TI 2023-023/V
Tinbergen Institute Discussion Paper

The Fast and The Studious? Ramadan Observance and Student Performance

Kyra Hanemaaijer[1]
Olivier Marie[1,2]
Marco Musumeci[1]

1 Erasmus University Rotterdam and Tinbergen Institute
2 ROA, CEPR, IZA, and CESIfo.

The Fast and The Studious? Ramadan Observance and Student Performance *

Kyra Hanemaaijer[†] Olivier Marie[‡] Marco Musumeci[†]

This version: April 2023

Abstract: What are the consequences of religious obligations conflicting with civic duties? We investigate this question by evaluating changes in the performance of practicing Muslim students when end-of-secondary-school exams and Ramadan overlapped in the Netherlands. Using administrative data on exam takers and a machine learning model to individually predict fasting probability, we estimate that the grades and pass rate of compliers dropped significantly. This negative impact was especially strong for low achievers and those from religiously segregated schools. Investigating mechanisms, we find suggestive evidence that not being able to sleep in the morning before an afternoon exam was particularly detrimental to performance.

Keywords: Religion, Productivity, Ramadan, Education, The Netherlands

JEL Codes: I2, I24, Z12, J15

*The authors are grateful for financial support from the Netherlands Organization for Scientific Research (NWO 016.Vidi.185.049). We thank Robin Verbeek and Elena Castellaro for providing excellent research assistance on this project. We are grateful to Gabriela Smarrelli and Zihan Ye for taking the time to give detailed comments on an earlier version of the paper. We also wish to thank participants at multiple seminars, workshops, and conferences for their valuable suggestions. Hanemaaijer: hanemaaijer@ese.eur.nl; Musumeci: musumeci@ese.eur.nl; and Marie: marie@ese.eur.nl (corresponding author).

†Department of Economics, Erasmus University Rotterdam (EUR) & Tinbergen Institute (TI).

‡EUR & ROA, TI, CEPR, IZA, and CESIfo.

1 Introduction

An individual's religious obligations and civic duties are sometimes misaligned. This internal conflict may be particularly pertinent for religious minorities, including the quarter of Muslims who reside outside of Muslim-majority nations (Lugo et al., 2011). A relevant case for this group is the obligation to fast during the Ramadan month, while simultaneously carrying on with their regular productive activities in countries that do not accommodate for this. During Ramadan, all healthy post-pubertal Muslims must refrain from eating and drinking between sunrise and sunset for up to 30 days.[1] While some may be able to adapt their schedule to minimize productive disruptions during the Ramadan month (e.g. by changing working hours or taking holidays), it may not be an option for most. This may be especially problematic, and with potential long term consequences, when its timing coincides with periods in which an individual's relative productivity is tested: i.e. during high stakes exams.

The religion-productivity relationship our paper explores closely relates to two recent studies which have provided compelling macro-level evidence of negative impacts of religious practices on economic growth. Campante and Yanagizawa-Drott (2015) exploit differences in the length of days during Ramadan in Muslim countries and Montero and Yang (2022) the timing of Catholic saints festivals in Mexico. In this paper, we focus instead on the effect of religious practices on productivity at the micro-level, in particular when the macro-context in which those impacted live is otherwise not affected. This matters as it could make a religious minority group appear relatively less productive than the majority group. Exacerbating already existing (large) differences in outcomes as a consequence, the impact of religious practices on relative productivity of minorities provides an important new angle to consider when investigating the causes of educational inequalities across groups (Blanden et al., 2022).

[1] Some exceptions to this obligation exist and are mainly related to health conditions, pregnancy, menstruation, old-age, or traveling.

The start of the Ramadan month moves backwards every year, because the Islamic calendar is shorter than the Gregorian calendar. In recent years, the Muslim holy month occurred in May and June, the period when high school graduation examinations take place in most Western countries with no accommodation made for the fasting obligation of Muslim students.[2] As a consequence, Muslim students - who are already part of an under-privileged population group - are required to take educational trajectory defining exams while observing Ramadan. The potential negative repercussions on these students' academic performance have been highlighted in the media of most Western countries.[3] Despite the media attention this generated and the growing share of Muslims living in affected countries,[4] the causal evidence of the impact of Ramadan exposure on such academic tests is very limited. The objective of this paper is to address the existing gap in knowledge by conducting a meticulous investigation into the impact of Ramadan observance on students' performance during high-stakes exams in the Netherlands.

In the medical literature, short-term fasting and dehydration are associated with cognitive impairment (Adolphus et al., 2016; Wittbrodt and Millard-Stafford, 2018). However, fasting is not the only factor that could impact the performance of Muslim students during Ramadan, as their daily schedule radically changes as well at this period. In order to respect the fast when days are long, meals have to be consumed very early in the morning (*Suhoor*) and late at night (*Iftar*). Consequently, Ramadan observance has been associated with important disruptions in sleep patterns (Bahammam, 2003, 2006, 2013; Margolis and Reed, 2004) potentially leading to sleep deprivation, which is itself strongly linked to loss of

[2]In some Muslim-majority countries, schools change their schedule to take into account the Ramadan month. For example, in Morocco, a reduction of classes duration is often introduced, while in Saudi Arabia, schools are closed for (most of) the duration of Ramadan (this was not the case in 2022).

[3]The debate on this topic has risen in several countries, such as the UK ("Popular exams in UK to be rescheduled to avoid Ramadan", The Guardian), the USA ("Testing and Fasting", Inside Higher ED), France ("*Jeûner pendant le bac : le défi des lycéens qui font le Ramadan*", France24), Germany ("Hungry Students? Postponed Exams? Ramadan In German Schools", Worldcrunch), Italy ("*La Maturità incrocia il Ramadan*", Corriere della Sera), and the Netherlands ("*Vasten en eindexamens, gaat dat wel samen?*", NOS).

[4]Muslims represent today around 5% of all the Europeans and by 2050 this number is expected to grow to 11% (Lugo et al., 2011)

cognitive functions (Curcio et al., 2006; Csipo et al., 2021). One would thus expect the existence of clear evidence of a negative impact of observing the Ramadan fast on performance. Surprisingly, the existing (very) small scale experiments looking at the effect of Ramadan on cognitive functions in a lab setting provide mixed evidence. Tian et al. (2011) find that psycho-motor function, vigilance, verbal learning and memory are impaired in male athletes complying with Ramadan obligations, while Ghayour Najafabadi et al. (2015) and Yasin et al. (2013) find no adverse effect on cognitive functions.

The few studies which, as in our setting, have used changes in student performance at real-world exams to explore the impact of Ramadan observance also come to inconsistent conclusions. Oosterbeek and van der Klaauw (2013) find that each week of overlap between a university microeconomics class and Ramadan is associated with a decrease in the exam grade of students with Muslim sounding names of almost 10% of a standard deviation. Hornung et al. (2023) and Kökkizil (2022) use cross-country data on TIMMS and PISA tests to compare performance at tests taken up to one year after the end of Ramadan of students with parents from Muslim majority countries. Hornung et al. (2023) find positive effects which they explain with Ramadan observance having a positive impact on social capital formation of young Muslim students. On the other hand, Kökkizil (2022) finds that Ramadan makes the existing gender norms and stereotypes more salient: in the three months after the end of Ramadan, the academic performance of female Muslim students in STEM subjects significantly worsens compared to their male peers. Our study builds on this small literature, while seeking to improve on it in a number of ways.[5] We are the first to look at performance changes at exams that take place *during* rather than *after* Ramadan, which we believe can much better capture productivity changes. Second, we rigorously explore which students are likely to comply with Ramadan obligations, by developing a

[5] There is a much more developed literature that has exploited the timing of Ramadan on other outcomes. Almond and Mazumder (2011), Van Ewijk (2011), Almond et al. (2015) have shown that prenatal exposure to Ramadan results in worse health outcomes at birth and later in life in various contexts. Schofield (2020) considers the impact of Ramadan on agricultural output in India. Colussi et al. (2021) look at voting behaviour and violent attacks against Muslim communities in Germany, and Bertoli et al. (2022) find that working accidents increase during Ramadan in Spain.

machine learning algorithm that assigns compliance probabilities to students, rather than relying on nationality or country of origin or name. Third, rich administrative data enables us to obtain precise average estimates, and to look into heterogeneities and peer effects. Fourth, we attempt to distinguish between the impact of eating and sleeping disruptions of Ramadan by considering the exact timing of exams. This allows us to make relevant policy recommendations on how to attenuate the negative effect of Ramadan on Muslim students by only changing the exams schedule rather than avoiding the entire Ramadan period. Finally, we use data on a standardized end-of-primary-school test at the national level to control for changes in ability. We show that this is especially relevant for students with a migration background, as their performance improves relative to native students over time.

Our focus is on the impact of Ramadan exposure on (compliant) students' performance at the secondary education graduation exams in the Netherlands. This examination consists of several written tests, covering different subjects, that are the same at the national level. All high school students, in order to graduate and potentially enter tertiary education, need to pass these exams. We focus on the final exams that took place from 2014 to 2019, with the Muslim holy month overlapping with the exams in both 2018 and 2019. In all previous years included in our analysis (2014 - 2017), Ramadan took place after the final exams and thus barely affected the instruction of previous cohorts as this was mostly during the summer holidays. This Ramadan-exams overlap is the first difference we exploit in our difference-in-differences identification approach. The second difference stems from the variation in Ramadan observance of students. This strongly depends on correctly identifying the students that are most likely to comply with Ramadan obligations. Based on Dutch survey data on religiosity and fasting behavior, we show that simply considering as treated all students with a migration background from a Muslim majority country would miss-classify many students. We instead assign treatment based on a machine learning (ML) model that predicts Ramadan compliance probability based on a large set of observable characteristics for all individuals in our sample. This reveals a large difference in compliance between and within the two main Muslim minorities in the Netherlands (individuals with a Turkish and Moroccan migration background). While most students with a Moroccan background are very likely to be compliers, those with a Turkish background have a much wider distribution in their strict adherence to the Ramadan fast. This novel and more precise classification into treatment is the second difference our identification strategy relies on.

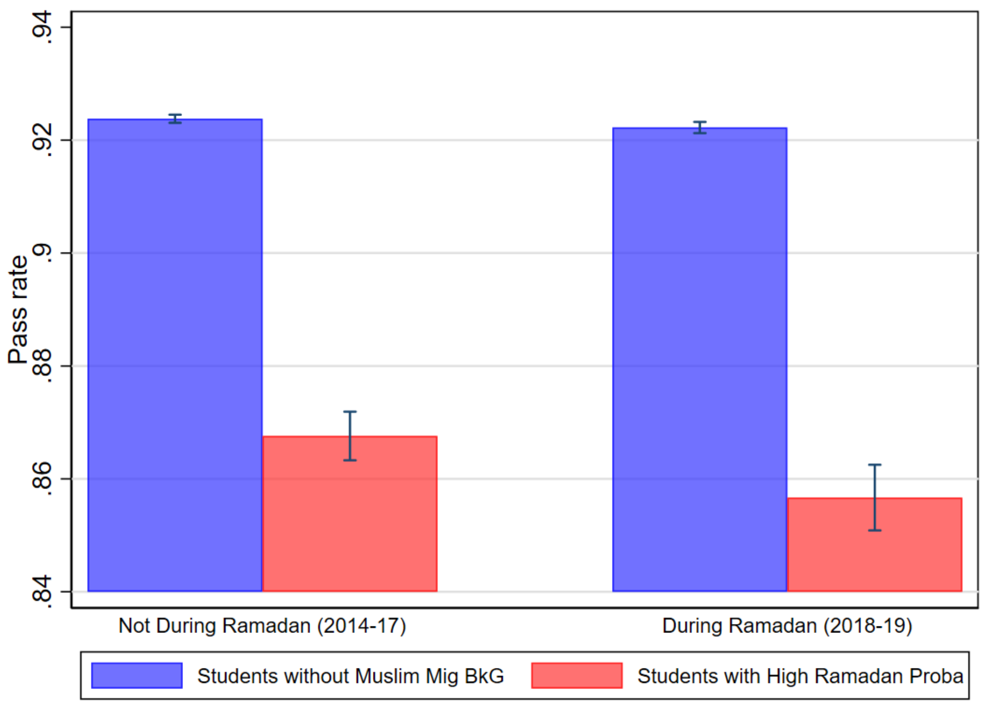
The most basic presentation of this difference-in-differences approach is shown in Figure 1 which reports average pass rates at final exams that did and did not overlap with Ramadan for treated and non-treated students respectively. Since most minority students in the Netherlands are from a socioeconomically disadvantaged population group, their pass rates are always much lower on average. Still, the existing large achievement gap between Muslim and non-Muslim students was strongly and significantly exacerbated - widening by almost 17% - when the high school graduation exams took place during Ramadan.

These findings are confirmed using statistical models which interact treatment status of students with Ramadan exam years, while controlling for a large set of individual, family, school, neighbourhood, and time characteristics. We estimate that taking a high stakes exam during the Ramadan month decreases the grade of students most likely to fast by 5.8% of a standard deviation (SD). This in turn leads to an increase of 8.3 percent in the probability of these students failing to graduate from high school.

We corroborate the validity of these findings by performing a number of robustness checks to confirm that Ramadan observance is what is driving our results. First of all, we check that these results are not driven by a change in academic ability of students. To do so, we consider as dependent variable the average grade obtained in all tests and homework performed in the last years of high school before the final exam.[6] These 'continuous' exams should not be affected by Ramadan, as they do not overlap with it, but are a very good proxy for the academic ability of students. No effect can be found in 2018 and 2019, confirming that our findings are not driven by lower skilled students in the treatment group

[6] The change in academic ability of students is also controlled for by including the standardized end of primary school test score in all regressions.

Figure 1: Average pass rate at final exams by Ramadan compliance probability



Notes: Bar graph with the average pass rate and 95% confidence interval of students with a Moroccan or Turkish migration background and a high Ramadan compliance probability Moroccan migration background and students with no Muslim-majority country migration background in the years that Ramadan did not overlap with the final exams (2014-2017) and when it did overlap with the final exams (2018-2019).

taking the exams those years. Another revealing check is a placebo test which considers as treated those students that are very unlikely to observe Ramadan, because they do not have a migration background from a Muslim-majority country. We select this group based on individual characteristics that our ML model identifies as the most relevant to predict Ramadan compliance. Hence, these students are similar in almost all aspects to our 'real' treated group, but should not be affected by Ramadan. Here again we do not find any effect for placebo students which shows that the results are unlikely to be driven by some shock affecting students with specific characteristics (e.g. disadvantaged background) in the years of the Ramadan-exams overlap. Finally, we check the robustness of our results by including: i) sibling fixed effects, and ii) family fixed effects, which includes cousins as well. In this way we can control for unobserved characteristics that are shared among siblings and cousins.

These alternative specifications produce results that are quantitatively similar to our main results. We thus argue that we can causally attribute the drop in performance that we observe to the overlap of final exams with the Muslim holy month.

We then perform a large number of investigative heterogeneity analyses to uncover who is most affected by Ramadan. There are no significant gender differences and students of all ability and income levels are relatively similarly affected when it comes to the drop in grades obtained. However, the lower performance at the final exam is especially detrimental for the graduating probability of students in the lowest quartile in terms of pre-final exam performance. School level heterogeneity analyses reveal that students in the most segregated schools, with respect to parental income or to the proportion of students with a high Ramadan compliance probability, are driving our results. Those in schools in the top quartile of the 'complier segregation' saw their grades drop by 7% of a standard deviation and were 12.1% more likely not to graduate. This strong segregation effect leads us to explore the possibility of peer-effects in Ramadan observance likelihood. To do so, we exploit within-school across-cohort differences in the proportion of students with high Ramadan compliance probabilities. We find that there is a strong additional impact on treated students' test scores, when students have more peers in their cohort that are treated as well.

Finally, we attempt to disentangle the impact of fasting itself and the important disruptions to sleep patterns that usually accompany Ramadan observance. For this we first consider exam results depending on the type of exam taken as lack of sleep does not affect all type of cognitive performance similarly Lim and Dinges (2010). Short memory based tests are somewhat more affected, suggesting sleep deprivation matters. We then consider how far into the Ramadan period each exam is taken. We hypothesize that fasting would mostly affect exams in the first week of Ramadan, while sleep deprivation would accumulate and would mostly impact exams in later weeks. However, we do not find any evidence of such a different effect depending on the week of Ramadan in which the exam is taken. We then investigate how time of day when an exam is taken, i.e. morning or afternoon, may matter for treated student's results. The main finding here is that afternoon results are especially negatively affected, but only when there was already an exam taken by the same student in the morning. We believe this is because those affected by Ramadan are able to counter the effects of sleep deprivation by taking (long) naps after the early meal when they have no exams in the morning. Consequently, we put forward a simple recommendation of only scheduling tests in the afternoon with no morning exam before, in case it is not politically feasible to altogether move performance evaluations away from the Ramadan period. This will at least reduce the very negative impact we uncover on testing Muslim-minority students whose religious beliefs conflict with their civic duties, even when these concern key moments that can have long term effects on their lives.

The rest of the paper is structured as follows. The next section describes the background of Ramadan and Muslims in the Netherlands, the Dutch secondary school system and the overlap between the central exams and Ramadan. Section 3 describes the data we use. In section 4 we discuss the identification of the 'treated' group, our difference-in-differences identification strategy and the machine learning model we use to predict Ramadan compliance. Section 5 describes our main results and in Section 6 we explore potential heterogeneity by student and school characteristics and by subject. In Section 7 we look into underlying mechanisms and propose policy implications. Section 8 concludes.

