these regressions because of the small sample of students with more than one test score in these ranges.

that the probability of passing the exam increases by close to one percentage point when the examiner has the same religious status as the student.

Remarkably, the in-group bias estimates and their precision are more definitive when the sample is stratified by examiner's gender. The estimates for male examiners are presented in columns 5–8 and for female examiners in columns 9–12. The picture is very clear and consistent with the patterns in Table 5: male examiners discriminate in favor of students from their own group by increasing exam scores around the passing threshold while female examiners appear quite neutral in this respect. In terms of size, in-group bias among male examiners is particularly large when focusing on the two ranges closest to the passing threshold: the likelihood of "bumping" a student from one's religious group from 54 to 55/56 or from 54 to 55-57 is 4.3 and 3.2 percentage points, respectively. This effect is sizeable, about 5–6% of the mean passing rate in the whole sample. By contrast, the estimated in-group bias of female examiners in these two ranges is zero.

Table 10 presents in-group bias estimates at the margin of scoring 100. The table reports estimates of a linear probability model where the dependent variable equals 1 for scoring 100, based on samples restricted to exam booklets with test scores within the following ranges: 90–100, 95–100, 98–100, and 99–100. Columns 1–4 pool all examiners. Note that there is no evidence that religious students are overall more likely to receive a grade of 100 rather than any grade in the 90-100 range (first row of first column). More importantly, all four estimates of in-group bias are positive, but those derived from the first two ranges are small. The estimates based on the two ranges nearest to 100 are larger but both are still only marginally statistically different from zero. They imply that the probability of getting a score of 100 versus 98 or 99 is higher by 2.5 percentage points when the examiner has the same religious status as the student.

Columns 5–12 of Table 10 present estimates based on the separate samples of male and female examiners. Again, sharp differences emerge between the two sets of estimates of in-group bias. The male estimates are positive and significant in all four ranges but they are again largest where bunching is from 98 or 99 to 100. These estimates suggest that among male examiners the test scores at the top of the distribution are inflated sharply when the student and the examiner have the same religious orientation. In this case the likelihood of getting 100 versus 99 is higher by almost 11 percentage points. Strikingly, in-group bias estimates among the female examiners in all four ranges are zero.

Before continuing, it is important to note that the overall in-group bias is not limited to these ranges. The in-group bias estimate (based on the preferred specification reported in column 4 of Table 4) remains 0.10 (SE=0.06), even when we remove from the sample test scores in the range 55–60 and 95–100.

5.3 Identifying Who Discriminates: Secular or Religious Examiners?

The difference-in-differences estimate that we obtain for our natural experiment is a *relative* measure of in-group bias. We cannot tell whether the sources of this discriminating behavior are secular or religious examiners. The difficulty of identifying the relative contribution of religious and secular examiners to the in-group bias is due to the lack of an objective test score for each exam. For example, it may be the case that secular students do in fact perform better on exams and hence the extent to which secular examiners give them higher grades is no indication of bias, and the bias is entirely due to the religious examiners. But, of course, the reverse is also possible: exams written by religious students might not be as bad as the grades indicate and the bias might be entirely due to the secular examiners. This limitation is common in studies that attempt to identify in-group bias in naturally occurring (non-experimental) data. For example, Shayo and Zussman (2011) find evidence of in-group bias among Arab and Jewish judges in Israel, but in the absence of an objective measure of the strength of the cases, they cannot definitively determine whether the bias is driven by Jewish judges, Arab judges, or both. Similarly, Anwar Bayer and Hjalmarsson (2012) find that in Florida, the presence of a member of one's race in the jury pool for the trial entails a better outcome for the defendant, but again absent information on the relative strength of the evidence brought against white and black defendants, they cannot pin down whether the bias detected is due to black or white jurors (or both).

In the present paper, however, we propose a way to help address this limitation. Our approach is based on the evidence of bunching of test scores near the 55 and 100 scores. In particular, we examine whether the likelihood of increasing test scores above the failing grade or to the 100 score is higher among, say, religious examiners when they grade exam booklets of religious students versus secular students. Note that while secular and religious students may well write different quality exams on average, it is less likely that they systematically vary in the likelihood of writing an exam worth 99 versus 100 (or 54 versus 55). This allows us to test for discrimination separately for secular and religious examiners in these ranges.

We use the same ranges of test scores around the passing threshold and the 100 score that we defined in the previous section. In Table 11 we focus on the probability of passing the exam. The dependent variable is an indicator for scoring 55 or higher and the main explanatory variable is a dummy for religious student. We stratify the sample by secular examiners (top panel) and religious examiners (bottom). In Table 12, we present similar estimates at the margin of scoring 100.

Consider first the male examiners. Looking at the passing threshold regressions in Table 11, Columns 5-8, we note that among secular examiners (panel A), the coefficient on religious student is negative in all four columns, consistent with discrimination against religious students. However, all the estimates are imprecisely measured and, for the most part, are not statistically different from zero. At the same time, the estimated coefficients on *religious* examiners (panel B) are all positive, implying a pro-religious student bias, but again only one of the estimates is statistically different from zero. Note that the difference between the estimated pro-religious bias of the religious and secular examiners equals the in-group bias that we reported in Table 9. For example, the difference between the estimates for the 54–56 range (0.027 - (-0.016)) is equal to 0.043, the estimate reported in Table 9, which is significantly different from zero (p=0.101). The plausible conclusion here is that *both* the religious and the secular examiners contribute to the in-group bias that pushes students from own group above the failing grade.

The evidence in Table 12 regarding in-group bias toward the best students is remarkably different: the estimates in columns 5-8 are positive, high, and significant for male religious examiners, while negative, much smaller, and less significant for male secular examiners. The bias toward religious students among male religious examiners is positive and large in all four ranges, but it is highest in the 99–100 range. The probability of a score of 100 is higher by almost 10 percentage points when it is a religious student. The respective in-group bias of a male secular examiner is much lower, 0.017 (se=0.021). Clearly, the religious examiners drive most of the in-group bias at this bunching of test scores.

Next consider the female examiners. The evidence presented in columns 9–12 of Tables 11 and 12 show little evidence of religion-based in-group bias. This is true both in the passing threshold and in the upper end of the test score distribution. These results complement the evidence presented in Table 5, based on which we concluded that on average female examiners do not discriminate their grading on the basis of the religious status of students. Not only is

there no evidence of overall bias among women, but the results in both tables suggest that this is true for *both* religious and secular women. Thus, the lack of overall bias among women in Tables 5 and 7 is unlikely to be masking differences between religious and secular women (e.g., due to in-group bias in one group and out-group bias in the other).

We note that our result regarding the impartiality of women with respect to religious orientation is similar to the finding reported in Gneezy and Fershtman (2001): based on an experimental trust game in which students in Israel participated, women's trust in their game partners was not based on ethnic affiliation or on gender while men clearly discriminated in favor of men and women of Western ethnic origin (Ashkenazi) and against men and women of Eastern ethnic origin (Sephardi). Similar evidence is documented in Angerer et al. (2017) who find that girls tend to discriminate less than boys when having to allocate a fixed endowment between two other children where only one speaks the same language as the child making the allocation (see also Croson and Gneezy 2009 for a review of the literature on gender differences in social preferences).

A remaining question about the nature of the discrimination of male religious examiners is whether they increase the grades of students from their own group ("in-group love") or whether they lower the grades of students from the other group ("out-group hate"). The surplus mass at test scores 55 and 100 and the "hole" in the test score distribution at 54 and 99 suggest that male religious examiners inflate test scores of religious students and do not lower test scores of secular students.[18]

5.4 Examiners' Characteristics and In-Group Bias

In this section we briefly discuss results (shown in the appendix) on the sensitivity of the in-group bias estimates to examiners' characteristics. The evidence presented in Appendix Table A10 relates to the following characteristics: STEM subjects' examiners, examiner's age, and examiner's education (M.A. or Ph.D.). The effect of STEM subjects is very small, negative, and not significantly different from zero. This is consistent with the evidence presented in Table 7 where we estimate a similar in-group bias in STEM and non-STEM subjects. Older examiners have zero in-group bias, which suggests that the estimate presented in Table 4 reflects the behavior of younger examiners. The estimate regarding the examiner's education is perhaps

[18] Feld, Salamanca, and Hamermesh (2016) use a field experiment that assigns examiners randomly to students' examinations that did/did not contain the students' names, and find that the examiners' favoritism toward their own group, rather than discrimination against the other group, explains their estimates of relative in-group bias by nationality and by gender.

surprising, as it indicates that the average in-group bias estimate presented in Table 4 is mainly a result of the behavior of examiners with a high academic education (M.A. or Ph.D.).

Perhaps more interesting are the results on differences across religious orientation *within* the religious group. Some of the examiners in our sample are "Ultra-Orthodox" Jews (*Haredim*). Their children attend special schools that belong to an independent education system.[19] We are therefore able to distinguish Ultra-Orthodox Jewish examiners from other religious examiners by the type of school their children attend. We can thus examine whether the in-group bias of the former group is different from that of the latter group. This is an interesting distinction because there exists a major rift between the Ultra-Orthodox and the "Religious-Zionist" Jews in Israel, which may lead Ultra-Orthodox examiners not to favor (or even to disfavor) students from the latter group. The Religious-Zionist population is different from the Ultra-Orthodox. On the one hand, Religious-Zionists share common values with the Ultra-Orthodox such as dedication to the family and observance of religious holidays, dietary laws, and prayers. But, on the other hand, they have a strong commitment to the general society, secular education, and work. These values bring them closer to the secular population than to the Ultra-Orthodox.[20] The results presented in Appendix Table A11, column1, show that in-group bias of Ultra-Orthodox examiners is small and not significantly different from zero. This implies that Ultra-Orthodox examiners are neutral between religious and secular students, and it is consistent with the often expressed opinion that Ultra-Orthodox Jews do not view the Religious-Zionist Jews as "truly" religious.

5.5 Does In-Group Bias Decline When Exposure to the Out-Group Increases?

The hypothesis that intergroup contact might reduce intergroup prejudice dates back to at least the 1940s, and has been studied intensively ever since (see Pettigrew and Tropp 2006 for a review and a meta-analysis of 515 studies). The thrust of this literature suggests that, at least

[19] These schools are semi-private and receive partial funding from the government. While under the authority of the Ministry of Education, they have a deputy Minister of Education who is from an Ultra-Orthodox political party whenever such a party participates in a government coalition. While none of the students in this system are part of our analysis, some of our teachers are Ultra-Orthodox.

[20] The Religious-Zionists are also generally averse to the poverty that characterizes the Ultra-Orthodox and oppose their extremism. They also oppose the control that the Ultra-Orthodox groups have exercised over state religious institutions for over three decades. These tensions are very much alive, and have perhaps even intensified since the Israeli evacuation of settlements in the Gaza Strip in 2005, which had been predominantly populated by religious Zionists, as the ultra-Orthodox political parties were part of the coalition government at the time.

under favorable conditions, such effects do in fact exist and that they extend beyond racial and ethnic groups.[21] In this section, we examine whether religious-orientation-based bias decreases with examiners' exposure to people of different levels of religiousness at home (the neighborhood where they live) and at work (the school where they teach). The analysis in this section should be taken as suggestive, since we do not have a random assignment of peers.

In addition to studying separately in different schools, Israeli secular and religious Jews often live in separate neighborhoods within large cities or in separate localities such as *kibbutzim*, *moshavim* (farming communities), or in small towns. All teachers in religious schools are religious and only a small proportion of teachers in secular schools are religious.

We examine first the hypothesis that in-group bias varies with the extent of exposure to the other group by allowing the in-group bias estimates to be different for examiners who teach in totally segregated religious localities. Ninety percent of the Jewish settlements in the West Bank are such communities and three percent of the examiners teach in one of them. In the second column of Appendix Table A11 we augment the baseline estimation of in-group bias with interactions with an indicator for religious examiners who teach in a religious settlement in the West Bank. The main effect of the in-group bias estimate is 0.005 (se=0.006). The interaction term of the main effect term (*Religious Student x Religious Examiner*) with this indicator is 0.038 (se=0.016). Both estimates are positive but the interaction term is large and significantly different from zero. The net in-group bias of examiners from religious communities in the West Bank is 0.038 (se=0.016), about four time larger than the mean effect of 0.010. We also note that leaving both types of religious groups discussed in the table out of the sample (i.e., Ultra-Orthodox examiners and examiners who teach in segregated religious localities in the West Bank) yields an estimated in-group bias of 0.011 (se=0.0068), which is very similar to the baseline in-group bias of the entire examiner population. We should be cautious in interpreting this estimate as a net effect of "contact" because people who teach in settlements tend to have more right-wing views about the Israeli-Palestinian conflict and therefore this estimate of in-group bias may not be generalizable to all religious examiners. Below we present estimates of exposure in regular cities and towns in Israel.

[21] Alport (1954) argued that contact between groups under "optimal conditions" would reduce intergroup prejudice. These conditions include four features: equal status between the groups in the situation, common goals, intergroup cooperation, and the support of authorities, law, or custom. Pettigrew and Tropp's (2006) meta-analysis finds support for the added benefit of these conditions.

Table 13 presents results using four different definitions of exposure, measured in two environments: the neighborhood in which one lives and the school in which one works. Exposure is measured as a dummy variable indicating an above-median proportion of neighbors or peers in the environment with a different religious status (see Appendix Table A6 for descriptive statistics). In Appendix Table A12 we report results when exposure is measured as the proportion of neighbors or peers with a different religious orientation. We start, in panel A, with neighbors within the examiner's home zip code. The next three panels examine exposure to peers (other teachers) at school. Panel B looks at the overall proportion of peers with a different religious orientation at school, whereas panels C and D look at peers at school with the same gender or who teach the same subject.