2 Background

2.1 The Ramadan month

The Ramadan month, which is the ninth month of the Islamic calendar, imposes a religious fast that is considered one of the five pillars of Islam. It requires all healthy Muslim individuals who have reached puberty to abstain from eating and drinking between sunrise and sunset. Some exceptions to this obligation exist, such as for health conditions, pregnancy, menstruation, age, or travel. The Ramadan month lasts for 29 or 30 days and its starting date is determined by the Islamic calendar. This calendar is shorter than the Gregorian one by approximately 11 days, causing the Ramadan month to move backwards relative to the Gregorian calendar every year by 11 days. As Ramadan moves over time, the daily duration of fasting, that depends on sunrise and sunset, varies by year. In the Netherlands, the sun rises at 5.20 AM and sets at 22 PM on the longest day of the year (21 June), while it rises at 8.50 AM and sets at 16.30 PM on the shortest day (21 December). This causes the Ramadan month to have a very different impact on daily life depending on the timing during the year. On the shortest day of the year, Muslims do not have to adjust their breakfast and dinner time and only have to fast for 7 hours and 40 minutes. On the contrary, if Ramadan takes place during the summer, Muslims need to have breakfast before 5.20 AM and dinner after 22 PM, and hence will have to abstain from eating and drinking for 16 hours and 40 minutes.

In some Muslim-majority countries, working hours are reduced during Ramadan and schools arrange changes to their schedule to accommodate students' needs deriving from the fast. For example, classes duration can be shortened, as it happens in Morocco, or schools are closed completely for the entire duration of Ramadan, which is the case in Saudi Arabia. In most European countries, employers and schools do not implement any change to help Muslim individuals who are fasting during the Ramadan month. However, Muslims represent today around 5% of all the Europeans and by 2050 this number is expected to grow to 11% (Lugo et al., 2011), which will lead to a rising need for accommodation for Ramadan also in non-Muslim countries.

2.2 Muslims in the Netherlands

According to estimates of Statistics Netherlands, in 2021 around 5% of the population aged 15 or above in the Netherlands was Muslim.[7] Moreover, more than 450 mosques can be found around the country.[8] This is not surprising considering that the two largest groups of migrants in the Netherlands come from Muslim majority countries: Turkey and Morocco. These two groups represent respectively 2.6% and 2.5% of the population. The percentage provided above refers to individuals who are first, second and third-generation immigrants[9]

Individuals with Moroccan or Turkish migration backgrounds tend to concentrate in large cities. As of 1st January 2022, around 45% of the population with a Moroccan migration background and 35% of the population with a Turkish migration background were living in one of the 4 large cities in the Netherlands (Amsterdam, Rotterdam, Utrecht, The Hague).[10] This is evident from the map presented in Figure 2 that shows the percentage of individuals with a Moroccan or Turkish migration background at municipality level.[11] The darker areas correspond to large cities in the Netherlands. Each dot represents a mosque where the language of the sermon is Moroccan or Turkish.

The age profile of the population groups with a Moroccan or Turkish migration background is quite different from the population group without a migration background. Individuals with a Moroccan or Turkish migration background are relatively younger (on average in their mid-thirties) than those without a migration background (on average around fifty

[7]Data come from Statistics Netherlands.

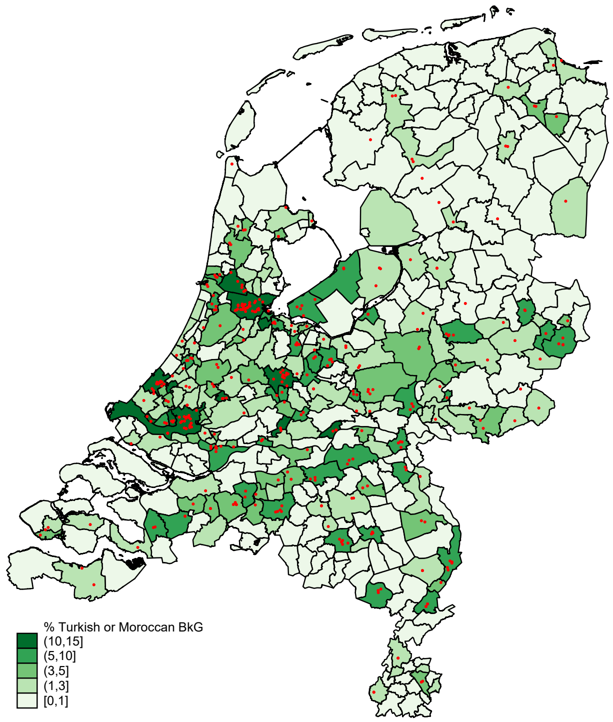
[8]Data was collected from Moskeewijzer.nl in 2021.

[9]See data section for an explanation on how individuals are classified as 1st, 2nd or 3rd generation immigrants. Numbers are author's calculations based on data from Statistics Netherlands on 1st and 2nd generation migrants and 3rd generation migrants.

[10]Within these cities, there is a lot of variation in the level of segregation across neighbourhoods. In certain neighbourhoods, up to 65% of the residents have a Moroccan or Turkish background.

[11]See Figure A1 for a map only on individuals with a Moroccan migration background and Moroccan mosques and see Figure A2 for a map only on individuals with a Turkish migration background and Turkish mosques.

Figure 2: The share of municipalities with a Moroccan or Turkish migration background and the location of mosques in the Netherlands



Notes: Municipality level proportion of people with a Moroccan or Turkish migration background in 2017 in the Netherlands. The locations of mosques where the language of the sermon is either Turkish or Moroccan are indicated wit red dots. Source: CBS Statline, statline.cbs.nl; Moskeewijzer, https://moskeewijzer.nl/.

years old).[12] Focusing on the population of high school students in 2017, we therefore see that these two ethnic groups represent a larger share than in the general population. They jointly represent 7.1% of the students, with the students with a Moroccan migration background accounting for 3.6% and those with a Turkish migration background for 3.5%. The majority of the students with a Moroccan or Turkish migration background are second generation migrants, meaning that respectively 95.1% and 92.7% have at least one parent that was born in Morocco or Turkey. Most students have parents that migrated to the Netherlands in the end of the 1980s, when a large wave of labour migrants was followed by follow-on migration, as the wives and children of these labour migrants also moved to the Netherlands.

[12] Data come from Statistics Netherlands.

Approximately 3.2% (3.6%) of the students with a Moroccan (Turkish) background migrated themselves, and 1.7% (3.7%) has at least one grandparent born in Morocco (Turkey).

2.3 Secondary education in the Netherlands

After completing primary school, pupils in the Netherlands start secondary school usually at the age of 12.[13] They can choose (advised by the results of a standardized national test called CITO test and their teacher) to go to one of the following three types (tracks) of secondary education: pre-vocational secondary education (*VMBO*), senior general secondary education (*HAVO*) or pre-university education (*VWO*). The duration of these programs is respectively four, five and six years. At the end of their third year (or second, in pre-vocational secondary education) students have to choose a certain subtrack (*profiel*) that corresponds to a selection of topics. In the years after the subtrack choice, students will mainly focus on the topics that correspond to this subtrack. However, certain subjects are compulsory in every subtrack, such as Dutch, English language and literature and physical education. Students are obliged to stay in school up to the age of 16. After the age of 16, compulsory education is replaced with a compulsory "start qualification" up to the age of 18. A start qualification is obtained when students obtain a senior general secondary education or pre-university education diploma, or a diploma in secondary vocational education (which students can attend after graduating from pre-vocational secondary education).

To obtain the secondary school diploma, students have to pass the secondary school graduation exam, that covers various subjects. They usually take this exam at the age of 16 to 18, depending on the track they are in. The final grade is for 50% determined by the average of decentralized continuous school exams and for 50% by the average of the centralized final exams in May of the final year.[14] The results of the continuous school exams depend on a set of assignments and tests that take place during the last two or three

[13]In this section we focus only on secondary education in Netherlands but in Appendix Table A1 we provide a detailed graphical overview of the entire Dutch educational system.

[14]Some subjects do not have a central final exam (for example physical education) and therefore the final grade is completely determined by the school interim examination.

years of high school. These tests and assignments are different for each school because they are formulated and graded by the student's teachers. On the contrary, the central exam consists of several written exams, that are the same at national level and are corrected by the student's teacher but also by a teacher that is randomly selected from another school in the Netherlands. The number of subjects that are tested during the central exam depends on the track and on the student's choices, but it usually ranges from 5 to 8. The grades of both the central and continuous exams are on a scale from 1 to 10, with the latter being the maximum grade that a student can achieve and 5.5 being sufficient.

The central exam usually takes place in May over a period of three weeks. During these weeks, the students must take all their exams according to a schedule determined by the government.[15] In June there is a retake session of the central exam. During this session, students can retake exams if they were severely ill during the main session of May or they had another serious personal reason not to take the exam. Moreover, in the retake session a maximum of one exam can be taken for a resit, if a student wants to improve the obtained grade. In this case, the final grade obtained by the student in that exam is the highest grade (so either the original or resit grade). In July there is a second retake session of the central exam, which is only available in special cases, when students could not take the exam in the original exam period and the first retake session.

A student graduates from secondary education when the average grade of the central exam is a 5.5 or higher. Depending on the student's track, there are additional requirements for the final grades (average of the continuous exam grade and central exam grade) needed to graduate.[16]

[15] See Examenblad.nl for the detailed schedule of (and more information on) all exams.

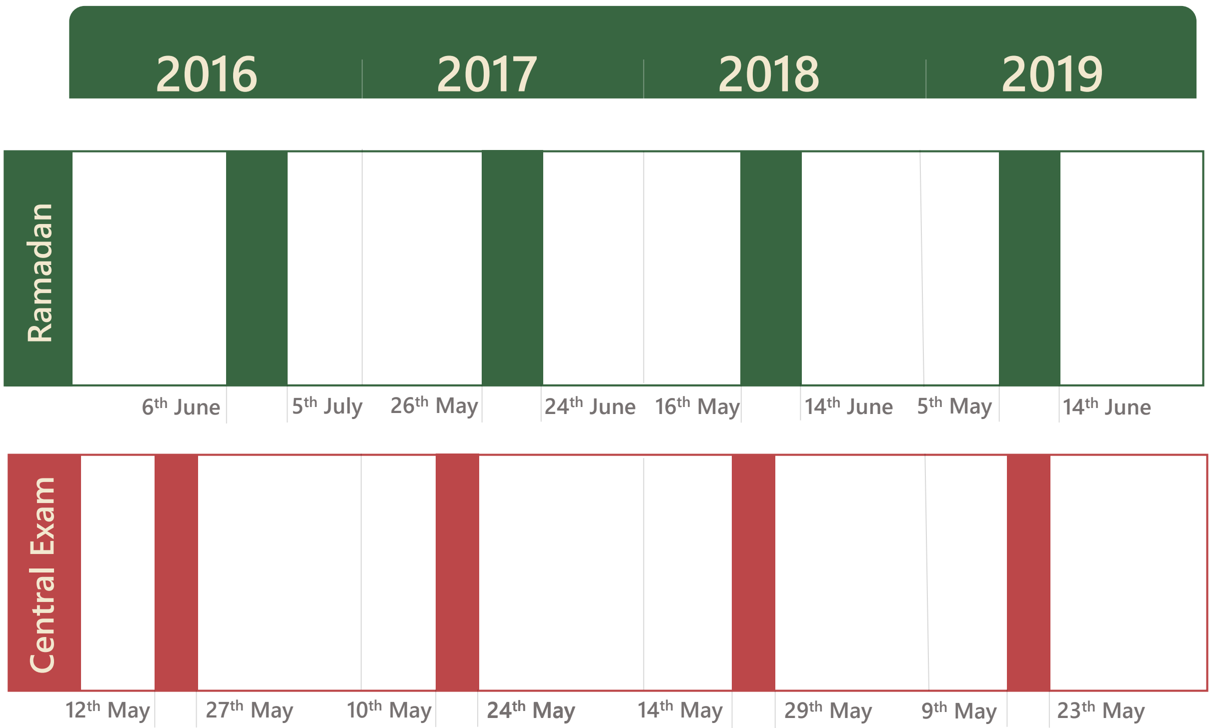
[16] In addition to the requirement of a 5.5 as overall average grade of the central exam, students in senior general secondary education and pre-university education can only have one 5 as a final grade (which consists for 50% of the continuous exam grade and for 50% of the central exam grade) for Dutch, English and Math, and they need to have an average final grade overall of 6 or higher, no final grade lower than a 4 and a pass for physical education. In addition to the requirement of a 5.5 as overall average grade of the central exam, students in pre-vocational secondary education need a 5 or higher as a final grade for Dutch, an average final grade overall of 6 or higher, no final grade lower than a 4, a pass for physical education, art education and research project (this latter requirement only applies to students in the theoretical type of pre-vocational secondary education (*VMBO-TL*)).

2.4 Central Exam and Ramadan

In recent years, there has been an overlap between the central exam, which is always held in May, and Ramadan, as illustrated in Figure 3. Prior to and including the year 2017, the month of Ramadan occurred after the central exam had concluded (so neither did Ramadan overlap with the central exam, nor with the preparation right before the central exam). However, in the years 2018 and 2019, the central exam overlapped with the ninth month of the Islamic calendar.[17] This overlap was complete in 2019 while in 2018 only partial, as the first three days of the central exam took place before the beginning of Ramadan. When Ramadan overlapped with the final exams in May, sunrise was at approximately 5.50 AM and sunset at approximately 21.30 PM, meaning that students had breakfast before 5.50 AM and dinner after 21.30 PM, and had to fast for about 16 hours per day. Morning exams started at 9 AM and afternoon exams at 13.30 PM. Hence, students did not have (much) time to sleep in after breakfast when they had a morning exam, and had abstained from drinks and food for more than 7.5 hours when they started an afternoon exam.

A member of the parliament of left-wing political party "DENK" already requested the rescheduling of the central final exams in 2016 to avoid overlap with the Ramadan month. The State Secretary of Education then answered to this request by saying that "it is practically impossible to reschedule the final exams to accommodate Ramadan. ... It could be that students find religious obligations important, even in periods when they take tests. I

Figure 3: Graphical representation of overlap between final exams in secondary school and Ramadan.



Notes: Graphical representation of the overlap between the central exams and the Ramadan month. In the years before and including 2017, there is no overlap between the two, even if in 2017 they are very close. From 2018, the central exam and Ramadan occur in the same period (only partially in 2018). Source: Examenblad.nl, https://www.examenblad.nl/.

leave that decision to the individual.”.[18]

[17] Continuous school exams take place at multiple moments (continuously) over the year and are scheduled by schools themselves. Continuous exams have to end before the central exams start for students in their final year who take the central exams. Students in the years before may also take continuous exams in June or July, and hence Ramadan might have partly overlapped in the years before 2019 (as it took place in May-July in 2014-2018) with one of their continuous exams. However, as the continuous exam grade that we use is constructed out of many tests and assignments over 2 to 3 years, this would mean that only a very small fraction of the continuous exam grade may have been affected.

[18] The original and complete answer in Dutch in the parliamentary document was: “Jaarlijks doen ongeveer 200.000 leerlingen het centrale examen in het vastgestelde tijdvak. Het is de individuele verantwoordelijkheid van een leerling om zich zo goed mogelijk voor te bereiden op toetsen en examens. Het kan zijn dat leerlingen religieuze verplichtingen belangrijk vinden, ook in periodes waarin toetsen worden afgenomen. Die afweging laat ik aan het individu. Islamitische leerlingen die graag deelnemen aan het vasten kunnen er overigens ook voor kiezen om - eventueel in overleg met hun geestelijke - de vastenperiode te verplaatsen. Er zijn leerlingen die dat ook doen.’

3 Data

3.1 Administrative data

We use administrative data from different registries compiled by Statistics Netherlands.[19] We focus on secondary school students that took the final exam between 2014 and 2019 and for whom we observe all relevant characteristics[20] (841,376 observations, or 789,259 students)[21] in 549 schools. For those students, we observe their central exam grade (both including and excluding retakes) and continuous school exam grade both as an average of all courses together and at the subject level, whether they passed the exams/graduated, in which sub-track and track they were in (and in which tracks they were before, if they moved up or down, or whether they repeated a year), their result on the national end of primary school (*CITO toets*) test (as a standardized measure for ability), and demographic characteristics such as birth date and migration background. We also observe neighbourhood characteristics (share with a Turkish or Moroccan migration background, share that uses benefits, average household size, share under 45 years old, fertility), school characteristics (share with a Turkish or Moroccan migration background) and characteristics of their parents (income, unemployment, illness),

We classify individuals as having a foreign background if they are first, second or third-generation immigrants. An individual is considered a first-generation immigrant if he is born outside the Netherlands and one of his parents is born abroad as well. A second-generation immigrant is an individual who is born in the Netherlands but has at least one parent who is born abroad. A third-generation immigrant is an individual who is born in the Netherlands and whose both parents are also born in the Netherlands, but at least one of

[19] See Appendix Section B for a detailed description of the data set-up, sample selection process and variable definitions.

[20] The characteristic that causes the largest selection in sample is the national end of primary school test that we use as a standardized measure for ability. We can only link about 75% of the exam takers to their result on this test.

[21] We have for approximately 6.5% students multiple observations, as they do not graduate from high school in one attempt and have to repeat the final year including the final exams.

his grandparents is born abroad. If a person does not belong to any of the above categories, then he is considered to have no migration background.[22]

Figure 4 shows the distribution of secondary school central exam grades (standardized at track level) in 2017 for students with the two most common migration backgrounds (Moroccan migration background in red and Turkish migration background in green) and students with no migration background or a migration background from a non-Muslim majority country (in blue). This figure shows an existing gap in the standardized average exam grade between Moroccan (Turkish) background students and the students without a Muslim-majority migration background of approximately 48% (50%) of a standard deviation. Notice that this gap is measured in 2017, when no overlap occurred of Ramadan and the final exams.