The regressions include year and student fixed effects as well as, importantly, examiner by environment (zip code or school) by questionnaire fixed effects. Thus for example, in panel A the interaction picks up the variation in in-group bias for a given examiner living in the same neighborhood, whose neighborhood's religious composition changed over time.

The estimates presented in the first column of Table 13 are based on the full sample and show no clear pattern for the association between exposure and bias. However, the estimates for male examiners (column 2) suggest that in-group bias declines sharply when examiners encounter high numbers of the other group in their neighborhood. To some extent this is also the case for peers at school, especially when the peer group includes teachers who teach the same subject or are of the same gender. In panel A the main effect of in-group bias is 0.064 (se=0.021) and the interaction estimate when the proportion of "others" in the neighborhood is above the median is -0.074 (se=0.030). In other words, in-group bias is positive and large when the examiner is not highly exposed to neighbors with a different religious orientation the, but drops to zero when the examiner is highly exposed to the other group in the neighborhood.[22]

Male in-group bias is also associated with changes in exposure to "others" at work, especially to teachers who teach the same subject or are of the same gender. In the second column of panel C, for example, the main in-group bias estimate is 0.050 (se=0.020) and the interaction term with high exposure to same-subject teachers is -0.050 (se=0.032); thus, they offset each other. Similarly, in panel D, the main in-group bias estimate is 0.052 (se=0.022)

[22] The estimates in panel A suggest that at a very high level of exposure, the in-group bias estimate even reverses sign, meaning that an examiner might even show some out-group bias.

and the interaction term with high exposure to same-subject teachers is -0.050 (se=0.030); thus, they too offset each other.

For female examiners, the estimates in column 3 show an interesting pattern. The main in-group bias in all four panels is small and not significantly different from zero. However, in-group bias appears to emerge among female examiners when they are in the minority in terms of religious orientation at school, and in particular among female teachers at school. In both cases the in-group bias is positive and significant, around 0.021 with a t-statistic of about 2. This is inconsistent with a simple version of the contact hypothesis that ignores the importance of the conditions under which contact takes place.

6. Conclusions

Religious doctrine often favors believers over non-believers. While secularization—and its opposite, resacralization—have drawn enormous attention, the economic effects of religion-based discrimination have gone largely unnoticed. Using data from Israel's high-stakes matriculation exams we are able to identify the religious status of both students and examiners, and thus study discrimination across religious and secular groups within the same religion.

We have three main findings. First, we document the existence of significant in-group bias in grading decisions. This bias is detectable among professional graders who are making highly consequential decisions. Second, we find that the bias is almost entirely driven by male examiners: female examiners show little if any bias. One possibility is that females are in general less prone to group-based behavior, across different cultural, situational, and contextual domains (Sidanius et al. 2000). Another possibility, however, is that this result is more specific to religious discrimination. Religion and its precepts are possibly more salient for (religious) Jewish men—who are required to pray three times each day, and to engage in a lifelong study of religious teachings—than for women. This brings us to our third result. Using bunching in the grading distribution we find evidence that bias, at least at the top of the distribution, is largely driven by religious examiners. Male religious examiners are six to ten percentage points more likely to bump a grade to 100 when the exam is written by a religious student, while male secular examiners are between one and three percentage points less likely to do so when grading a religious student.

Such biases can have significant long-term implications for the allocation of talent and human capital formation. However, we do find suggestive evidence that contact across

religious and secular groups may attenuate these biases. Even though our study looks within a given religion and emphasizes the implications of a rift between secular and religious groups, it sheds light on the potential consequences of conflict across different religious groups. Such heterogeneity surfaced recently in many European countries where immigration flows brought groups with a different religion (and probably also different degree of religiousness) than the native population.

7. References

Allport, G. W. (1954). *The Nature of Prejudice*. Boston: Addison-Wesley.

Angerer, S., E. G. Dutcher, D. Glätzle-Rützler, P. Lergetporer, and M. Sutter (2017). "Gender Differences in Discrimination Emerge Early in Life: Evidence from Primary School Children in a Bilingual City." *Economics Letters*, 152, 15–18.

Anwar, S., P. Bayer, and R. Hjalmarsson (2012). "The Impact of Jury Race in Criminal Trials." *Quarterly Journal of Economics*, 127(2), 1017–1055.

Bar, R. and A. Zussman (2017). "Customer Discrimination: Evidence from Israel." *Journal of Labor Economics*, 35 (4), 1031–1059.

Barro, R. J. and R. M. McCleary (2003). "Religion and Economic Growth across Countries." *American Sociological Review*, 68(5), 760–781.

Bertrand, M. and E. Duflo (2017). "Field Experiments on Discrimination." In E. Duflo and A. Banerjee (eds.), *Handbook of Economic Field Experiments*, Volume 1 (pp. 309–393). Amsterdam: North Holland.

Bryan, G., J. J. Choi, and D. Karlan (2018). "Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes." *NBER Working Paper* 24278.

Burgess, S. and E. Greaves (2013). "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities." *Journal of Labor Economics*, 31, 535–576.

Charles, K. K. and J. Guryan (2013). "Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots." *The Economic Journal*, 123(572), 417–432.

Chen, Y. and S. X. Li. (2009). "Group Identity and Social Preferences." *American Economic Review*, 99(1), 431–457.

Cox, D. and R. P. Jones (2017), "America's Changing Religious Identity." Public Religion Research Institute (PRRI).

Croson, R. and U. Gneezy (2009). "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2), 1–27.

Diamond, R. and P. Persson (2016). "The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests." NBER Working Paper 22207.

Ebenstein, A., V. Lavy and Sefi Roth (2016). "The Long Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution." *American Economic Journal: Applied Economics*, 8(4): 36–65.

Eckel, C. C. and P. J. Grossman (2005). "Managing Diversity by Creating Team Identity." *Journal of Economic Behavior & Organization*, 58(3), 371–392.

Feld, J., N. Salamanca, and D. S. Hamermesh (2016). "Endophilia or Exophobia: Beyond Discrimination." *The Economic Journal*, 126(594), 1503–1527.

Fisman, R., Paravisini, D., and V. Vig (2017). "Cultural Proximity and Loan Outcomes." *American Economic Review,* 107(2), 457–492.

Gneezy, U. and H. Fershtman (2001). "Discrimination in a Segmented Society: An Experimental Approach." *The Quarterly Journal of Economics,* 116(1), 351–377.

Gruber, J. and D. Hungerman (2008). "The Church vs the Mall: What Happens When Religion Faces Increased Secular Competition?" *Quarterly Journal of Economics,* 123(2), 831–862.

Hanna, R. N., and L. L. Linden (2012). "Discrimination in Grading." *American Economic Journal: Economic Policy*, 4(4), 146–168.

Hjort, J. (2014). "Ethnic Divisions and Production in Firms. *The Quarterly Journal of Economics*, 129(4),1899–1946.