It also provides additional descriptive statistics for these three groups. Students with a Moroccan or Turkish background have on average poorer parents, live in more segregated neighborhood and performed worse in the standardized end of primary school test. Put together, the descriptive statistics presented underline how students with a Moroccan or Turkish background belong to a strongly under-privileged population group.

3.2 Survey data

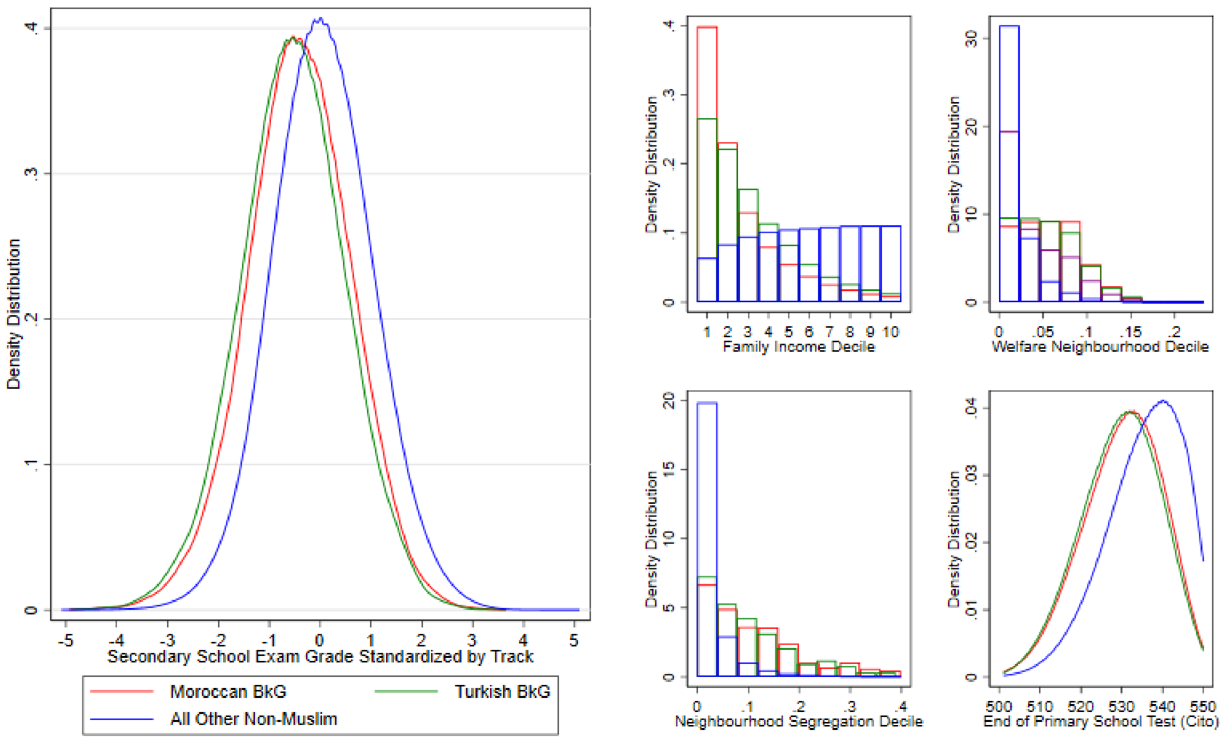
To determine who is observing Ramadan in the Netherlands, we make use of the so-called SIM Survey (*Survey Integratie Minderheden*, meaning survey on minorities' integration). This survey was conducted in 2006, 2011, 2015 and 2020. It contains questions on religiosity, and specifically on how strictly individuals comply with fasting during Ramadan, by 19,002 respondents in total. We are able to link the respondents to our administrative data, and therefore have information on a rich set of characteristics of the respondents. There are 5,587 respondents with a Turkish or Moroccan migration background (after cleaning and linking the data).[23] Descriptive statistics on the SIM survey data are provided Table A3.

[22] See Appendix Section B for a detailed description of the classification of migration background.

[23] See Appendix Section B for a detailed description of the SIM survey sample (selection) and cleaning of the SIM data.

Figure 4: Distribution of characteristics of students by migration background

A: Standardized final exam grade **B:** Other characteristics



Notes: The left graph shows the density distribution of the average central exam grade standardized by track and exam year in baseline year 2017. The right graph plots the density distribution of family income divided in deciles, the share of a neighbourhood using welfare (*bijstand*), the share of a neighbourhood with a Moroccan migration background and the grade of the end of primary school test (standardized by year). The distributions are given for students with a Moroccan migration background (red), Turkish migration background (green) and a non-Muslim migration background or no migration background (blue).

3.3 Mosque data

Finally, we collect data on the location and type of mosques in the Netherlands. We do so by scraping the geolocation of mosques, together with other information, from the website moskeewijzer.nl.[24] We then link the obtained dataset with the administrative data from Statistics Netherlands. This allows us to compute the distance between the neighbourhood where a student is living and the neighbourhood of the closest mosque.

[24]The data were collected in 2021.

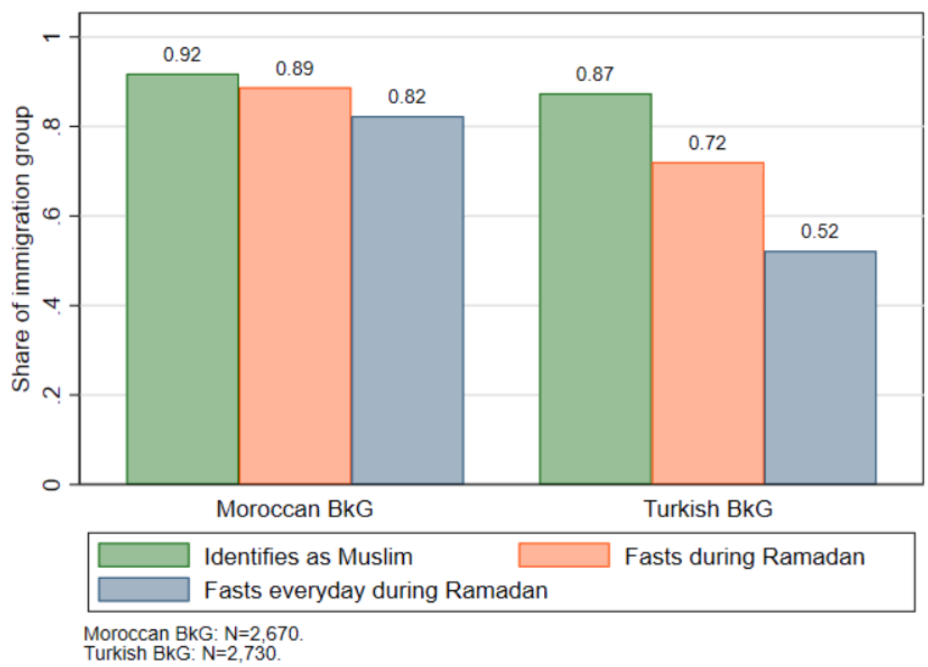
4 Identification strategy

4.1 Identification of the 'treated' group

To causally identify the effect of Ramadan on educational outcomes, we want to compare students that observe Ramadan during their secondary school exams with students that do not. Ideally, we would want to know which students comply with Ramadan during their final exams to identify who are 'treated'. While we do not have this information, we do have information on self-reported religiosity of 2,670 people in the Netherlands with a Moroccan migration background and 2,730 with a Turkish migration background. We focus on these groups for two reasons. First, they are the two largest migration groups in the Netherlands and they jointly represent more than 7% of the high school graduation exam takers (in 2017). Secondly, for other migration groups from Muslim-majority countries, no detailed information is known about Ramadan compliance. Considering these groups as treated would identify the effect on a group with unknown, and potentially very low, Ramadan compliance. The individuals who migrated in the Netherlands might have different religious attitudes from individuals in the country of origin and using the major religion in the country of origin might therefore not be a good proxy. For example, a relatively large part of the Iranian Dutch are not religious and around 20% of them are Christian, while the majority of the people living in Iran are Muslim (Huijink, 2018).

Figure 5 shows the share of individuals with a Turkish or Moroccan migration background who identify as Muslim and report to fast (every day) during Ramadan. These two groups are almost equally likely to identify as Muslim. However, when being asked specifically how strict they are with fasting, the behaviour of the groups diverge significantly. 84% of the respondents with a Moroccan background say they fast everyday during Ramadan, compared to only 54% of the respondents with a Turkish background. When we take this as a proxy for how strict people are with complying to Ramadan, we would expect students with a Moroccan background to be more impacted during the overlap of Ramadan with the high-stake secondary school exams. Other measures of compliance with the rules of Islam are shown in Table A2, confirming that on average individuals with a Moroccan migration background in the Netherlands comply more with Islamic rules than individuals with a Turkish migration background. The sample of respondents in the SIM data is very similar to our sample of exam takers (see Table A3 in the Appendix), except for age. The sample of exam takers is older than the sample of respondents in the SIM data. However, Figure A3 in the Appendix shows that everyday Ramadan compliance is very stable over age, which ensures us that our sample of 16 to 18 year old exam takers is not very different in terms of compliance to Ramadan.

Figure 5: Self-reported religiosity of SIM respondents with a second generation Moroccan or Turkish background



Notes: Graphical representation of the average self-reported religiosity (being Muslim, fasting during Ramadan and fasting everyday during Ramadan) of respondents with a Moroccan migration background (on the left) and a Turkish migration background (on the right). Source: Author's calculations based on integration survey (*Survey Integratie Minderheden*) data on 2006, 2011, 2015, and 2020.

Given the high compliance with Ramadan fasting of individuals with a Moroccan migration background, an obvious choice is to consider students with such a migration background as the treated group. However, with this approach, it remains unclear how students with a Turkish migration background should be considered in our analysis. If we would treat them as treated or control group, we would incorrectly classify half of them. If we would want to avoid this by dropping them from the analysis, we would drop 3.5% of the students, even though half of them might be treated by the Ramadan fast. For this reason, rather than relying solely on the migration background to classify individuals as treated, we develop a classification algorithm based on machine learning that assigns to each individual with a Moroccan or Turkish migration background a probability of being treated, that is, a probability of complying with Ramadan obligations. This classification algorithm exploits individual and neighbourhood characteristics to produce such probabilities. The model is trained on the SIM survey respondents and then used to populate the sample of high school exam takers with individual compliance probabilities. More details about this model are presented in section 4.3.

In our main specification, the treated individuals are those with a predicted probability that is higher than a certain threshold. In an alternative specification we use the continuous probability itself as a measure of treatment intensity. Finally, we also implement a third specification where students with a Moroccan migration background form the treated group and Turkish background students are dropped from the sample. In all of our specifications the control group consists of all other students that do not have a migration background from a Muslim-majority country (non-Muslim background) (see Data Appendix).

4.2 Main Econometric Specification

In order to estimate how Ramadan compliance affects students performance in the high school graduation exams, we exploit a natural experiment in which Ramadan exposure during the exams varies over the years, as described in Section 2.4. We apply a difference-in-differences approach comparing 'treated' students, those with a high probability of complying with Ramadan obligations, with the control group composed of students without a migration background from a Muslim-majority country. Equation (1) describes our difference-in-differences main specification model:

$$y_{i,t,s,p} = \alpha_{s,p} + \sum_{T=2014}^{2019} \beta_T \, Treated_i * T_t + \zeta \, Treated_i + \delta X_{i,t} + \gamma_t + \epsilon_{i,t,s,p} \quad (1)$$

The outcome variable $y_{i,t,s,p}$ indicates the final exam grade (standardized by track and exam year) obtained in the central exam by student i, in year t, school s and subtrack p, or a pass dummy taking value one if the student graduated from high school. In our preferred specification, the variable $Treated_i$ is a dummy equal to one for individuals with a high probability of Ramadan compliance. In the alternative specifications, $Treated_i$ is the continuous probability predicted by the machine learning model or a dummy that takes value one if the individual has a Moroccan migration background. T_t represents a dummy equal to one if the exam is taken in year T. This will generate 5 estimates β which we can plot graphically to check for consistent differences in outcomes and pre-trends, which will help us to validate the common trends hypothesis. This specification also serves as a 'placebo in time' as it will reject or not the possibility that we would detect an impact of Ramadan on treated students before Ramadan actually coincided with the final exams. The coefficients that measure the average effect of Ramadan on exam outcomes are β_{2018} and β_{2019}. $\alpha_{p,s}$ are school-track-subtrack fixed effects, $X_{i,t}$ is a vector of controls and γ_t are exam year fixed effects. Controls included in vector $X_{i,t}$ are on the individual level: gender, age at exam, ability as measured by the end of primary school test (standardized by test year), continuous exam grade (standardized by track and exam year, which is dropped when it is used as outcome variable instead), moved up from lower track or failed the final exams last year, number of siblings, a dummy for having a Moroccan migration background (which is dropped when treatment is based on this), a dummy for having a Turkish migration background (which is dropped when treatment is based on Moroccan migration background), a dummy for a different migration background from Muslim-majority countries, time spent abroad, parental income, unemployment or illness benefits, and on the neighbourhood level: the share receiving assistance benefits, the average household size, the ratio under 45 years old, fertility rate, share with Turkish migration background, share with Moroccan migration background. Standard errors are clustered at school-track-subtrack exam year level.

As we know from Figure 4 that there are important level differences in characteristics between the students in our treatment and control group, we validate that students in our treatment group are not changing in other respects - compared to the students in the control group - in those years that Ramadan overlapped with the final exams. We do not find evidence of a compositional change in parental and neighbourhood characteristics over time (family income, parental unemployment, and neighbourhood segregation) (see Figure A4). In particular, one might be worried about selection in ability over time of the treated students. Figure A4 suggests positive selection in ability over time, as measured by the end of primary school test (standardized by year), of the students in the treatment group. If we would not control for ability, we would capture this trend, that might be explained by the integration of migrants over time, which could bias our estimates. Hence, it is very important for the validity of our estimates to control for ability in our analysis, and we have therefore controlled for this in every analysis.

We estimate a second model where the year dummies ($Year_t$) are replaced by a single dummy ($Ramadan\ Year_t$) that takes value one if the central exam is taken in one of the years when it overlaps with Ramadan. The model is the following:

$$y_{i,p,s,t} = \alpha_{p,s} + \beta\, Treated_i * Ramadan\ \ Year_t + \zeta\, Treated_i + \delta X_{i,t} + \gamma_t + \epsilon_{p,s,t} \tag{2}$$

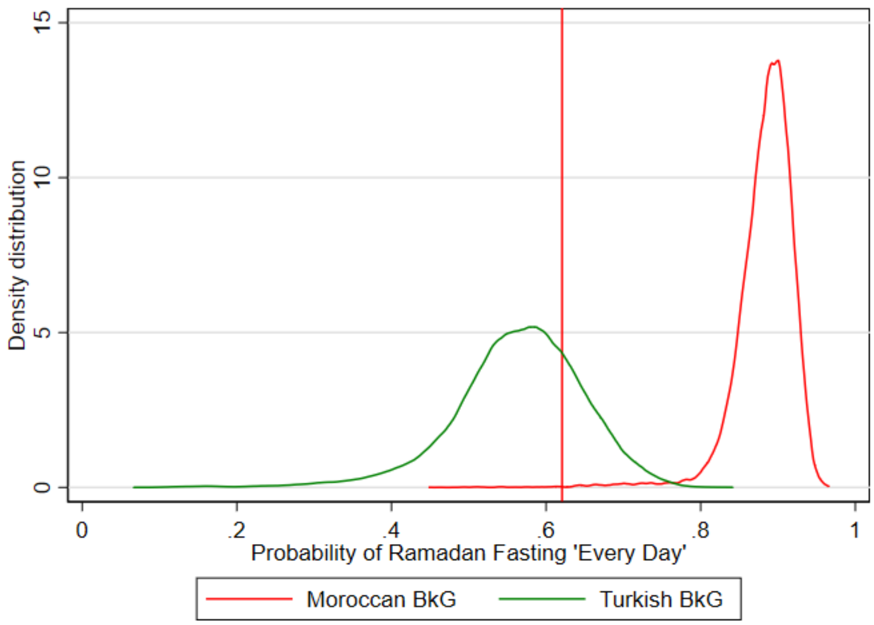
Standard errors clustering, fixed effects and the included controls in equation (2) are the same as in equation (1).

4.3 Machine Learning (ML) prediction of Ramadan compliance

Before moving to the results, we provide in this section more details on how the treated group is constructed. Our aim is to identify a group with high compliance with Ramadan fasting. We do so by building two machine learning classification algorithms: one for individuals with a Moroccan migration background and one for individuals with a Turkish migration background. These models are trained to predict the probability of fasting everyday. We expect this measure to represent a proxy for the individual involvement in Ramadan, not only in terms of fasting compliance, but also in terms of participation to the Muslim community activities during the Ramadan month. The model is based on individual and neighbourhood level characteristics and the training is performed on the respondents of the SIM survey. We implement two different ML techniques: Random Forest and penalised logistic regression. As the latter outperforms the former in terms of out-of-sample prediction accuracy, we only use the penalized logistic regression to predict the exam takers' compliance probability. We produce these probabilities as follows: first, we train the classification algorithm on the SIM survey respondents. Then, we use this model to predict the probability of everyday compliance with Ramadan fasting for individuals with a Moroccan or Turkish migration background. Finally, we assign individuals in the control group (students with no or other migration background from non-Muslim majority countries) a probability of zero.