Hout, M., C. S. Fischer, and M. A. Chaves (2013). "More American Have No Religious Preference: Key Findings from the 2012 General Social Survey." *Institute for the Study of Societal Issues*.

Iyer, S. (2016). "The New Economics of Religion." *Journal of Economic Literature*, 54(2), 395–441.

Lavy, V. (2008). "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics,* 92, 2083–2105.

Lavy, V. and E. Sand. "On the Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases." Forthcoming, *Journal of Public Economics*.

Pettigrew, T. F. and L. R. Tropp (2006). "A Meta-analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology*, 90, 751–783.

Price, J. and J. Wolfers (2010). "Racial Discrimination among NBA Referees." *Quarterly Journal of Economics,* 125(4), 1859–1887.

Sandberg, A. (2018) “Competing Identities: A Field Study of In-group Bias Among Professional Evaluators.” The Economic Journal, 10.1111/ecoj.12513,

Shayo, M. and A. Zussman (2011). “Judicial Ingroup Bias in the Shadow of Terrorism.” *Quarterly Journal of Economics*, 126(3), 1447–1484.

Shayo, M. and A. Zussman (2017). “Conflict and the Persistence of Ethnic Bias.” *American Economic Journal: Applied Economics*, 9 (4), 137–165.

Sidanius, J., S. Levin, J. Liu, and F. Pratto (2000). “Social Dominance Orientation, Anti-egalitarianism and the Political Psychology of Gender: An Extension and Cross-cultural Replication.” *European Journal of Social Psychology*, 30(1), 41–67.

Tajfel, H., M. G. Billig, R. P. Bundy, and C. Flament (1971). “Social Categorization and Intergroup Behavior.” *European Journal of Social Psychology*, 1, 149–178.

Terrier C. (2016), “Boys Lag Behind: How Teachers’ Gender Biases Affect Student Achievement.” *SEII Working Paper* 2016.07.

Table 1: Summary Statistics of Students' Characteristics

	All Students	Religious Students	Secular Students
	(1)	(2)	(3)
Proportion of Boys		0.376	0.492
		(0.484)	(0.499)
Mean Father's Education	12.525	12.568	12.402
	(4.693)	(5.339)	(4.536)
Mean Mother's Education	12.899	12.134	13.066
	(4.208)	(5.173)	(3.893)
Mean Number of Siblings	1.341	2.250	0.943
	(1.475)	(2.051)	(0.978)
Proportion of Asian/African Ethnicity	0.123	0.152	0.112
	(0.329)	(0.359)	(0.316)
Proportion of European/American Ethnicity	0.104	0.140	0.092
	(0.305)	(0.347)	(0.288)
Proportion of Israeli Ethnicity	0.641	0.622	0.646
	(0.480)	(0.484)	(0.478)
Proportion of Former Soviet Union	0.112	0.056	0.131
	(0.315)	(0.232)	(0.337)
Proportion of Religious Students	0.257		
	(0.437)		
Number of Students	423,002	108,594	314,408

Notes: The sample includes students in Jewish schools who were born in Israel and took at least one matriculation test in an identical questionnaire for both the religious and secular sectors. Religious students are defined by the degree of religiousness of the students' school (dummy=1 if the school is a religious school). Standard deviations are reported in parentheses.

Table 2: Summary Statistics of Examiners' Characteristics, by Gender

	All Examiners	Religious Examiners	Secular Examiners	Male Examiners	Female Examiners
	(1)	(2)	(3)	(4)	(5)
Proportion of Male Examiners	0.173	0.167	0.175	1.000	0.000
	(0.378)	(0.374)	(0.378)	(0.000)	(0.000)
Proportion of Science Examiners	0.478	0.477	0.475	0.650	0.440
	(0.500)	(0.499)	(0.499)	(0.478)	(0.497)
Proportion of Religious Examiners	0.338	1.000	0.000	0.336	0.338
	(0.473)	(0.000)	(0.000)	(0.473)	(0.473)
Proportion of Ultra-Orthodox Examiners	0.111	0.374	0.000	0.055	0.122
	(0.315)	(0.484)	(0.000)	(0.228)	(0.327)
Proportion of Examiners who Teach in Schools Located in Segregated Religious Areas	0.030	0.128	0.000	0.026	0.030
	(0.169)	(0.334)	(0.000)	(0.159)	(0.170)
Examiners' Age	51.880	49.906	51.374	54.832	51.260
	(9.741)	(10.574)	(9.402)	(10.509)	(9.460)
Proportion of Highly Educated Examiners	0.656	0.571	0.670	0.689	0.649
	(0.464)	(0.500)	(0.470)	(0.456)	(0.466)
Proportion of Examiners of Asian/African Ethnicity	0.050	0.050	0.053	0.082	0.044
	(0.218)	(0.218)	(0.223)	(0.275)	(0.205)
Proportion of Examiners of European/American Ethnicity	0.120	0.163	0.106	0.119	0.120
	(0.325)	(0.369)	(0.308)	(0.324)	(0.325)
Proportion of Examiners from Former Soviet Union	0.108	0.057	0.126	0.159	0.098
	(0.310)	(0.232)	(0.332)	(0.366)	(0.297)
Proportion of Examiners of Israeli Ethnicity	0.720	0.728	0.713	0.638	0.736
	(0.449)	(0.445)	(0.452)	(0.481)	(0.440)
Number of Examiners	2,508	715	1,400	431	2,064

Notes: Religious examiners are defined by the level of religiousness of their children school (dummy=1 if the school is a religious school). Ultra-Orthodox religious examiners are also defined by the religious status of their children school (dummy=1 if the school is an Ultra-Orthodox religious school). Highly educated examiners are examiners with an M.A. or a Ph.D. Note that some examiners have missing values for religious status or for gender. Standard deviations are reported in parentheses.

Table 3: Balancing Tests for the Assignments of Students' Tests to Examiners, by Examiners' Gender

	All Examiners	Male Examiners	Female Examiners
	(1)	(2)	(3)
Gender (Boy = 1)	-0.0020	-0.0020	-0.0020
	(0.0010)	0.0030	0.0010
Number of siblings	-0.0010	0.0060	-0.0010
	(0.0030)	0.0080	0.0040
Father's years of schooling	0.0090	-0.0070	0.0100
	(0.0130)	0.0260	0.0140
Mother's years of schooling	0.0010	0.0001	0.0010
	(0.0120)	0.0250	0.0130
Asian/African Ethnicity	0.0000	-0.0003	0.0000
	(0.0004)	0.0010	0.0005
European/American Ethnicity	0.0010	-0.0010	0.0010
	(0.0005)	0.0010	0.0010
Israeli Ethnicity	-0.0010	0.0010	-0.002*
	(0.0010)	0.0020	0.0010
Former Soviet Union Ethnicity	0.0010	0.0003	0.0010
	(0.0010)	0.0020	0.0010
Ethiopian Ethnicity	0.0000	-0.0004	0.0001
	(0.0002)	0.0005	0.0002
Religious Student	-0.0020	-0.0020	-0.0020
	(0.0020)	0.0050	0.0020
N	3,590,116	508,324	3,081,792

Notes: The dependent variable in each regression is the characteristic of the student and the explanatory variable is a dummy for religious examiner. Column 1 includes all examiners, column 2 includes only male examiners, and column 3 includes only female examiners. Each regression controls for questionnaire and year fixed effects. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 4: Estimated Effect of In-Group Biases of Examiners on Test Scores

	Religious Examiners	**Secular Examiners**	**All Examiners**		
	Questionnaire and Year Fixed Effects	Questionnaire and Year Fixed Effects	Questionnaire and Year Fixed Effects	Questionnaire, Year, and Student Fixed Effects	Exam Booklet Fixed Effects
	(1)	(2)	(3)	(4)	(5)
Religious Student	-0.041***	-0.051***	-0.051***		
	(0.010)	(0.007)	(0.006)		
Religious Examiner			0.019***	0.011**	0.014***
			(0.007)	(0.005)	(0.004)
Religious Student x Religious Examiner			0.011	0.010*	0.008***
			(0.012)	(0.006)	(0.003)
Number of Observations	1,201,625	2,388,491	3,590,116	3,590,116	3,590,116

Notes: The first two columns of the table present the difference in grades given to religious and secular students, separately by religious (column 1) and secular examiners (column 2). The estimates of the religious student indicator are from a specification that includes questionnaire and year fixed effects. The other columns present the difference-in-differences in-group bias estimates, from different specifications: in column 3 the specification includes only questionnaire and year fixed effects; in column 4 the specification includes also student fixed effects; and the last specification includes only exam booklet fixed effects. The number of observations is twice the number of exam booklets, since each exam booklet appears twice (once for each examiner). Dependent variables are standardized scores. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 5: Estimated Effect of In-Group Biases of Examiners on Test Scores, by Examiners' Gender

	Male Examiners		Female Examiners	
	Questionnaire, Year, and Student Fixed Effects	Year, Student, and Examiner by Questionnaire Fixed Effects	Questionnaire, Year, and Student Fixed Effects	Year, Student, and Examiner by Questionnaire Fixed Effects
	(1)	(2)	(3)	(4)
Religious Examiner	0.017		0.011*	
	(0.013)		(0.0060)	
Religious Student x Religious Examiner	0.030**	0.027*	0.010	0.006
	(0.015)	(0.015)	(0.006)	(0.006)
Number of Observations	508,324	508,324	3,081,792	3,081,792

Notes: The table presents the estimated in-group bias of examiners according to two specifications, separately for male and female examiners. The first specification in columns 1 and 3 includes year, questionnaire, and student fixed effects; the second specification in columns 2 and 4 includes also examiner by questionnaire fixed effects instead of questionnaire fixed effects. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 6: Sensitivity of the Results to Students' Characteristics

	Boy	Highly Educated Mother	Highly Educated Father	Highly Number of Siblings	Israeli Ethnicity	European/American Ethnicity	Asian/African Ethnicity	Former Soviet Union	All Characteristics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Religious Examiner	0.006	0.007	0.010	0.011**	0.013**	0.010*	0.012**	0.009*	0.015
	(0.006)	(0.008)	(0.008)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.014)
Religious Student x Religious Examiner	0.011*	0.010*	0.010*	0.011**	0.010*	0.010*	0.011*	0.011*	0.012**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Student Characteristic x Religious Examiner	0.009	0.000	0.000	-0.001	-0.003	0.006	-0.010***	0.009*	
	(0.008)	(0.001)	(0.000)	(0.001)	(0.003)	(0.006)	(0.004)	(0.005)	
Number of Observations	3,590,116	3,547,780	3,541,390	3,551,430	3,590,116	3,590,116	3,590,116	3,590,116	3,496,361

Notes: The table presents the sensitivity of the in-group bias estimate to students' characteristics. All columns present the results from separated regressions based on the preferred specification (which includes year, questionnaire, and student fixed effects), where each regression additionally includes the interaction between the dummy for religious examiner and a different student characteristic. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 7: Estimated In-Group Biases of Examiners in STEM and Non-STEM Subjects, by Examiners' Gender

	All Examiners	Male Examiners	Female Examiners
	(1)	(2)	(3)
A. STEM Test Scores			
Religious Examiner	0.007	0.019	0.004
	(0.007)	(0.016)	(0.008)
Religious Student x Religious Examiner	0.012*	0.033*	0.013*
	(0.007)	(0.018)	(0.008)
Number of Observations	1,652,315	320,764	1,331,551
B. Non-STEM Test Scores			
Religious Examiner	0.01**	-0.01**	0.02**
	(0.007)	(0.032)	(0.007)
Religious Student x Religious Examiner	0.010	0.048	0.006
	(0.008)	(0.052)	(0.009)
Number of Observations	1,937,801	187,560	1,750,241

Notes: The table presents the estimated in-group bias of examiners, separately for STEM (panel A) and non-STEM (panel B) subjects. All columns present the results from separated regressions based on the preferred specification (which includes year, questionnaire and student fixed effects). In the first column all examiners are included; in the second and third columns the sample is stratified by examiner's gender. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 8: Estimated Effect of In-Group Biases of Examiners on Related Exam Outcomes

	Average External Exam Grade	Internal Exam Grade: Placebo Test	Average Final Exam Grade	Probability of Passing the Exam		
				All Students	Students with Low Parental Education	Students with High Parental Education
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of Religious Examiners	0.005**	0.000	0.003	-0.002*	-0.003*	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Religious Student x Proportion of Religious Examiners	0.020***	-0.002	0.010**	0.005***	0.009**	0.001
	(0.005)	(0.005)	(0.004)	(0.002)	(0.004)	(0.002)
Number of Observations	1,565,252	1,535,550	1,535,550	1,535,556	627,818	883,892

Notes: The table presents the estimated effect of in-group bias of examiners on additional outcomes: 1) the average external exam grade (the average of the two examiners' normalized scores); 2) the normalized internal exams, which are exams examined by students' school teachers; 3) the final exam score (the average of the external and internal exams' normalized scores); 4) probability of passing the exam (if final grade >=55); 5) probability of passing the exam from a subsample of students with low parental education (low parental education is equal to one if both parents have 12 or less years of schooling); 6) and the probability of passing the exam from a subsample of students with high parental education. The proportion of religious examiners is measured in each exam booklet. The number of observations is the number of exam takers, since each exam appears only once. All columns present the results from separated regressions based on the preferred specification (which includes year, questionnaire, and student fixed effects). Standard errors are corrected for clustering at the student level and are presented in parentheses.