Figures A6 and A7 in the appendix show the coefficients of the models. The distance to the closest mosque is the strongest predictor in both models. This is consistent with the idea that the distance to mosque represents a strong proxy for the level of religiosity of individuals. Figure 6 shows the probabilities obtained with the two different models. The green line represents the distribution of the predicted probabilities of the Turkish background specific model and the red one of the Moroccan background specific model. We observe that the probabilities assigned to individuals with a Moroccan background are much more concentrated than the probabilities of individuals with a Turkish background, that show more variation. The vertical red line indicates the threshold that we use in our main specification

Figure 6: Predicted Ramadan compliance probabilities



Notes: Density distribution of the Ramadan compliance probability as predicted by the ML model by migration background (Moroccan BkG in red and Turkish BkG in green). The red vertical line defines the threshold for a student being classified as having a high Ramadan compliance probability.

to consider a student as treated (having a high probability of Ramadan compliance). This threshold is defined at 0.62074, which corresponds to the top quartile of the predicted probability among students with a Turkish migration background (who have a lower probability than the students with a Moroccan migration background). The sensitivity of the results to the use of alternative thresholds, as well as the use of the probability itself as a treatment intensity variable, is extensively investigated in our robustness analysis.

Table A4 compares the different compositions of the treatment group between our main specification, students with a high probability of Ramadan compliance, and the alternative specifications: i) individuals with Moroccan background considered as treated, and ii) the continuous probability of Ramadan compliance used as treatment intensity variable. If we use our main specification, students with a high Ramadan probability, our treatment group includes practically all students with a Moroccan migration background are treated, and 25% of the students with a Turkish migration background. By using the alternative specification of Moroccan migration background as treatment group, all students with a Moroccan migration background are treated and we drop students with a Turkish migration background. Using continuous treatment intensity, the average Ramadan probability of students with a Moroccan migration background is 0.88 and for students with a Turkish migration background 0.57. In all treatment specifications we exclude students that come from a Muslim-majority country (see Table A13) not being Morocco or Turkey from our sample. We do not have information on their exact Ramadan compliance probability so cannot use them as treated, but do know that they might observe Ramadan, so neither want to use them as control group.

5 The Impact of Ramadan on Student Performance

5.1 Main Results

In this section, we present the results for the effect of Ramadan on secondary school outcomes, based on equation (1). Panel A of Figure 7 shows that the standardized final (central) exam grade of students with a high Ramadan probability drops significantly in the years that the exams took place during Ramadan.[25] The size of the effect is 5.9 (in 2018) to 6.6% (in 2019) of a standard deviation, while the lack of evidence for pretrends in the years before 2018 confirms the common trends hypothesis.

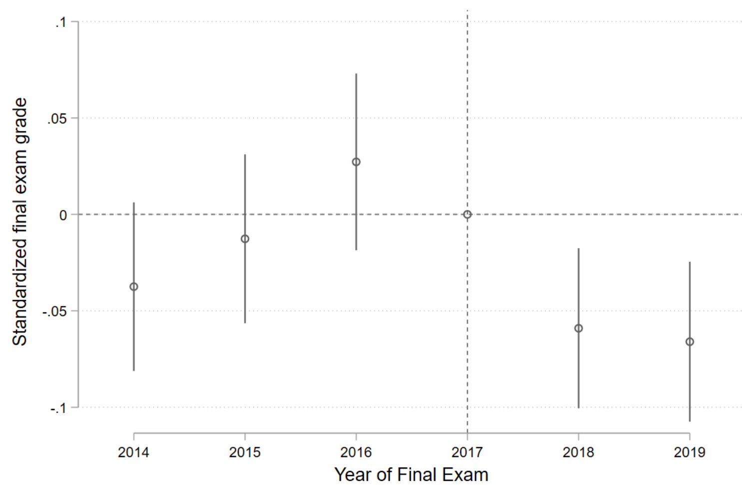
As the final exam grade of students with a high Ramadan probability is lower during Ramadan, the pass rate of those students could be impacted as a consequence, as they are more likely to fail their final exams. Panel B of Figure 7 shows a small drop in pass rate for students with a high Ramadan probability in 2018 and 2019, which suggests that students with a high Ramadan probability might be slightly less likely to pass their final exams.

Table 1 shows the simple DiD estimates based on equation (2) with students with a high

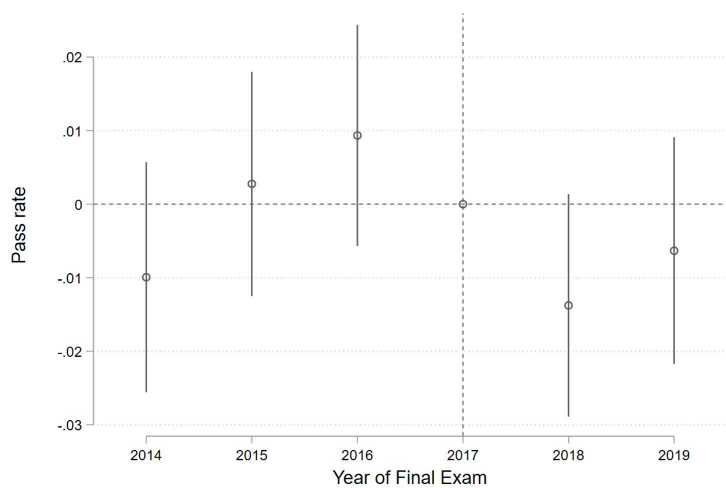
[25] Results are robust to using Moroccan migration background or continuous Ramadan probability as treatment instead of high Ramadan probability. We have also checked whether Ramadan observing students avoid the final exams by not showing up, but find no evidence for this.

Figure 7: Impact of Ramadan on secondary education outcomes of students with a high Ramadan compliance probability

A: Standardized final exam grade



B: Pass rate



Notes: The figures plot the point estimates and 99% confidence intervals for the coefficients (β_t) that are estimated using our main specification (equation (1)). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. The exams in years 2014 to 2017 did not overlap with Ramadan (and hence we would not expect an effect here), while exams in 2018 and 2019 did overlap with Ramadan. Final exam grade is measured as the average of the central exam grades standardized with mean zero and standard deviation one by track and exam year. Probability of overall pass is measured by a dummy that indicates whether the student passed the final exams (and graduated from secondary education). All controls are included, as in Table 1. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

Ramadan probability as treated. The final grade of those students decreased by 5.8% of a standard deviation during the years that Ramadan coincided with the final exams (Column 1). The probability that students with a high Ramadan probability graduated from high school also dropped during Ramadan, as they were 8.3% more likely to fail their final exams (Column 2). Table A5 shows that the results are robust to using the alternative specifications.

Table 1: Impact of Ramadan on education outcomes of students with a high Ramadan probability

	Secondary Education Outcomes	
	Final	Pass
	(1)	(2)
Ramadan Exam Years *Treated	- 0.058***	- 0.011***
	(.010)	(.004)
School-Subtrack f.e.	Yes	Yes
Exam year f.e.	Yes	Yes
All controls	Yes	Yes
Mean outcome before	6.10	86.8%
Share treated	4.7%	4.7%
Observations	776,219	776,284

Notes: The table shows point estimates for the coefficient β estimated using equation (2) with students with a high Ramadan probability being treated and students without a Muslim-majority migration background being the control group. Final is the average of the central exam grades standardized with mean zero and standard deviation one by track and final exam year. Pass is a dummy indicating that a student passed the final exams (including retakes) and graduated from secondary education. The difference in number of observations between the analyses with Final and Pass as outcomes can be explained by students not taking the exam during the first period, but during the first or second retake due to e.g. illness. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. Controls included are: gender, age at exam, end of primary school test score (standardized by year), continuous exam grade (standardized by track and exam year), moved up from lower track or failed the final exams last year, number of siblings, a dummy for being Moroccan or Turkish, time spent abroad, parental income, unemployment or illness benefits, and on the neighbourhood level: the share receiving assistance benefits, the average household size, the ratio under 45 years old, fertility rate, share with Turkish migration background, share with Moroccan migration background. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Robustness and Placebos

First, we validate that students in our treatment group are not changing in other respects. In particular, one might be worried that the negative impact can be explained by negative selection in ability over time of the treated students. Therefore, we look at whether the standardized 'continuous' exam grade, as a measure for ability, changes for students in the treatment group compared to the control group in those years that Ramadan overlapped with the final exams. This grade is an average of several decentralized exams that students take in the last two or three years (depending on the track) of high school. Figure A8 shows that, once we control for standardized end of primary school test score (our other proxy for ability), there is no evidence for a compositional change in ability of exam takers with a high Ramadan probability (apart from a small jump in 2014, which is far from our baseline and treated years).[26] Column 1 in Table 2 confirms this.

Table 2: Impact of Ramadan on secondary school outcomes

	Secondary School Outcomes		
	Cont-inuous	Improved	Final incl. retakes
	(1)	(2)	(3)
Ramadan Exam Years	0.019	0.007	- 0.059***
***High Ramadan Proba**	(.012)	(0.005)	(.010)
School-Subtrack f.e.	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes
All controls	Yes	Yes	Yes
Mean outcome before	6.27	26.3%	6.18
Share treated	4.7%	4.7%	4.7%
Observations	776,284	776,284	776,284

Notes: The table shows point estimates for the coefficient β estimated using equation (2). Treated students in this table are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. Continuous is the average exam grade of the continuous school exams standardized with mean zero and standard deviation one by track and exam year. Improved is a dummy that takes value 1 if the student improved his or her final exam grade in a retake. Final incl. retakes is the average final exam grade of the central exam standardized with mean zero and standard deviation one by track and exam year including the grades of the first and second retake. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included as in Table 1. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, we check whether students can compensate for the negative impact of Ramadan

[26] Figure A4 does show evidence for a compositional change in ability of exam takers with a high Ramadan probability when we take the standardized end of primary school test score as an outcome variable. Hence, it is very important to control for ability. This test, which takes the continuous exam grade as an outcome variable while controlling for the standardized end of primary school test score, shows that conditional on this earlier test score, there is no evidence of a compositional change.

by choosing to retake an exam to improve their grade in a specific topic. Column 2 of Table 2 shows that there is only a very small (and insignificant) increase in the share of students with a high Ramadan probability that improve their final grade with a retake during the years in which the final exam overlapped with Ramadan. As a consequence, the impact of Ramadan on the final grade including retakes is almost identical to the impact on the final grade excluding retakes (Column 3). Hence, students do not compensate for the negative impact of Ramadan with retakes. Possible explanations for this lack of compensation might be that students can only take one retake, which does not leave them a lot of room to improve their average over all final exams, and fasting might have a longer term impact on performance via for example loss in study time, so that it also affects the exam retaken a few weeks after Ramadan.

Third, we check the robustness of the results by i) including sibling fixed effects; and ii) (extended) family fixed effects. We link each exam taker to its siblings and cousins and compare exam takers within families to take out any family fixed effects. As not every exam taker has a sibling or cousin that also took an exam in 2014-2019, and as we are not able to identify families for all exam takers (for exam takers with a migration background we can only observe family if their grandparents also migrated to the Netherlands),[27] the number of observations drops with approximately 47.4% when including sibling fixed effects and 36.4% when including (extended) family fixed effects. Table A6 shows that our results are robust to using sibling and family fixed effects (the effect size even becomes slightly larger).

Fourth, we conduct two placebo in treatment group tests with groups that are similar in socioeconomic characteristics to our treatment group, but that we do not expect to be Muslim. We compute the Ramadan compliance probability, using both the within-Turk machine learning model and the within-Moroccan machine learning model, for students who do not have a migration background from a Muslim-majority country and hence should not observe Ramadan. We use non-Muslim students with a high placebo Ramadan probability as treat-

[27] See Data Appendix for more details on the creation of family links.

ment group. Using these placebo Ramadan probabilities, we do not find any evidence of a change in secondary education outcomes when Ramadan overlapped with the final exams (see Table 3). The results of these placebo tests provide reassurance that the decrease in secondary education outcomes we find when Ramadan overlaps with final exams for students that are likely to observe Ramadan is due to Ramadan and not due to another shock that affected students with a migration background or a socioeconomically disadvantaged background.

Table 3: Placebo impact of Ramadan on secondary school outcomes

	Secondary School Outcomes	
	Final	Pass
	(1)	(2)
Panel A: Ramadan Probability Based on Within-Turk ML Model		
Ramadan Exam Years	- 0.001	- 0.000
***High Ramadan Proba**	(.008)	(.003)
Mean exam grade	6.02	84.7%
Share treated	4.9%	4.9%
Observations	707,501	707,561
School-Subtrack f.e.	Yes	Yes
Exam year f.e.	Yes	Yes
All controls	Yes	Yes
Panel B: Ramadan Probability Based on Within-Moroc ML Model		
Ramadan Exam Years	0.013	- 0.000
***High Ramadan Proba**	(.008)	(.003)
Mean exam grade	6.13	87.4%
Share treated	4.9%	4.9%
Observations	707,501	707,561
School-Subtrack f.e.	Yes	Yes
Exam year f.e.	Yes	Yes
All controls	Yes	Yes

otes: The table shows point estimates for the coefficient β estimated using equation (2). In Panel A, 'treated' students are defined as students with a non-Muslim migration background or without a migration background that are classified as high Ramadan probability according to predictions made by our within-Turk ML model. In Panel B, 'treated' students are defined as students with a non-Muslim migration background or without a migration background that are classified as high Ramadan probability according to predictions made by our within-Moroccan ML model. Final is the average of the central exam grades standardized with mean zero and standard deviation one by track and final exam year. Pass is a dummy indicating that a student passed the final exams and graduated from secondary education. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table 1. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Lastly, we replicate our main analysis for students with a high Ramadan probability by setting different thresholds for the Ramadan probability being high. Figure A9 shows that our results are robust to using different thresholds and get stronger the higher we set the threshold.

6 Heterogeneity Analysis

6.1 Student and school characteristics, peer effects

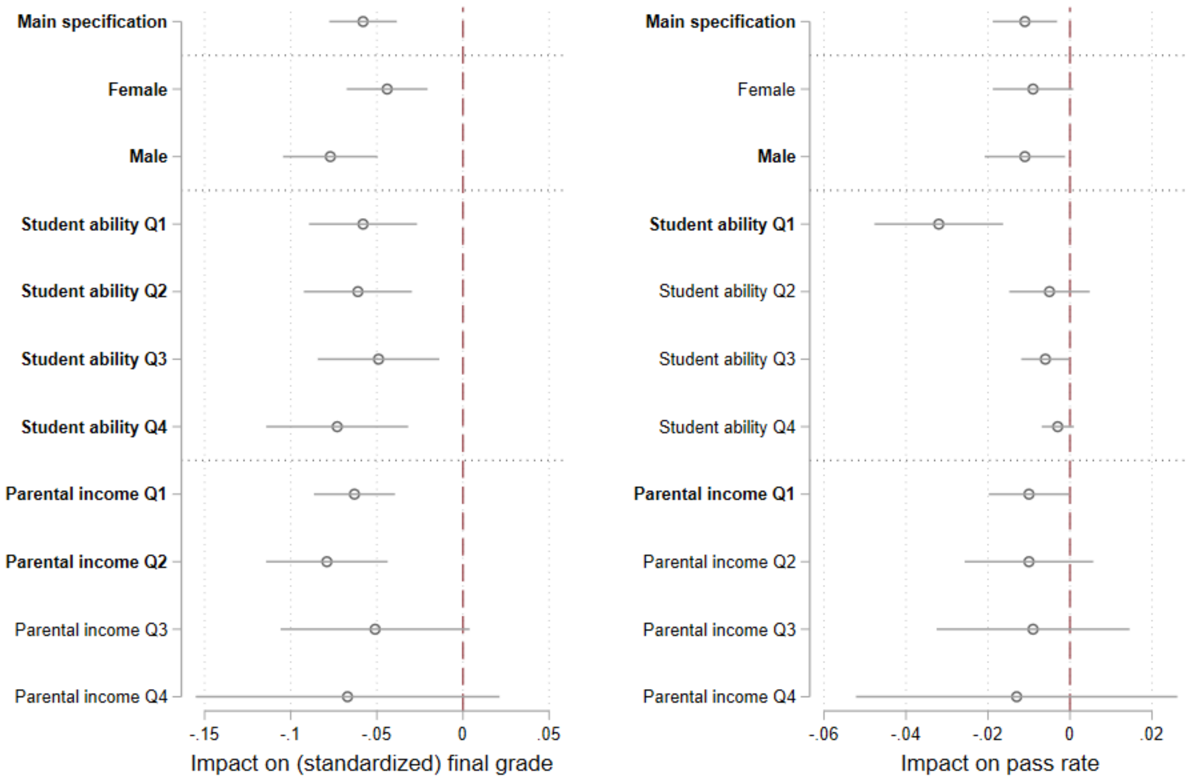
In this section, we explore who is impacted most by Ramadan by examining heterogeneity of the effect by student and school characteristics. First, we check whether the effect on secondary education outcomes differ by gender. We find a very similar impact of Ramadan on the final grade and pass rate of female and male students, with a slightly larger impact on the final grade of male students (see Figure 8).

Second, we check whether the effect on final central exam grade and pass rate differs for high and low performers at the individual and school level.[28] At the individual level, we measure ability with the continuous exam grade and find no heterogeneity by student ability on the impact on the final grade, but we do find heterogeneity in the impact on the pass rate (see Figure 8). The effect of Ramadan on the pass rate is (naturally) driven by the 25% worst performing students. Those students, who are already on the margin of graduating, are 11% more likely to fail their exams and not graduate if the final exams overlap with Ramadan.

At the school level, we measure mean school performance by the average final exam grade of the years before Ramadan overlapped with the final exams. The effect seems to be slightly larger for the lowest performing schools, and there is no impact at all on Ramadan compliant

[28] Results by track show a negative coefficient in each track, but its size and precision varies. While in pre-university education we do not find a significant negative impact, in senior general secondary education we do find a large significant negative impact, and in pre-vocational secondary education we also find a (slightly smaller) significant negative impact. The share treated varies a lot across tracks, from 1.6% in pre-university education to more than 6.5% in pre-vocational secondary education.