Table 9: Estimated Effect of In-Group Biases of Examiners on the Probability of Passing the Exam, by Examiners' Gender

	All Examiners				Male Examiners				Female Examiners			
	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Religious Student	-0.009***	-0.001	0.001	-0.001	-0.012*	-0.006	-0.007	-0.016	-0.008***	0.000	0.003	0.002
	(0.0020)	(0.002)	(0.004)	(0.005)	(0.007)	(0.060)	(0.011)	(0.014)	(0.003)	(0.002)	(0.004)	(0.006)
Religious Examiner	-0.008**	0.002	0.001	-0.004	-0.017	-0.010	-0.028	-0.054*	-0.007**	0.004	0.005	0.003
	(0.003)	(0.003)	(0.007)	(0.009)	(0.011)	(0.012)	(0.022)	(0.028)	(0.004)	(0.004)	(0.007)	(0.009)
Religious Student x Religious Examiner	0.006	0.007*	0.010	0.009	0.012	0.017*	0.032*	0.043	0.005	0.005	0.006	0.002
	(0.004)	(0.004)	(0.007)	(0.009)	(0.012)	(0.010)	(0.018)	(0.026)	(0.004)	(0.004)	(0.007)	(0.010)
Number of Observations	371,094	255,779	127,998	84,110	51,394	42,279	18,070	11,722	319,700	220,236	109,028	72,388

Notes: The dependent variable is the probability of passing the exam (if score>=55). The coefficients in each column are from separated regressions that include questionnaire fixed effects, for four different subsamples: in the first column the subsample includes all tests with scores between 50 and 60; in the second column the subsample includes all tests with scores between 54 and 60; in the third column the subsample includes all tests with scores between 54 and 57; and in the last column the subsample includes all tests with scores between 54 and 56. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 10: Estimated Effect of In-Group Biases of Examiners on the Probability of Scoring 100, by Examiners' Gender

	All Examiners				Male Examiners				Female Examiners			
	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Religious Student	-0.001	-0.008	-0.023**	-0.002	-0.006	-0.011	-0.035**	-0.017	0.001	-0.006	-0.019	0.002
	(0.0030)	(0.006)	(0.011)	(0.013)	(0.0050)	(0.010)	(0.017)	(0.021)	(0.0040)	(0.007)	(0.013)	(0.014)
Religious Examiner	0.000	-0.005	-0.022	-0.025	-0.001	-0.015	-0.050	-0.070	-0.001	-0.004	-0.016	-0.012
	(0.005)	(0.009)	(0.016)	(0.018)	(0.016)	(0.029)	(0.049)	(0.053)	(0.005)	(0.009)	(0.015)	(0.018)
Religious Student x Religious Examiner	0.005	0.009	0.026	0.025	0.029***	0.046**	0.098***	0.109***	0.001	0.000	0.006	0.000
	(0.005)	(0.0090)	(0.016)	(0.018)	(0.011)	(0.0190)	(0.032)	(0.036)	(0.005)	(0.0090)	(0.018)	(0.019)
Number of Observations	557,641	243,970	105,919	68,332	89,101	42,158	20,001	13,894	468,540	201,812	85,918	54,438

Notes: The dependent variable is the probability of scoring 100 on the exam. The coefficients in each column are from separated regressions that include questionnaire fixed effects, for four different subsamples: in the first column the subsample includes all tests with scores between 90 and 100; in the second column the subsample includes all tests with scores between 95 and 100; in the third column the subsample includes all tests with scores between 98 and 100; and in the last column the subsample includes all tests with scores between 99 and 100. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 11: Estimated Effect of In-Group Biases of Examiners on the Probability of Passing the Exam, by Examiners' Gender and Religious Status

	All Examiners				Male Examiners				Female Examiners			
	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56	Test Scores between 60–50	Test Scores between 54–60	Test Scores between 54–57	Test Scores between 54–56
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Secular Examiners												
Religious Student	-0.009***	-0.001	0.001	-0.001	-0.012*	-0.006	-0.007	-0.016	-0.008***	0.003	0.003	0.002
	(0.002)	(0.002)	(0.004)	(0.005)	(0.007)	(0.006)	(0.011)	(0.014)	(0.003)	(0.004)	(0.004)	(0.006)
Number of Observations	250,814	173,779	87,446	57,752	33,476	23,929	11,996	7,862	217,338	150,487	75,450	49,890
B. Religious Examiners												
Religious Student	-0.003	0.006**	0.011**	0.008	0.000	0.011	0.024*	0.027	-0.003	0.005*	0.009	0.004
	(0.003)	(0.003)	(0.006)	(0.008)	(0.010)	(0.008)	(0.015)	(0.022)	(0.004)	(0.003)	(0.006)	(0.008)
Number of Observations	120,280	82,000	40,552	26,358	17,918	12,251	6,074	3,860	102,362	64,740	34,478	22,498

Notes: See Table 9. The coefficients in each column are from separated regressions that include a dummy for religious student and questionnaire fixed effects. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 12: Estimated Effect of In-Group Biases of Examiners on the Probability of Scoring 100, by Examiners' Gender

	All Examiners				Male Examiners				Female Examiners			
	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100	Test Scores between 90–100	Test Scores between 95–100	Test Scores between 98–100	Test Scores between 99–100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Secular Examiners												
Religious Student	-0.001	-0.008	-0.023**	-0.002	-0.006	-0.011	-0.034**	-0.017	0.001	-0.006	-0.019	0.002
	(0.003)	(0.006)	(0.011)	(0.013)	(0.005)	(0.010)	(0.017)	(0.021)	(0.004)	(0.007)	(0.013)	(0.014)
Number of Observations	361,929	156,690	67,505	43,667	55,150	25,384	11,863	8,233	306,779	131,306	55,642	35,434
B. Religious Examiners												
Religious Student	0.004	0.000	0.003	0.023*	0.023**	0.035**	0.065**	0.096***	0.001	-0.007	-0.013	0.000
	(0.003)	(0.006)	(0.012)	(0.013)	(0.009)	(0.016)	(0.027)	(0.031)	(0.003)	(0.007)	(0.012)	(0.013)
Number of Observations	195,712	87,280	38,414	24,665	33,951	16,774	8,138	5,661	161,761	70,506	30,276	19,004

Notes: See Table 10. The coefficients in each column are from separated regressions that include a dummy for religious student and questionnaire fixed effects. Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 13: Estimated Effect of High Exposure to a Different Religious Environment on In-Group Biases of Examiners, by Examiners' Gender

	All Examiners	Male Examiners	Female Examiners
	(1)	(2)	(3)
A. High Exposure to Neighbors with a Different Religious Orientation than that of the Examiner			
Religious Student x Religious Examiners	0.011	0.064***	0.004
	(0.008)	(0.021)	(0.009)
Religious Student x Religious Examiners x Dummy for Exposure to a High Proportion of Neighbors with a Different Religious Orientation	-0.007	-0.074**	0.004
	(0.012)	(0.030)	(0.013)
Observations	3,505,201	497,811	3,007,390
B. High Exposure to Peers at School with a Different Religious Orientation than that of the Examiner			
Religious Student x Religious Examiners	0.002	0.043**	-0.005
	(0.008)	(0.021)	(0.009)
Religious Student x Religious Examiners x Dummy for Exposure to a High Proportion of Peers at School with a Different Religious Orientation	0.011	-0.032	0.021*
	(0.011)	(0.029)	(0.012)
Observations	3,590,116	508,324	3,081,792

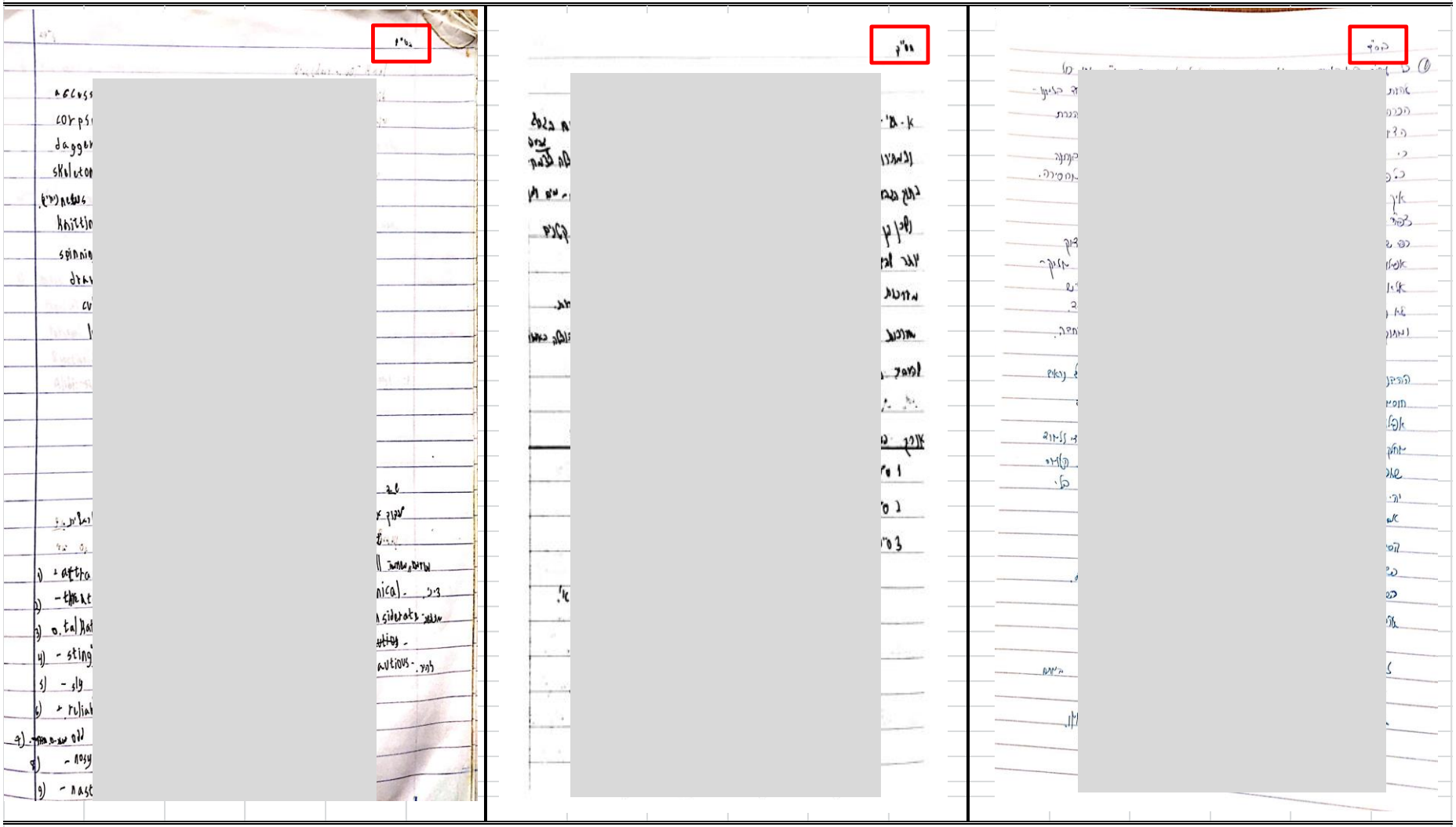
Notes: The coefficients in each column and panel are from separated regressions that include a dummy for different types of exposure and its interactions with the variables of the main specification. Each regression additionally includes year and student fixed effects and examiner by questionnaire and zip code/school fixed effects. The proportion of neighbours with a different religious orientation is based on the proportion of religious students in the examiner's zip code in each year. The proportion of peers at school with a different religious orientation is based on the proportion of peer teachers at school in each year. The dummy variable for high exposure equals one if the proportion of the examiner's neighbours or peers is higher than the median of each group (by religious status and gender). Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

Table 13: Estimated Effect of High Exposure to a Different Religious Environment on In-Group Biases of Examiners, by Examiners' Gender—Continued

	All Examiners	Male Examiners	Female Examiners
	(1)	(2)	(3)
C. High Exposure to Peers at School with a Different Religious Orientation than that of the Examiner but who Teach the Same Subject			
Religious Student x Religious Examiners	0.012*	0.050**	0.008
	(0.007)	(0.020)	(0.007)
Religious Student x Religious Examiners x Dummy for Exposure to a High Proportion of Same-Subject Peers at School with a Different Religious Orientation	-0.018	-0.050	-0.014
	(0.012)	-(0.032)	(0.013)
Observations	3,485,422	498,185	2,987,237
D. High Exposure to Peers with a Different Religious Orientation than that of the Examiner but of the Same Gender at School			
Religious Student x Religious Examiners	0.002	0.052**	-0.006
	(0.008)	(0.022)	(0.009)
Religious Student x Religious Examiners x Dummy for Exposure to a High Proportion of Same-Gender Peers at School with a Different Religious Orientation	0.010	-0.050*	0.022*
	(0.011)	(0.030)	(0.012)
Observations	3,590,116	508,324	3,081,792

Notes: The coefficients in each column and panel are from separated regressions that include a dummy for religious student and a dummy for religious teacher and their interactions with the different types of exposure variables. Each regression additionally includes year and student fixed effects and examiner by questionnaire and zip code/school fixed effects. The proportion of neighbors with a different religious orientation is based on the proportion of religious students in the examiner's zip code in each year. The proportion of peers at school with a different religious orientation is based on the proportion of peer teachers at school in each year. The dummy variable equals one if the proportion of the examiner's neighbors or peers is higher than the median of each group (by religious status and gender). Standard errors are corrected for clustering at the examiner level and are presented in parentheses.