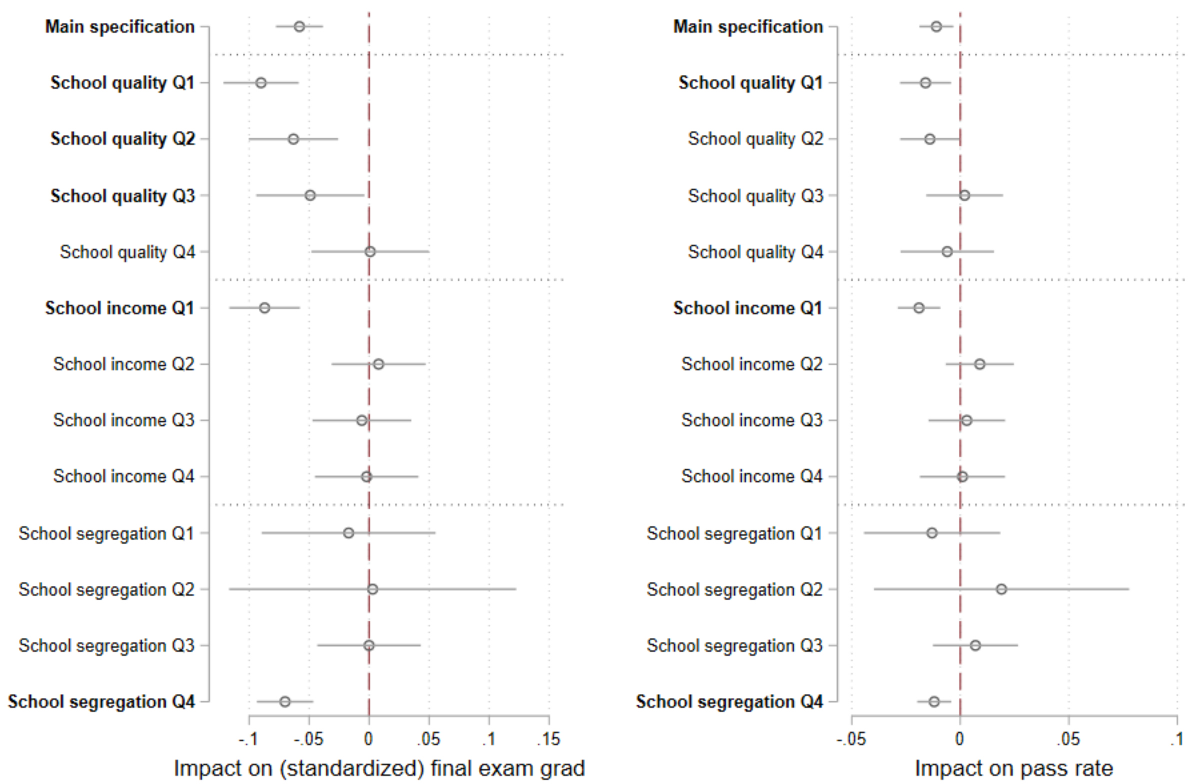
Figure 8: Heterogeneity in impact of Ramadan on secondary education outcomes of high probability Ramadan students by student characteristics



Notes: The figures plot the point estimates and 95% confidence intervals for the coefficient β that is estimated using equation (2), split by student characteristics. Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. The first row shows our main effect for the full sample of students. The second and third row show the main effect for female and male students separately. The fourth to seventh rows show the effect for students split into four quartiles by ability, which is measured by their standardized continuous exam score. The eight to eleventh rows show the effect for students split into four quartiles by parental income, measured in the year before the exam took place. Final exam grade is measured as the average of the central exam grades standardized with mean zero and standard deviation one by track and exam year. Probability of overall pass is measured by a dummy that indicates whether the student passed the final exams (and graduated from secondary education). Labels on the y-axis are in bold when the coefficient is significant on a 5% significance level. All controls are included, as in Table 2. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

students in the top 25% best performing schools (see Figure 9). Similarly, we find the largest effect on the pass rate (Ramadan compliant students in those schools are 11.3% more likely to not graduate during Ramadan) in the 25% worst performing schools and no impact at all

Figure 9: Heterogeneity in impact of Ramadan on secondary education outcomes of high probability Ramadan students by school characteristics



Notes: The figures plot the point estimates and 95% confidence intervals for the coefficient β that is estimated using equation (2), split by school characteristics. Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. The first row shows our main effect for the full sample of students. The second to fifth rows show the effect for students split into four quartiles by school quality, which is measured by the average standardized final exam grade in the years before Ramadan overlapped with the final exam. The sixth to ninth rows show the effect for students split into four quartiles by school income, measured as the average of parental income decile in schools. The tenth to thirteenth rows show the effect for students split into four quartiles by school segregation, measured as the average share of students with a high Ramadan probability in a school. Final exam grade is measured as the average of the central exam grades standardized with mean zero and standard deviation one by track and exam year. Probability of overall pass is measured by a dummy that indicates whether the student passed the final exams (and graduated from secondary education). Labels on the y-axis are in bold when the coefficient is significant on a 5% significance level. All controls are included, as in Table 2. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

in the 50% best performing schools.

Third, we check whether the effect differs by income at the school and individual level.

While the income of the family of the student does not seem to deferentially affect for the impact of Ramadan (see Figure 8), mean school income (of the student body) clearly does. The impact of Ramadan on final grade and pass rate is fully driven by students with a high Ramadan probability in the 25% poorest schools (see Figure 9).

Finally, we check whether the effect varies by school segregation level, where we measure segregation by the proportion of Ramadan compliant students in a school in the years before Ramadan overlapped with the final exams. We only find a negative impact of Ramadan on the final grades and pass rate of students with a high Ramadan probability in the 25% schools with the highest share of Ramadan compliant students. That we only find a Ramadan impact in the most segregated schools may reflect high self-selection, because compliers go to the same schools. However, it could also stem from peer effects in compliance such that Muslim students would follow Ramadan rules more strictly as there are more fasting students around them.

While anecdotal mentions of 'Ramadan peer effects' is frequent, there is no scientific evidence on its existence. We test for this mechanism by exploiting our DiD setting and using cohort variations in compliance probability. Figure A10 shows that there is a lot of cohort variation beyond school self selection, both before and after 2018. Cohort variation in the share of students with a high Ramadan compliance probability is more random than school variation, as cohorts suffer less from self selection than schools. Using the share of peers with a high Ramadan probability in a student's cohort might therefore gives us a better indication of whether the heterogeneity by segregation is fully driven by self-selection or also by Ramadan peer effects. When we interact our main coefficient of interest with the share of peers with a high Ramadan probability in a student's cohort, we find suggestive evidence of an additional impact of Ramadan on final grades and pass rate when there are more peers around a student that are also likely to fast, which suggests there might be Ramadan peer effects (see Figure A11).

6.2 Effect by Type of Exam Taken

In this section we implement a set of heterogeneity analyses based on the subject level data, meaning that we consider each test (e.g. the exam in math, English, etc.) in the central examination as the outcome variable, rather than the average final central exam grade obtained by a student[29] We first replicate our main analysis using this different data structure.[30] Our results are qualitatively the same obtained in the main specification, but the magnitude of our estimates is smaller (see Table A7).[31]

The availability of data at subject level allows us to show (in Table A9) the results for different groups of exams: languages (reading comprehension tests in Dutch, English, German and French), more memory-based exams (history, geography and biology), and math.[32] Memory-based exams were negatively affected by approximately 4.2% of a standard deviation. The impact on language exams is about 40% smaller (only 2.4% of a standard deviation) than the impact on memory-based exams. On the contrary, we find a very strong effect on math (grades dropped with 8.9% of a standard deviation), which is more than twice as large as the effect on memory-based exams. The first two findings are in line with the literature on sleep deprivation and cognition, that finds that sleep deprivation does negatively affect memory, but is much less likely to affect reasoning and crystallized intelligence (Lim and Dinges, 2010). This suggests that the impact of Ramadan on academic performance could be driven by sleep deprivation, more than by fasting. However, math exams are very

[29]Figure A12 provides an overview of the schedule of these exams.

[30]Given the different structure of the data, the regression employed to estimate the effects at exam subject level has some differences with equation(2). This regression is presented in equation(3) in the appendix.

[31]The smaller magnitude of the results using the subject level data can be explained by the fact that students in the highest track (pre-university education) take more exams than students in other track, causing a larger number of observations for students in the highest track. As a consequence, by focusing on subject level outcomes, we give more weight to students in the highest track, who are less affected by Ramadan exposure. Moreover, the use of a rich set of fixed effects to account for the timing of the exams might explain the smaller magnitude obtained with this specification. Finally, we standardize exams at subject-track-year level, so that for each subject in a certain track in a specific year the mean grade is zero with standard deviation one. This is different from the standardization adopted in our main specification, where we transform the final (comprehensive of all subjects) grade such that it has mean zero and standard deviation one at track-year level.

[32]Note that in this analysis we can only look at a selected sample that actually takes a final exam for one or more of those subjects.

strongly affected, while those exams might also be classified as dependent on reasoning and crystallized intelligence, which would not be in line with the idea that sleep deprivation is mainly driving our effect. Hence, some results point into the direction of sleep deprivation as the main mechanism, while others point into the direction of fasting as the main mechanism. In the following section, we will therefore explore the timing of exams over the weeks and during the day, to further try to explore whether our effect is driven by sleep or by fasting.

7 Mechanisms & Policy Implications

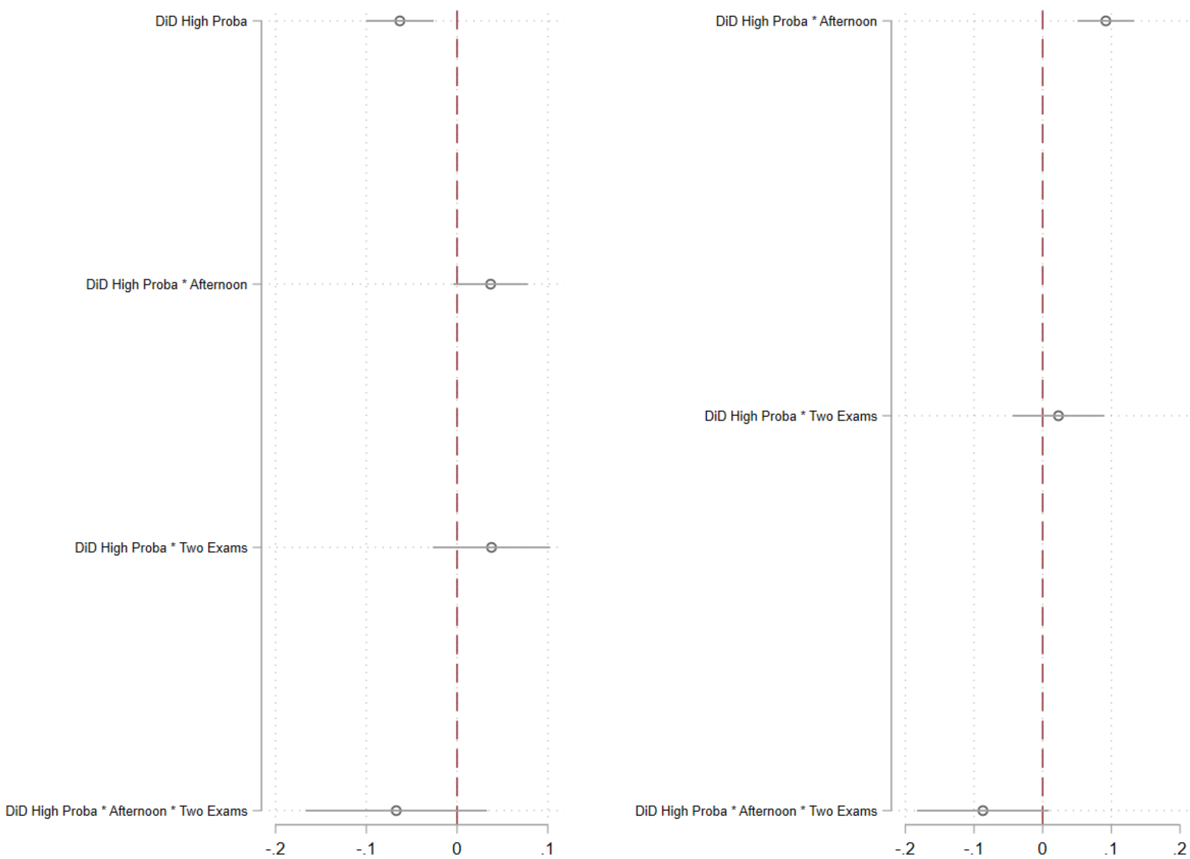
7.1 Fasting or sleep deprivation?

In this section we explore which factors might have contributed the most to the worsening in performance observed among students who are likely to adhere to Ramadan obligations. While fasting is the most obvious factor, other variables might have played a role. Muslims often change their sleep patterns during the Ramadan month (Bahammam, 2003; Bahammam, 2006; Margolis and Reed, 2004). The reason for this is that, to avoid having meals during daylight, Muslims have to eat very late at night and early in the morning. Hence, students might suffer from sleep deprivation during the exams. In addition, the time spent celebrating the holy month of Islam could affect the time that students can dedicate to studying, potentially explaining part of the negative impact of Ramadan exposure on academic performance.

First, we check whether the effect size estimated changes over time. Our assumption is that if the negative effect we find is mainly driven by a lack of sleep, we would see that the effect becomes more negative over time. On the contrary, if the effect is stronger in the first week of Ramadan, than we can attribute the decline in cognitive performance mainly to fasting. We make these assumptions based on the literature on sleep, which clearly shows how sleep deprivation accumulates over time Van Dongen et al. (2003), leading to worse cognitive functioning as the nights affected by lack of sleep accumulates. Our assumption on fasting affecting the first week more than later weeks is based on the anecdotal evidence that the first few days of fasting are those during which the body feels the effects of fasting the most and still has to adapt. To check the empirical relevance of these mechanisms, we decompose the effect of Ramadan on exams grade depending on whether an exam is taken in the first, second or third week of Ramadan. Figure A13 shows that the effect over those three weeks is quite constant. We also report the coefficient of the extra effect of the exact day into Ramadan each exam is taken. This is a well estimated zero, confirming that loss of performance is constant over-time. These results suggests that neither cumulative sleep deprivation nor fasting can alone explain our results. Finally, we also compute an effect for 'week zero' in 2019, to see that the performance at the few exams taken before Ramadan did not drop for complying students. This means that if any anticipation effect is present, for example because of a change of eating regime to transition smoothly to the Ramadan fast, this is not particularly strong.

Second, we check whether the effect is different for exams that take place in the afternoon compared to exams that take place in the morning. We hypothesize that if fasting is the main driver of the Ramadan effect, we would find a smaller effect in the morning, as the students had breakfast a few hours before the exam. On the contrary, if we expect the lack of sleep to be driving the effect, we would expect the effect to be larger in the morning, as students have no time to sleep in. In Table A8 we show that the coefficient capturing the additional effect of exams held in the afternoon is insignificant and close to zero in magnitude, showing that there is no difference in the impact of Ramadan depending on the timing of the exam. However, the previous findings do not take into account that on some days, students have to take an exam both in the morning and in the afternoon. When we consider in our model the interaction between timing of the exam and the number of exams taken in one day, we find different results, shown in Figure 10. Exams that are held in the afternoon, when no exam is scheduled in the morning, are affected the least. We believe that this is driven by the possibility for students to rest in the morning.

Figure 10: Effect on final grade in 2 exams days



Notes: The figure displays the point estimates and 95% confidence intervals for the coefficients estimated using equation (4) with (on the right) and without (on the left) individual fixed effects, based on the subject level data. The dependent variable is the grade obtained in the exam (after retakes) standardized at subject-track-year level. $Afternoon$ is a dummy that takes value 1 if an exam is held in the afternoon. $DiD\ High\ Proba$ is a dummy equal to one if the exam is taken during the Ramadan month by a student with high probability of Ramadan compliance. $Two\ Exams$ is dummy that takes value 1 if an exam is held on a day in which a student has to sit for two exams. The coefficient of interest is the triple interaction between these dummies, which captures the additional impact of sitting an exam in the afternoon after a student already took another exam in the morning. All controls are included, as in Table 2. We include school-track-exam subject and treatment type-exam subject-afternoon-two day exam fixed effects. Robust standard errors are clustered at the school-track-subject-year level.

Finally, the worsening in exams results might be driven by additional time constraints that students face during the Ramadan month. In this month, students often spend more time with their relatives and participate to the celebration of Ramadan. This time cannot be dedicated to preparing for the exams, which eventually could result in a worse outcome.

We check whether this might play a role in two ways. First, we check whether the effect is different depending on whether a student has a small or large family. We hypothesize that a large family might cause more social obligations during Ramadan. We do not find much evidence for this, with only slightly larger effects on outcomes for students with larger families than smaller families (see Table A10). Second, we focus again on the results split by type of subject. If we believe that our results are driven by time constraints that affect the preparation for the exams but not the exams themselves, we would only expect to find an effect on exams that depend a lot on short-term memory and not so much on other subjects (like languages that use reading comprehension tests as final exams or math). However, Table A9 shows that not only short-term memory exams, but also math exams - and language exams to a lesser extent - are affected. Math exams are even affected most of all exams, which suggests that our effect is not (solely) driven by less preparation time due to social obligations. These two findings together suggest that time constraints are not likely to play a large role in driving the negative impact of Ramadan on secondary education outcomes of students with a high Ramadan probability.

7.2 Policy implications

Failing the final exams has longer-term consequences for students, as they have to repeat the year or go into adult education in order to obtain their diploma. Table A11 shows that students that have a high Ramadan probability are 12.4% more likely to repeat the year if their final exam coincided with Ramadan. The costs of this are considerable, as the student will enter the labor market one year later and will therefore suffer from a year of earnings loss. The government also bears more costs if a student fails to graduate: it pays 7500 euro for an additional year of school for a student (Van Vuuren and Van der Wiel, 2017). Students have different options when they fail their exams, depending on their age. At age 18, they are no longer obliged to obtain a so-called "start qualification", which is a senior general secondary education or pre-university education diploma, or a diploma in secondary vocational education. We therefore also split our analysis by whether the student is age 18 on the 1st of September after the final exams, when secondary and tertiary education normally start again. We find that our impact on students repeating the year (or going into adult education is driven by students younger than 18, while students older than 18 are 22.5% more likely to drop out if their final exams overlapped with Ramadan (see Table A12).

Figure A13 shows that in what we have defined as week zero, the week before the beginning of Ramadan, we find no significant change in grades for students with a high Ramadan probability. The most straightforward policy implication would therefore be to move exams entirely so that they do not overlap with and, as there might be longer-term effects of Ramadan, take place before Ramadan. However, as exams take place during a multiple week period, it might not be (politically) feasible to fully avoid the Ramadan month. A second, and milder, policy implication would therefore be to take into account that the impact is largest for students with two exams on one day (see Figure 10). To mitigate the impact of Ramadan on secondary education outcomes, it would therefore be best not to schedule two exams on one day, but to spread them out over time.

8 Conclusion

Using a unique natural experiment combined with high quality administrative data, we causally estimate the impact of Ramadan observance on student performance during high stakes exams. Our main finding is that of a strong and significant negative impact on average grades and pass probability for this already disadvantaged population. Specifically, we find that those most likely to fast obtained grades 5.8% of a standard deviation (SD) lower during Ramadan. These are large effects which are quantitatively comparable to taking an exam during heat waves (1 SD higher temperature, Park, 2022) or during pollution spikes (1.5 SD higher, Ebenstein et al., 2016). In the Dutch context, it is equivalent to having a teacher with 6 years experience before an exam (Gerritsen et al., 2017) or twice the effect of students having legal cannabis access (Marie and Zölitz, 2017).

This drop in performance during Ramadan also strongly impacted pass rates. As a result of the lack of accommodation for the religious obligations of the affected minority students, there was an large increase in the existent achievement gap with their non-Muslim peers of 16.4%. These negative impacts on education outcomes were especially bad for low achieving students and almost entirely driven by those in religiously segregated schools. For those in schools with the highest quartile of Ramadan complying students, the achievement gap increased by a quarter. There were also potential longer term effects for affected students as 22.5% more of those above 18, who could legally drop out, ended up doing so.

Changing the timing of national exams to accommodate religious minority might be the simplest option but has proven practically and politically difficult to implement[33]. In our context, the exploration as to whether the Ramadan impact stemming from fasting or sleeping yields the interesting finding that examination results are especially negatively affected when taking place in the afternoon, following a morning exam. Adapting the exams schedule for this not to happen may provide a practical solution to attenuate Ramadan effects when testing performance in the future.

Another possibility would be for religious leaders to give permission to their members to (exceptionally) adapt to the rules of the country where they live when needed. In our case, it would involve Islamic authorities producing *fatwas* which would allow Ramadan postponements for those taking important tests of their performance, such as high stakes exams. Such exemptions exist for health reasons but not educational ones and those living in non-Muslim countries have received specific guidelines as to how to follow Ramadan in difficult contexts. [34] As such requests need to come from potentially affected Muslims, the clear findings of our paper might help make this happen. There is otherwise recent robust

[33] It has however happened before as, for example, the Seventh Day Adventists were able to force the Brazilian government in 2020, after a long legal battle that took them all the way to the Supreme Court, to allow them to take selective exams ('concursos') on other days than Saturday, which is their most holy day.

[34] Various fatwas were declared in the mid-2010s to give permission to Muslims living in northern countries, such as Iceland where the day could last up to 22 hours, to follow the fasting schedule of other (more southern) countries.

evidence that awareness of the negative impact of Ramadan observance on productivity can lead to changes in the fasting behaviour of those concerned (Wang et al., 2022). There is thus potential for the results presented here to have an an effect when individuals have to decide between their religious obligations and their requirement to perform in specific civic duties.

References

ADOLPHUS, K., C. L. LAWTON, C. L. CHAMP, AND L. DYE (2016): "The effects of breakfast and breakfast composition on cognition in children and adolescents: a systematic review," *Advances in Nutrition*, 7, 590S–612S.

ALMOND, D. AND B. MAZUMDER (2011): "Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy," *American Economic Journal: Applied Economics*, 3, 56–85.

ALMOND, D., B. MAZUMDER, AND R. VAN EWIJK (2015): "In utero Ramadan exposure and children's academic performance," *The Economic Journal*, 125, 1501–1533.

BAHAMMAM, A. (2003): "Sleep pattern, daytime sleepiness, and eating habits during the month of Ramadan," *Sleep and Hypnosis*, 5, 165–174.

——— (2006): "Does Ramadan fasting affect sleep?" *International journal of clinical practice*, 60, 1631–1637.

BAHAMMAM, A. S., A. M. ALASEEM, A. A. ALZAKRI, AND M. M. SHARIF (2013): "The effects of Ramadan fasting on sleep patterns and daytime sleepiness: An objective assessment," *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, 18, 127.

Bertoli, P., V. Grembi, and J. Vall-Castello (2022): "The Ramadan effect in the workplace," .

Blanden, J., M. Doepke, and J. Stuhler (2022): "Educational inequality," Tech. rep., National Bureau of Economic Research.

Campante, F. and D. Yanagizawa-Drott (2015): "Does religion affect economic growth and happiness?" *The Quarterly Journal of Economics*, 130, 615–658.

Colussi, T., I. E. Isphording, and N. Pestel (2021): "Minority salience and political extremism," *American Economic Journal: Applied Economics*, 13, 237–71.

Csipo, T., A. Lipecz, C. Owens, P. Mukli, J. W. Perry, S. Tarantini, P. Balasubramanian, Á. Nyúl-Tóth, V. Yabluchanska, F. A. Sorond, et al. (2021): "Sleep deprivation impairs cognitive performance, alters task-associated cerebral blood flow and decreases cortical neurovascular coupling-related hemodynamic responses," *Scientific reports*, 11, 20994.

Curcio, G., M. Ferrara, and L. De Gennaro (2006): "Sleep loss, learning capacity and academic performance," *Sleep medicine reviews*, 10, 323–337.

Ebenstein, A., V. Lavy, and S. Roth (2016): "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution," *American Economic Journal: Applied Economics*, 8, 36–65.

Gerritsen, S., E. Plug, and D. Webbink (2017): "Teacher quality and student achievement: Evidence from a sample of Dutch twins," *Journal of applied econometrics*, 32, 643–660.

Ghayour Najafabadi, M., L. Rahbar Nikoukar, A. Memari, H. Ekhtiari, and S. Beygi (2015): "Does ramadan fasting adversely affect cognitive function in young females?" *Scientifica*, 2015.

Hornung, E., G. Schwerdt, and M. Strazzeri (2023): "Religious practice and student performance: Evidence from Ramadan fasting," *Journal of economic behavior & organization*, 205, 100–119.

Huijink, W. (2018): *De religieuze beleving van moslims in Nederland.*

Kökkizil, M. (2022): "Religiosity and gender gap in STEM performances," *Unpublished manuscript.*

Lim, J. and D. F. Dinges (2010): "A meta-analysis of the impact of short-term sleep deprivation on cognitive variables," *Psychological bulletin*, 136, 375–389.

Lugo, L., A. Cooperman, E. O'Connell, and S. Stencel (2011): "The future of the global Muslim population," *Pew Research Center, USA*, 1–209.

Margolis, S. A. and R. L. Reed (2004): "Effect of religious practices of Ramadan on sleep and perceived sleepiness of medical students," *Teaching and learning in medicine*, 16, 145–149.

Marie, O. and U. Zölitz (2017): ""High" achievers? cannabis access and academic performance," *The Review of Economic Studies*, 84, 1210–1237.

Montero, E. and D. Yang (2022): "Religious festivals and economic development: Evidence from the timing of mexican saint day festivals," *American Economic Review*, 112, 3176–3214.

Oosterbeek, H. and B. van der Klaauw (2013): "Ramadan, fasting and educational outcomes," *Economics of Education Review*, 34, 219–226.

Park, R. J. (2022): "Hot temperature and high-stakes performance," *Journal of Human Resources*, 57, 400–434.

Schofield, H. (2020): "Ramadan fasting and agricultural output," *Unpublished manuscript.*

Tian, H.-H., A.-R. Aziz, W. Png, M. F. Wahid, D. Yeo, and A.-L. C. Png (2011): "Effects of fasting during Ramadan month on cognitive function in Muslim athletes," *Asian Journal of Sports Medicine*, 2, 145.

Van Dongen, H. P., G. Maislin, J. M. Mullington, and D. F. Dinges (2003): "The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation," *Sleep*, 26, 117–126.

Van Ewijk, R. (2011): "Long-term health effects on the next generation of Ramadan fasting during pregnancy," *Journal of health economics*, 30, 1246–1260.

Van Vuuren, D. and K. Van der Wiel (2017): *Zittenblijven in het primair en voortgezet onderwijs. Een inventarisatie van de voor- en nadelen*, CPB Policy Brief 2015/01.

Wang, A., S. Wang, and X. Ye (2022): "When Information Conflicts with Obligations: the Role of Motivated Cognition," Tech. rep., National Bureau of Economic Research.

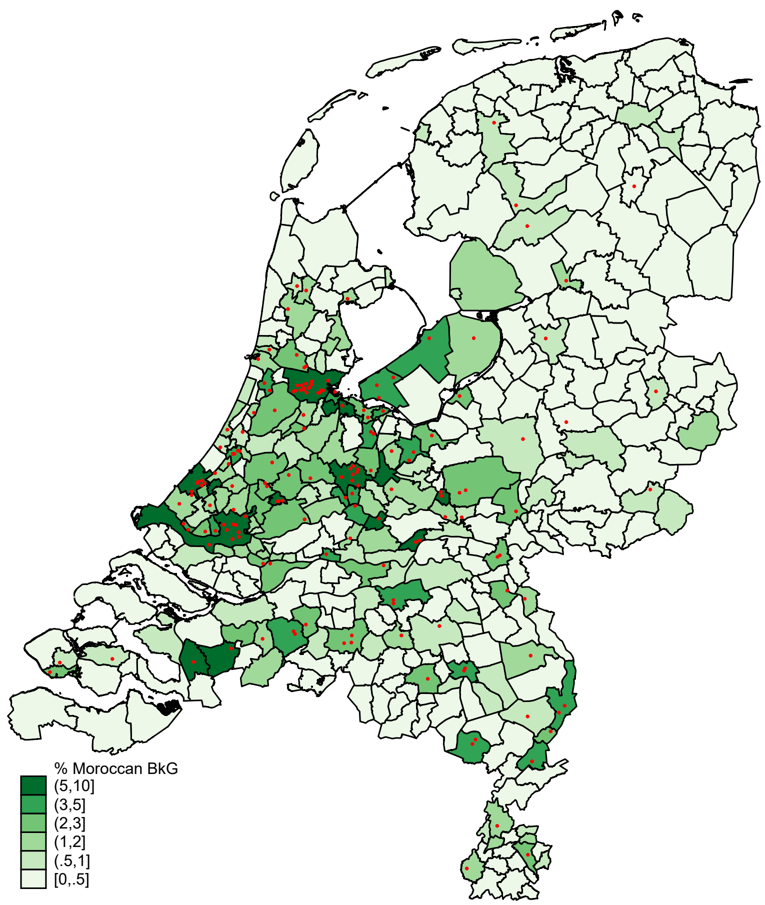
Wittbrodt, M. T. and M. Millard-Stafford (2018): "Dehydration impairs cognitive performance: a meta-analysis," *Med Sci Sports Exerc*, 50, 2360–2368.

Yasin, W. M., M. M. A. K. Khattak, N. M. Mamat, and W. A. M. A. Bakar (2013): "Does religious fasting affect cognitive performance?" *Nutrition & Food Science*.

Appendix

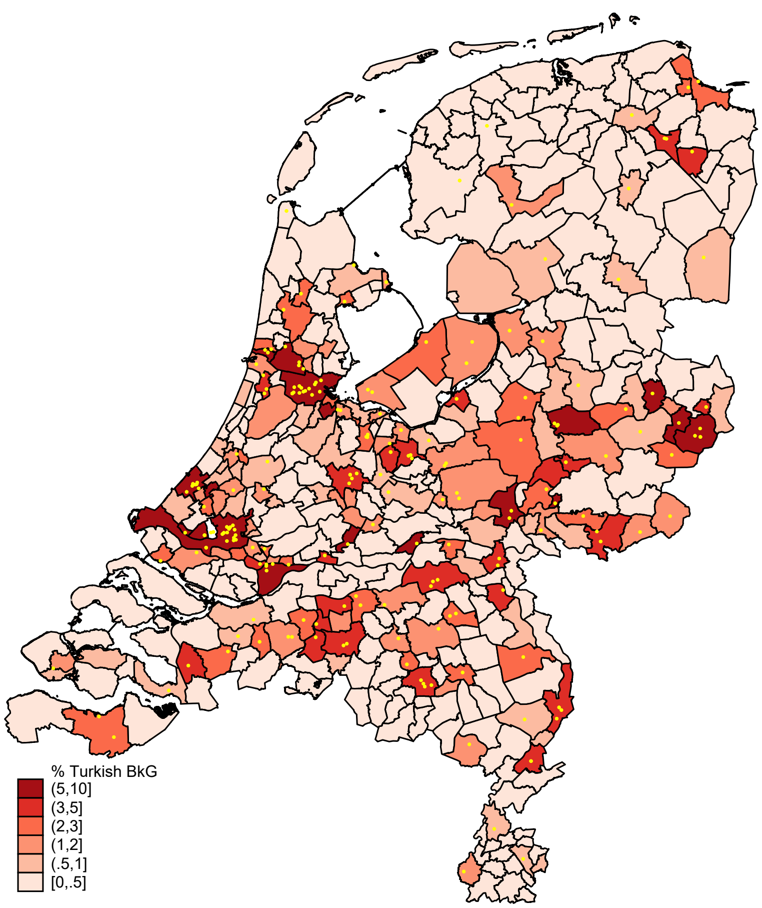
A Additional Tables and Figures

Figure A1: The share of municipalities with a Moroccan migration background and the location of mosques in the Netherlands



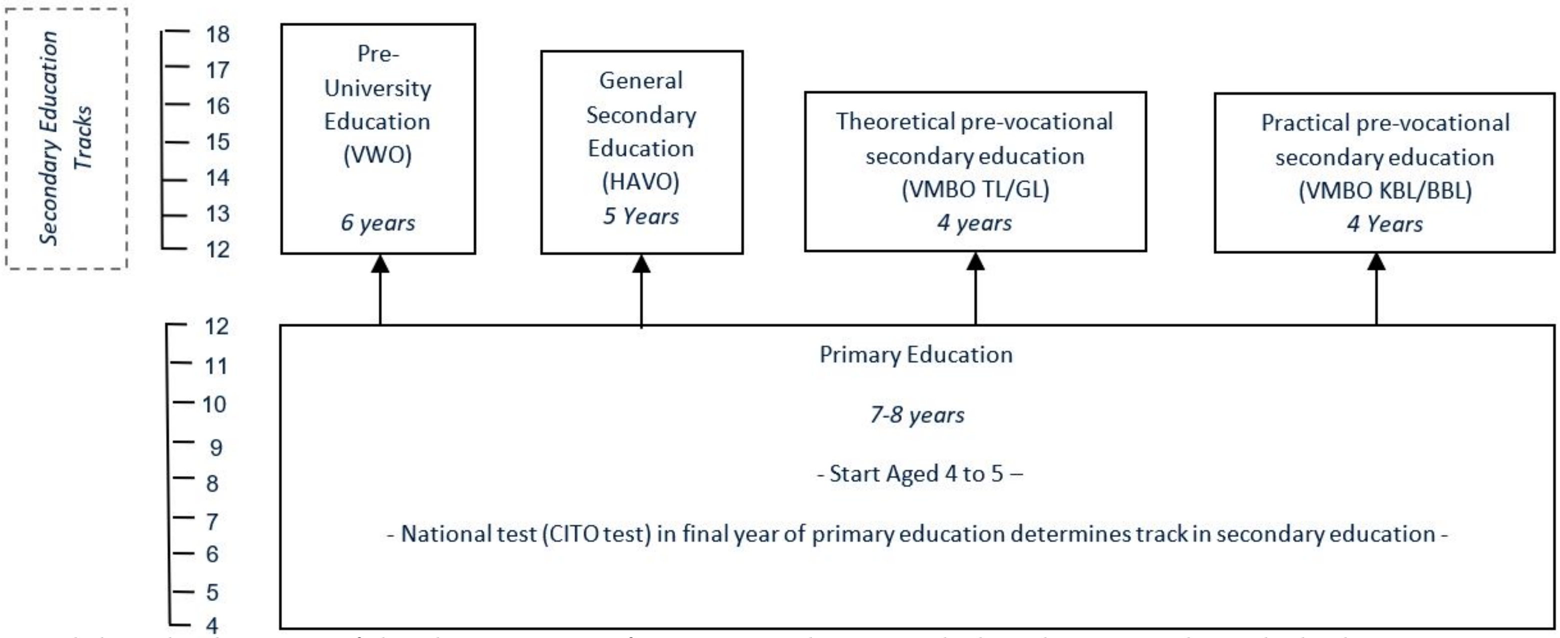
Notes: Municipality level proportion of people with a Moroccan migration background in 2017 in the Netherlands. The locations of mosques where the language of the sermon is Moroccan are indicated with red dots. Source: CBS Statline, statline.cbs.nl; Moskeewijzer, https://moskeewijzer.nl/.