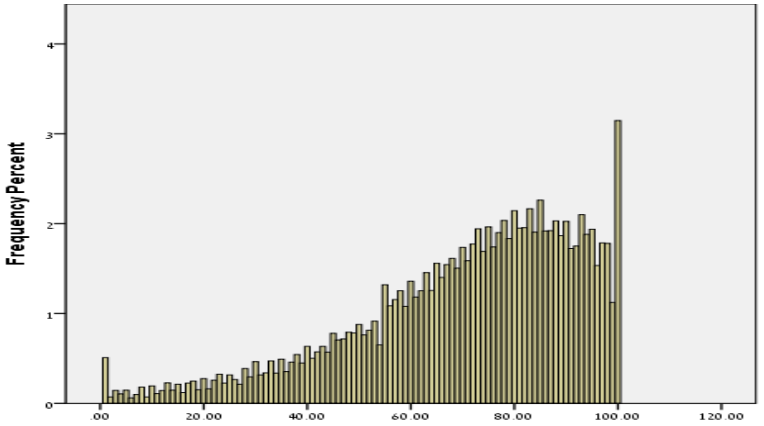
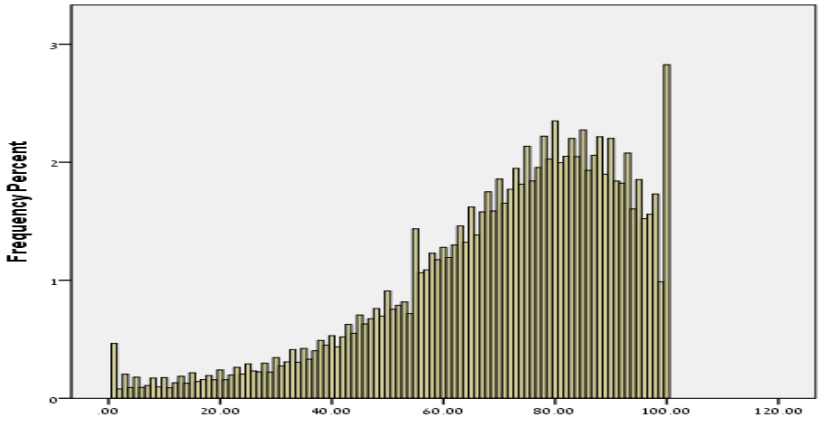
Figure 1: Sample of Religious Students' Notebooks



Notes: The inscription is `BS"D' (בס"ד in Hebrew)-- acronym for Besiyata Dishmaya, an Aramaic phrase, meaning "with the help of Heaven". Religious Jews write this notation at the top of every page in a written document as a reminder to them that all comes from God.

Figure 2: The Distributions of Religious Examiners Scores, by Students' Religiosity and Examiners' Gender

A. Male Examiners

Religious students	**Secular students**
	

B. Female Examiners

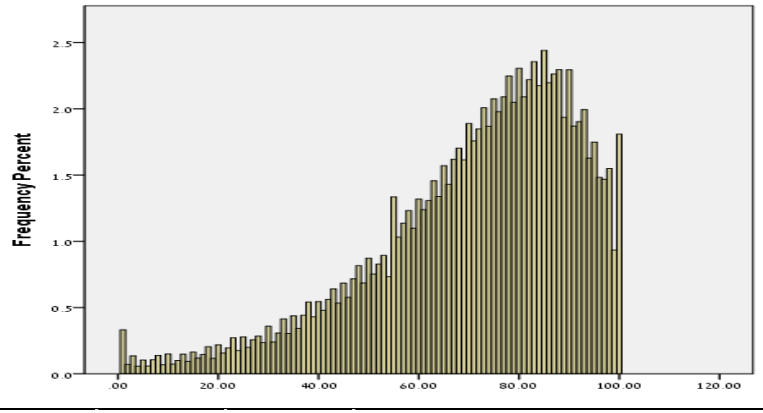
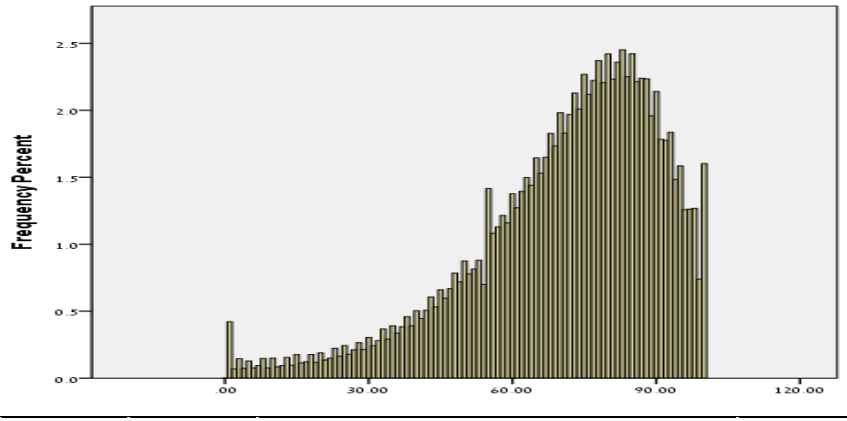
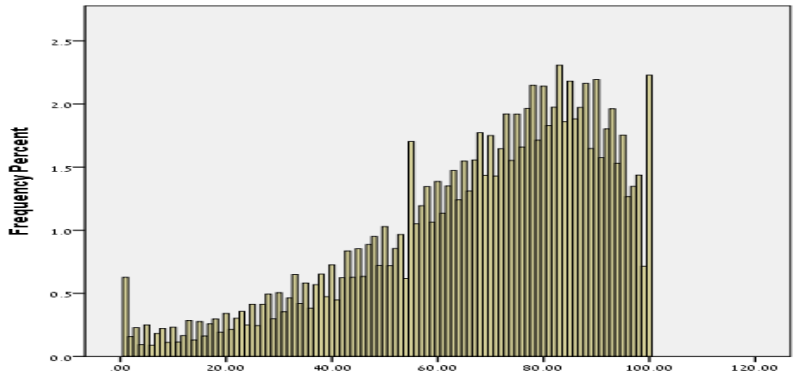
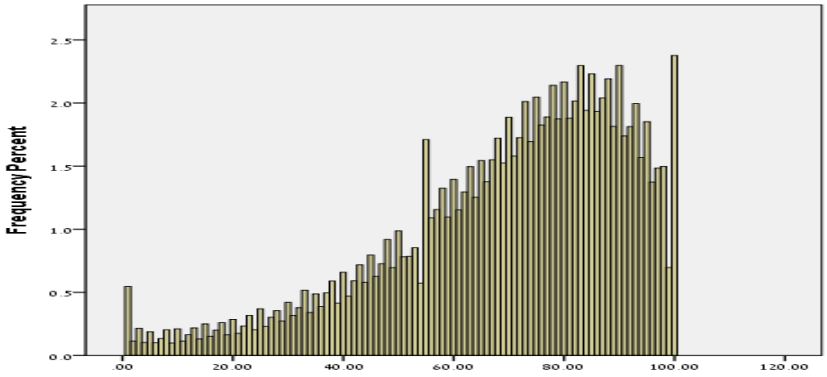
Religious students	**Secular students**
	

Figure 3: The Distributions of Secular Examiners Scores, by Students' Religiosity and Examiners' Gender

A. Male Examiners

Religious students	**Secular students**
	

B. Female Examiners

Religious students	**Secular students**