Figure A2: The share of municipalities with a Turkish migration background and the location of mosques in the Netherlands



Notes: Municipality level proportion of people with a Turkish migration background in 2017 in the Netherlands. The locations of mosques where the language of the sermon is Turkish are indicated with yellow dots. Source: CBS Statline, statline.cbs.nl; Moskeewijzer, https://moskeewijzer.nl/.

Table A1: Dutch education system



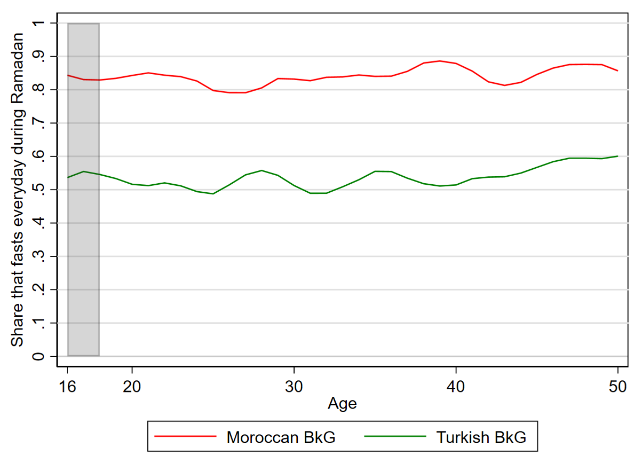
Notes: Detailed graphical overview of the education system from primary education to higher education in the Netherlands.

Table A2: Compliance with measures of Muslim religiosity by migration background

	SIM Respondents	
	Moroccan Migration BkG	Turkish Migration BkG
	(1)	(2)
Prays At Least Every Day	0.796	0.376
	(.403)	(.484)
Visits Mosque At Least Every Week	0.367	0.344
	(.482)	(.475)
Eats Halal Every Day	0.877	0.732
	(.328)	(.443)
Finds Religion Important	0.901	0.801
	(.299)	(.399)
Islam Rules Should be Followed Completely	0.601	0.501
	(.490)	(.500)
Observations	2,677	2,910

Notes: Self-reported compliance with measures of Muslim religiosity based on SIM (*Survey Integratie Minderheden*) data of respondents with a Moroccan migration background on the left and Turkish migration background on the right. Source: SIM data in 2006, 2011, 2015 and 2020.

Figure A3: Fasting behavior of SIM respondents with a Moroccan or Turkish BkG by age (SIM survey 2006-2020)



Notes: Graphical representation of the average share of respondents with a Moroccan migration background (in red) or Turkish migration background (in green) that report to fast everyday during Ramadan by age at the time of the survey. Source: SIM (*Survey Integratie Minderheden*) data on 2006, 2011, 2015 and 2020.

Table A3: Characteristics of SIM Survey Respondents and Exam Takers Sample

	Moroccan Migration BkG		Turkish Migration BkG	
	SIM	Exam takers	SIM	Exam takers
	(1)	(2)	(3)	(4)
Age	38.180	16.496	38.328	16.544
	(15.317)	(0.976)	(15.225)	(0.983)
Ratio under 45 neighbourhood	0.615	0.608	0.609	0.597
	(0.089)	(0.090)	(0.091)	(0.090)
Birth over fertility neighbourhood	0.028	0.028	0.027	0.027
	(0.007)	(0.008)	(0.008)	(0.009)
Average household size neighbourhood	2.107	2.165	2.145	2.188
	(0.303)	(0.316)	(0.321)	(0.310)
Share assistance benefits neighbourhood	0.059	0.058	0.058	0.056
	(0.034)	(0.034)	(0.035)	(0.035)
Share Moroccan BkG neighbourhood	0.107	0.114	0.076	0.074
	(0.091)	(0.096)	(0.078)	(0.082)
Share Turkish BkG neighbourhood	0.078	0.079	0.111	0.102
	(0.075)	(0.073)	(0.092)	(0.085)
Distance to Moroccan mosque	2,068.481	1,975.036	3,393.643	3,988.605
	(3,621.228)	(3,543.813)	(5,882.818)	(6,447.188)
Distance to Turkish Mosque	2,203.873	2,270.903	1,485.046	1,741.672
	(3,001.709)	(2,881.584)	(1,889.499)	(2,140.501)
Observations	2,104	30,413	2,208	29,546

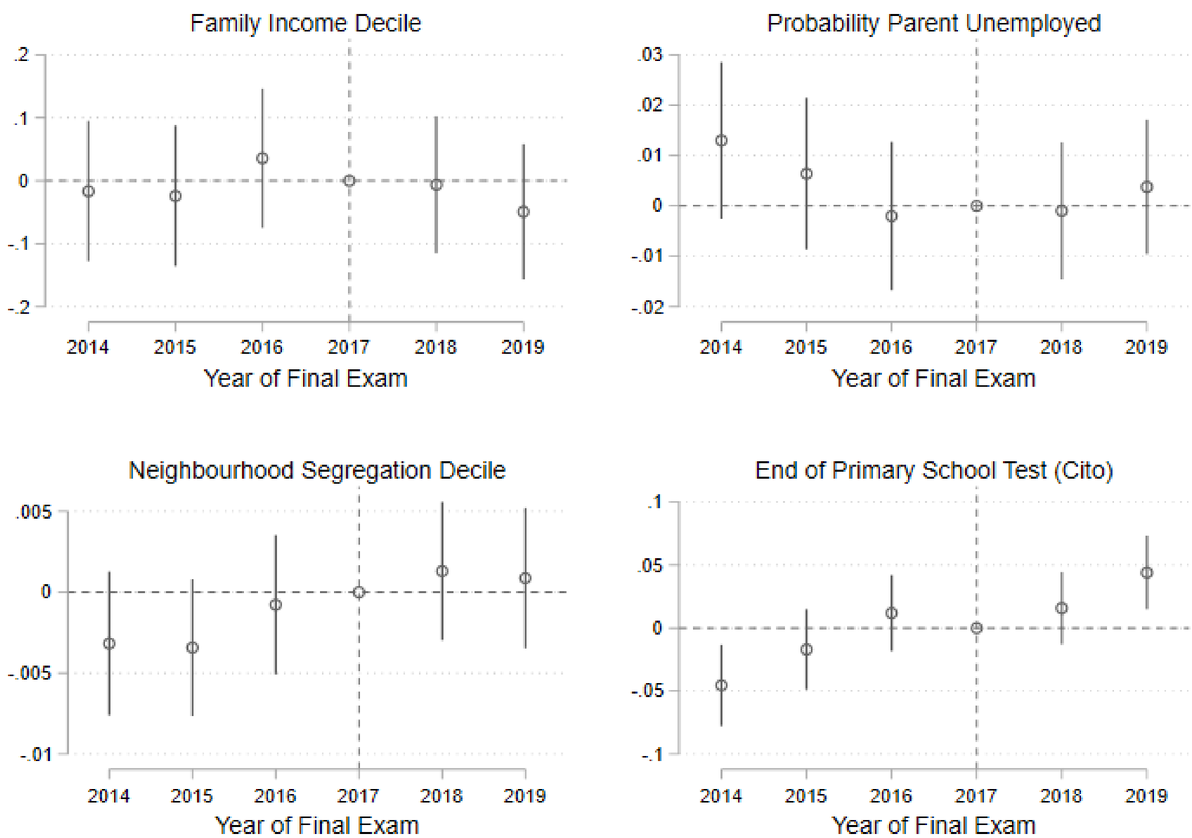
Notes: Balance table of characteristics of the sample of SIM (Survey Integratie Minderheden) respondents (in 2006-2020) with a Turkish or Moroccan migration background (that we use to train our ML model to predict Ramadan compliance with) and the sample of exam takers (in 2014-2019) with a Turkish or Moroccan migration background that we use in our analysis.

Table A4: Composition of treated group

	Students with	
	Moroccan Migration BkG	Turkish Migration BkG
	(1)	(2)
Moroc BkG	1.000	0.000
High Ramadan Proba	0.998	0.250
Continuous Ramadan Proba	0.883	0.565
Observations	30,413	29,546

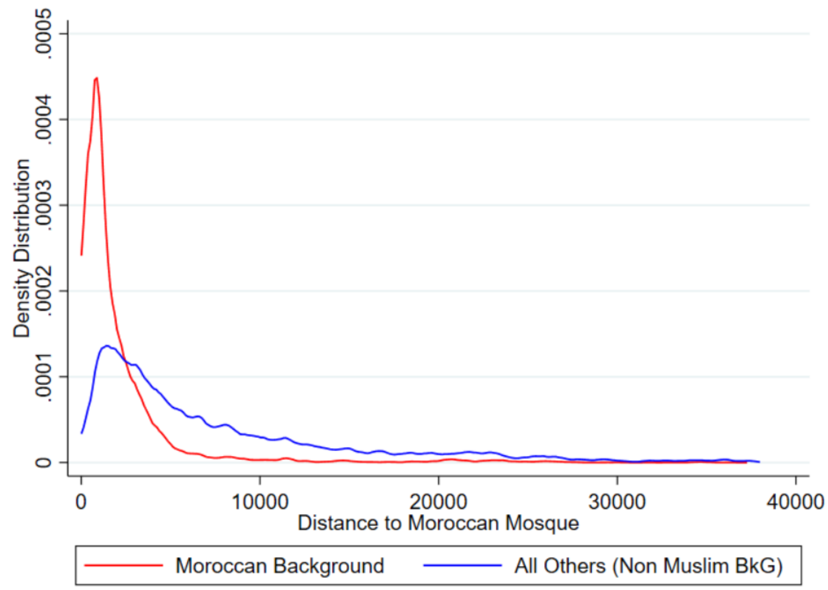
Notes: Share of students in our sample by migration background that are ‘treated’ (compliant to the Ramadan fast) by treated specification.

Figure A4: DiD with student characteristics as outcomes and students with high Ramadan probability as treated group



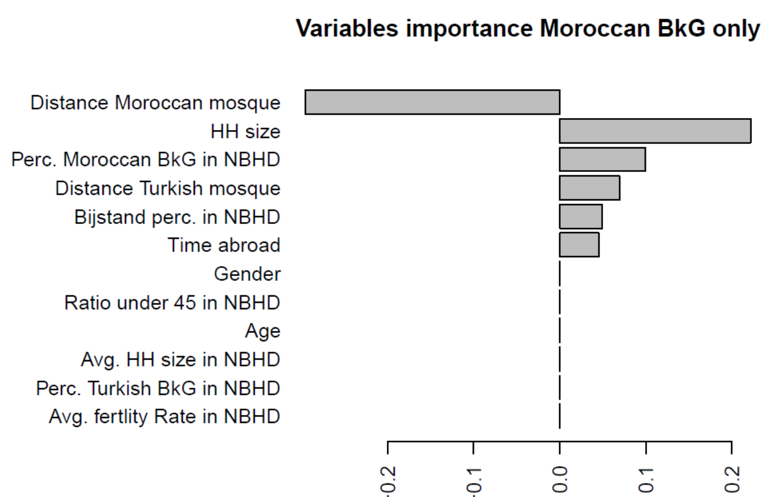
Notes: The figures plot the point estimates and 99% confidence intervals for the coefficients (β_t) that are estimated using our main specification (equation (1)). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. Family income is divided into deciles, probability parent unemployed is measured as having at least one parent being unemployed, neighbourhood segregation is measured as the share of people with a Moroccan or Turkish migration background in your neighbourhood and divided into deciles, the end of primary school test (CITO) is standardized by CITO year with mean zero and standard deviation one. Controls included are: gender, age at exam, end of primary school test score (standardized by year), continuous exam grade (standardized by track and exam year), moved up from lower track or failed the final exams last year, number of siblings, a dummy for being Moroccan or Turkish, time spent abroad, parental income, unemployment or illness benefits, and on the neighbourhood level: the share receiving assistance benefits, the average household size, the ratio under 45 years old, fertility rate, share with Turkish migration background, share with Moroccan migration background. We exclude a control variable if it overlaps with the outcome variable. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

Figure A5: Distance to Moroccan mosque



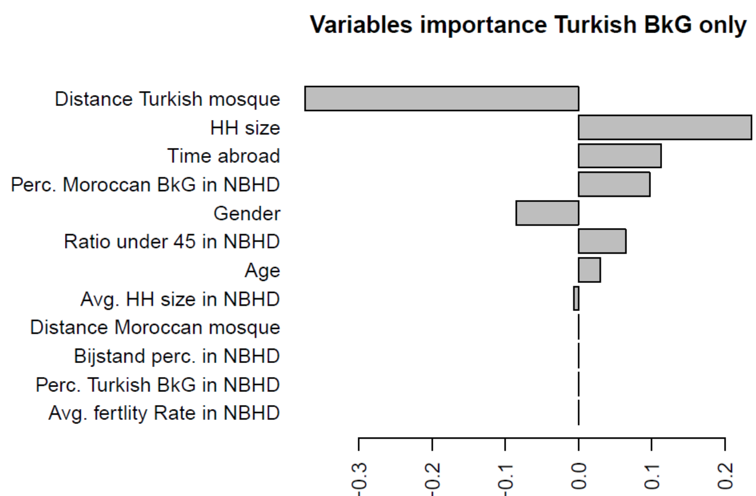
Notes: Density distribution of the distance from the neighbourhood of the student to the neighbourhood of the Moroccan mosque for students with a Moroccan migration background (red) and students with a non-Muslim migration background or no migration background (blue). Source: Moskeewijzer, https://moskeewijzer.nl/.

Figure A6: Penalized Logit model coefficients: Moroccan BkG



Notes: Importance of variables for prediction of Ramadan compliance probability as predicted with a penalized logit ML model including only SIM respondents with a Moroccan migration background. Variables used in the model are the share of the respondent's neighbourhood having a Moroccan neighbourhood or Turkish neighbourhood, age of the respondent, time spent abroad, gender of the respondent, average household size in the respondent's neighbourhood, distance to Moroccan or Turkish mosque, share of the neighbourhood using welfare (*bijstand*), fertility in the neighbourhood and the share of the respondent's neighbourhood below 45 years old.

Figure A7: Penalized Logit model coefficients: Turkish BkG



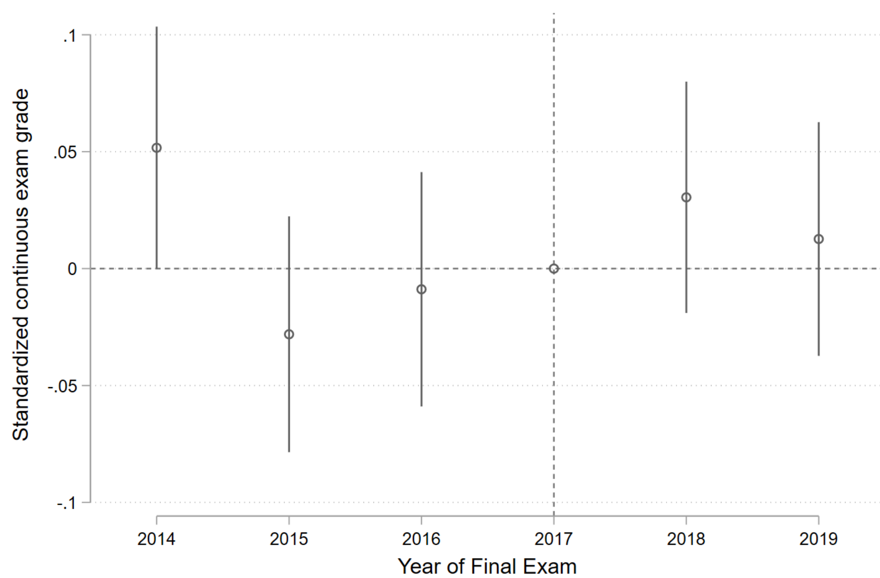
Notes: Importance of variables for prediction of Ramadan compliance probability as predicted with a penalized logit ML model including only SIM respondents with a Turkish migration background. Variables used in the model are the share of the respondent's neighbourhood having a Moroccan neighbourhood or Turkish neighbourhood, age of the respondent, time spent abroad, gender of the respondent, average household size in the respondent's neighbourhood, distance to Moroccan or Turkish mosque, share of the neighbourhood using welfare (*bijstand*), fertility in the neighbourhood and the share of the respondent's neighbourhood below 45 years old.

Table A5: Impact of Ramadan on education outcomes of students with a high Ramadan probability

	Secondary School Outcomes					
	High Ramadan Proba		**Moroc BkG**		**Cont. Ramadan Proba**	
	Final	Pass	Final	Pass	Final	Pass
	(1)	(2)	(3)	(4)	(5)	(6)
Ramadan Exam Years	- 0.058***	- 0.011***	- 0.059***	- 0.009**	- 0.052***	- 0.011**
***Treated**	(.010)	(.004)	(.011)	(.004)	(.011)	(.004)
School-Subtrack f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes	Yes	Yes	Yes
All controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome before	6.10	86.8%	6.13	87.4%	6.08	86.0%
Share treated	4.7%	4.7%	3.9%	3.9%	5.5%	5.5%
Observations	776,219	776,284	746,697	746,742	776,219	776,284

Notes: The table shows point estimates for the coefficient β estimated using equation (2) for the 3 different treatment group specifications. Final is the average of the central exam grades standardized with mean zero and standard deviation one by track and final exam year. Pass is a dummy indicating that a student passed the final exams (including retakes) and graduated from secondary education. The difference in number of observations between the analyses with Final and Pass as outcomes can be explained by students not taking the exam during the first period, but during the first or second retake due to e.g. illness. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. Controls included are: gender, age at exam, end of primary school test score (standardized by year), continuous exam grade (standardized by track and exam year), moved up from lower track or failed the final exams last year, number of siblings, a dummy for being Moroccan or Turkish, time spent abroad, parental income, unemployment or illness benefits, and on the neighbourhood level: the share receiving assistance benefits, the average household size, the ratio under 45 years old, fertility rate, share with Turkish migration background, share with Moroccan migration background. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A8: Impact of Ramadan on continuous exam grade of students with a high Ramadan probability



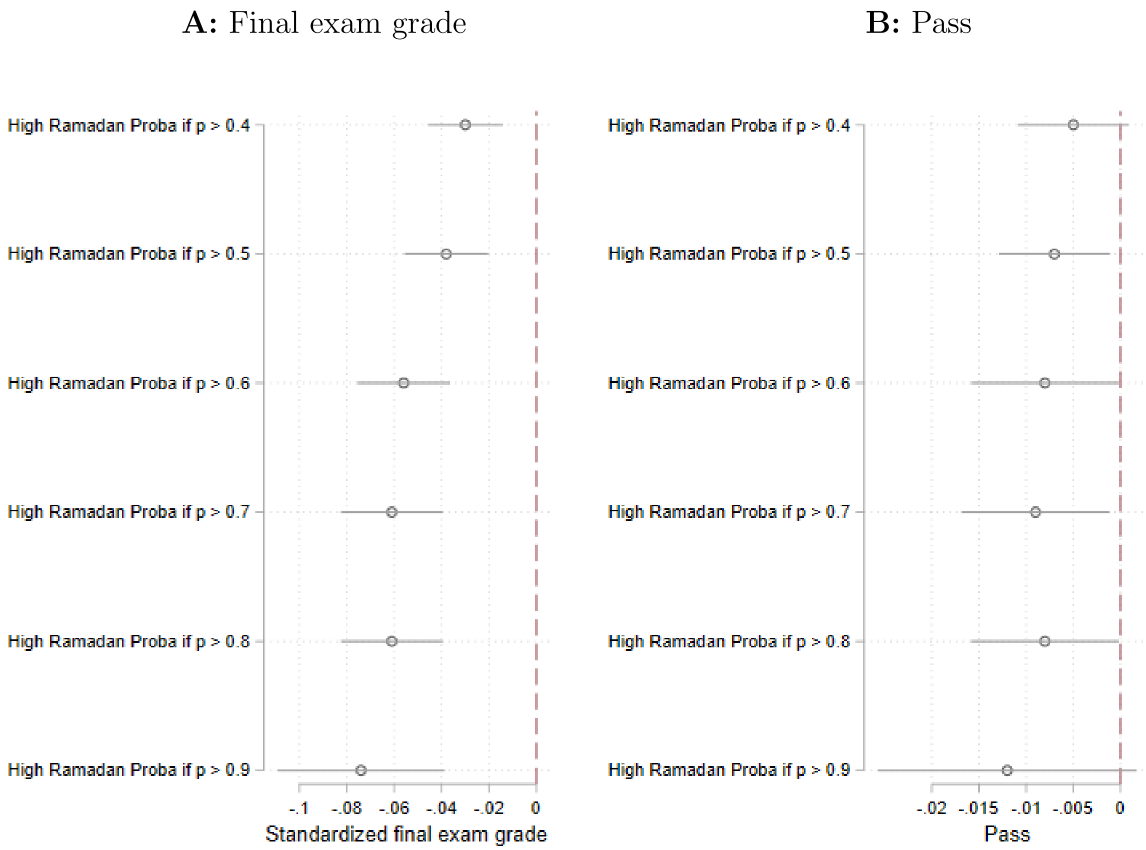
Notes: The figures plot the point estimates and 99% confidence intervals for the coefficients (β_t) that are estimated using our main specification (equation (1)). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. Continuous exam grade is measured as is the average continuous exam grade standardized with mean zero and standard deviation one by track and exam year. The continuous exams took place before the central final exams and hence did not overlap with Ramadan, this test serves as a robustness check and we should not observe any significant impact of Ramadan on this outcome. All controls are included, as in Table A5. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

Table A6: Impact of Ramadan on secondary school outcomes controlling for sibling and family fixed effects

	Secondary School Outcomes			
	Incl. Sibling f.e.		**Incl. Family f.e.**	
	Final	Pass	Final	Pass
	(1)	(2)	(3)	(4)
Ramadan Exam Years	- 0.059***	- 0.014**	- 0.068***	- 0.018***
*** High Ramadan Proba**	(.014)	(.007)	(.013)	(.005)
School-Subtrack f.e.	Yes	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes	Yes
Sibling f.e.	Yes	Yes	No	No
Family f.e.	No	No	Yes	Yes
All controls	Yes	Yes	Yes	Yes
Mean outcome before	6.03	81.8%	6.07	84.9%
Share treated	4.6%	4.6%	5.4%	5.4%
Observations	408,017	408,060	493,477	493,528

Notes: The table shows point estimates for the coefficient β estimated using equation (2). Treated students in this table are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. In Column 1 and 2 we include sibling fixed effects, in Column 3 and 4 family fixed effects (comparing students not only to their siblings but also to their cousins). Final is the average of the central exam grades standardized with mean zero and standard deviation one by track and final exam year. Pass is a dummy indicating that a student passed the final exams and graduated from secondary education. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table A5. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A9: Impact of Ramadan on secondary education outcomes of students with a high Ramadan probability at different thresholds



Notes: The figures plot the point estimates and 99% confidence intervals for the coefficients (β_t) that are estimated using our main specification (equation (1)). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. We define 'high' Ramadan probability based on multiple thresholds, ranging from a probability of 0.4 or higher to a probability of 0.9 or higher. The exams in years 2014 to 2017 did not overlap with Ramadan (and hence we would not expect an effect here), while exams in 2018 and 2019 did overlap with Ramadan. Final exam grade is measured as the average of the central exam grades standardized with mean zero and standard deviation one by track and exam year. Probability of overall pass is measured by a dummy that indicates whether the student passed the final exams (and graduated from secondary education). All controls are included, as in Table A5. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

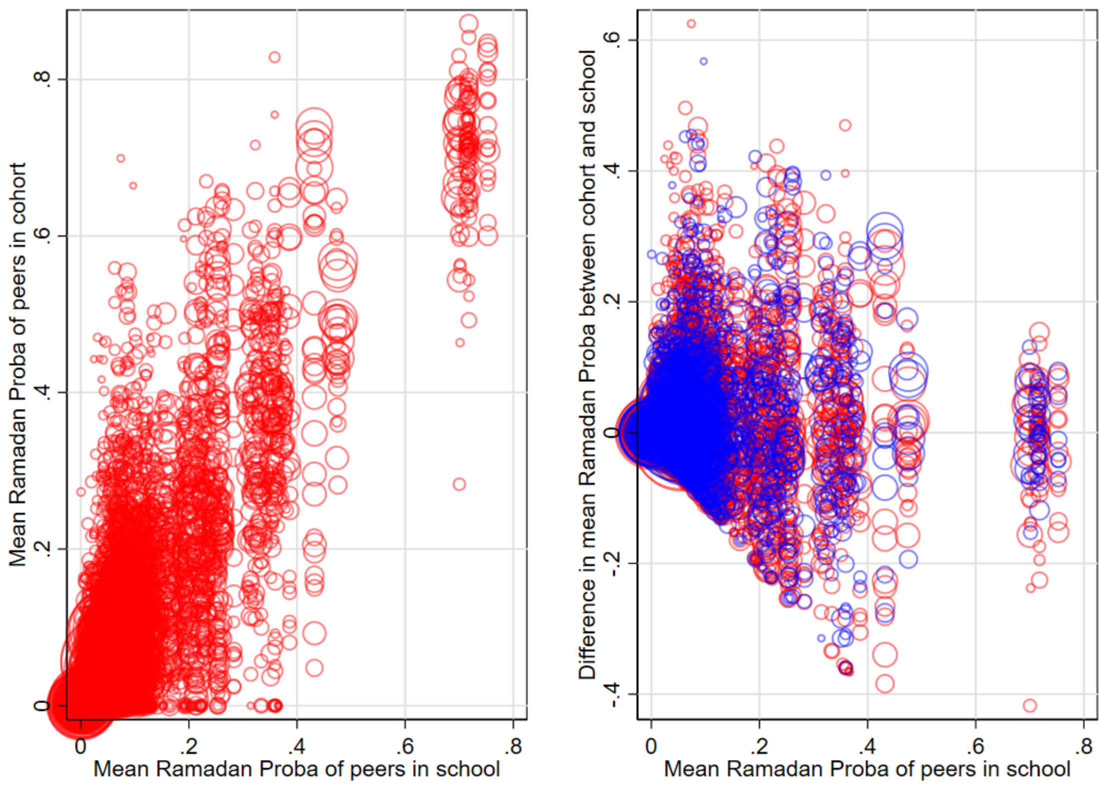
Table A7: Impact of Ramadan at subject level

	Secondary School Outcomes		
	High Ramadan Proba	**Moroc BkG**	**Cont. Ramadan Proba**
	Final	Final	Final
	(1)	(2)	(3)
Ramadan Exam Years	- 0.032***	- 0.037***	- 0.028***
***Treated**	(.006)	(.006)	(.007)
School-Track-Subject f.e.	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes
Treated-Subject-Afternoon-Two day f.e.	Yes	Yes	Yes
All controls	Yes	Yes	Yes
Mean outcome before	6.13	6.17	6.11
Share treated	4.13%	3.40%	4.83%
Observations	4,731,843	4,573,531	4,731,846

Notes: The table shows point estimates for the coefficient β estimated using equation (3) for the 3 different treatment group specifications, based on the subject level data. The dependent variable is the grade obtained in the exam (after retakes) standardized at subject-track-year level. All controls are included, as in Table A5. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. Robust standard errors are clustered at the school-track-subject-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A10: Variation in Ramadan compliance probability across and within schools

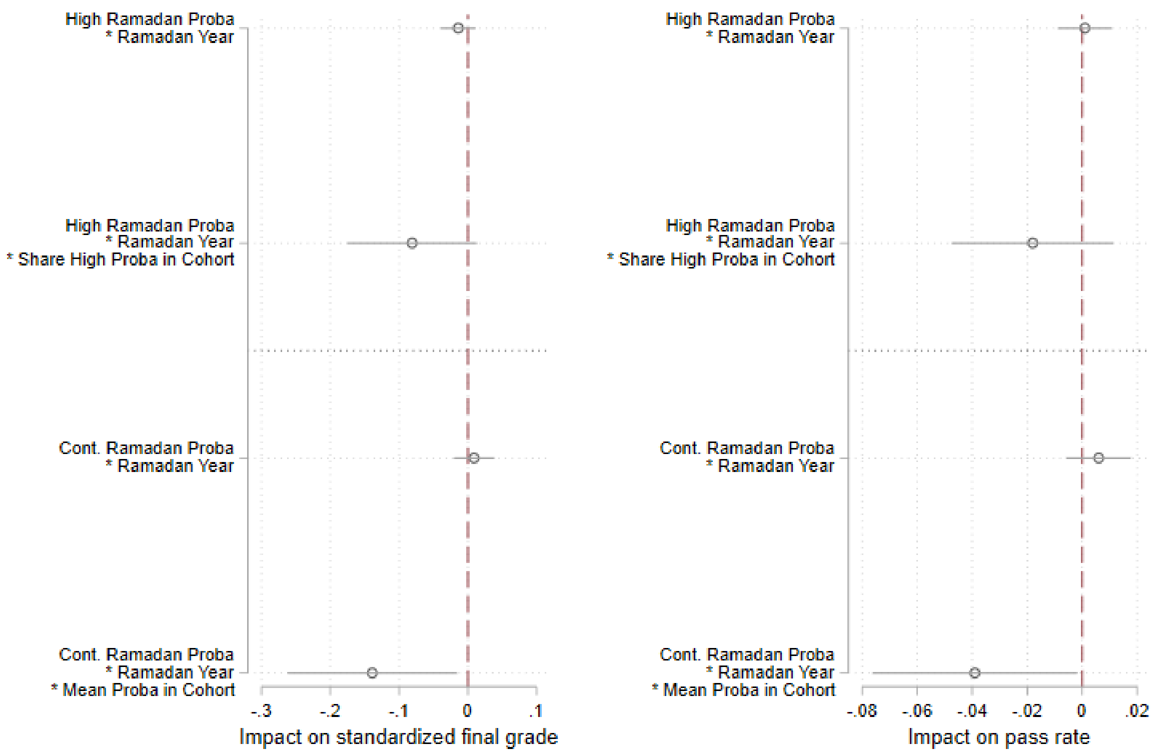
A: Probability in schools and cohorts **B:** Difference in probability within schools



Notes: The left figure plots the mean Ramadan compliance probability in a student's school on the x-axis (leave-one-out-mean in same school) and the mean proability in a student's cohort on the y-axis (leave-one-out-mean in same subtrack-track-school-exam year) for the years 2014 to 2019. The size of the circles is determined by size of cohort. The right figure plots the (leave-one-out) mean continuous Ramadan probability of peers in school (leave-one-out-mean in same school) on the x-axis and the difference between the (leave-one-out-mean) continuous Ramadan probability in the cohort and in the school on the y-axis. The size of the circles is determined by size of cohort. The red circles define the cohorts that took their final exams in 2014-2017, while the blue circles define the cohorts that took their final exams in 2018-2019 (when Ramadan overlapped with the final exams).

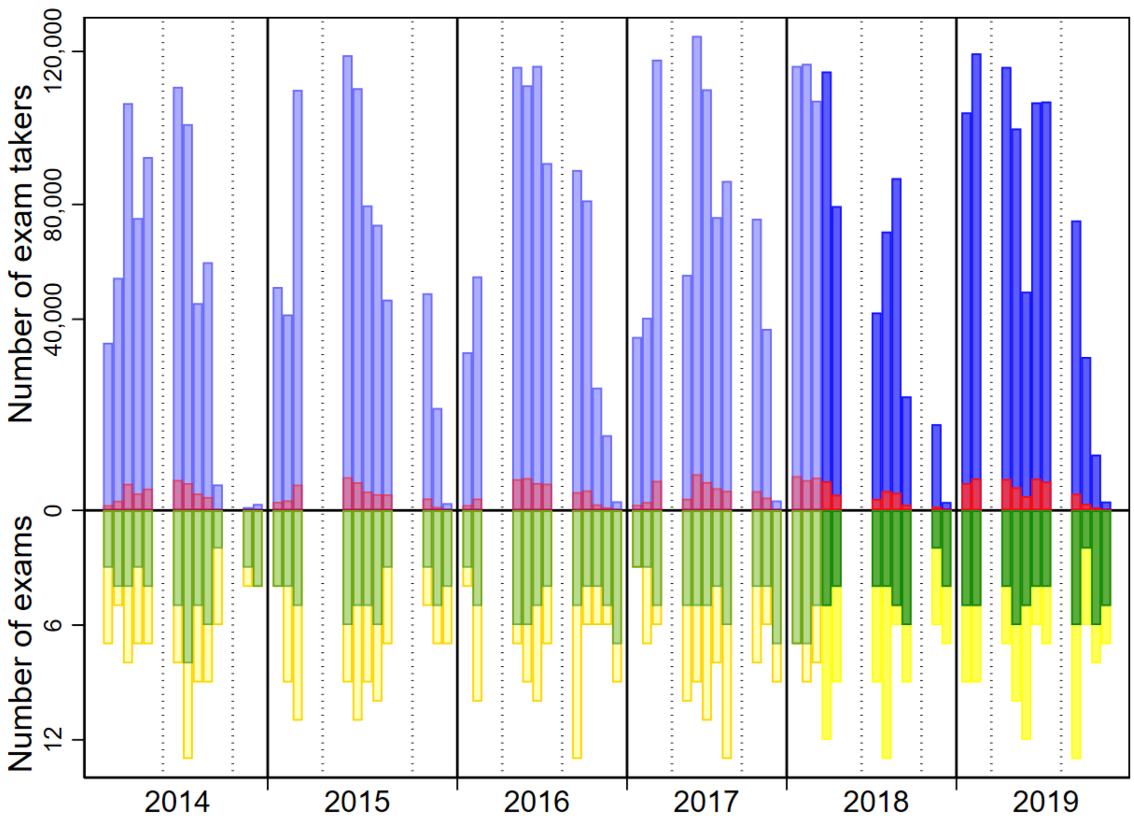
Figure A11: Heterogeneity in impact of Ramadan on secondary education outcomes by Ramadan probability of peers

A: Standardized final grade **B:** Pass rate



Notes: The figures plot the point estimates and 95% confidence intervals for the coefficient β that is estimated using equation (2). In the upper panels, treated students in are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. In the bottom panels, treated students are defined continuously by the Ramadan probability. We interact β with the average share of a student's peers that also have a high Ramadan probability (in the upper panel) or the average Ramadan probability of a student's peers (in the bottom panel). Final exam grade is measured as the average of the central exam grades standardized with mean zero and standard deviation one by track and exam year. Probability of overall pass is measured by a dummy that indicates whether the student passed the final exams (and graduated from secondary education). All controls are included, as in Table A5. School-subtrack and exam year fixed effects are included. Robust standard errors are clustered at the school-track-subtrack-year level. The horizontal dashed line marks the zero or no effect.

Figure A12: Number of exam takers and exams by timing and migration background



Notes: The figure plots on the upper panel the number of exam takers, with in red the number of students with a Moroccan or Turkish migration background and in blue the total number of exam takers on that day. In the bottom panel, it displays the number of exams per day, with in green the number of exams that take place in the afternoon and in yellow the number of exams that take place in the morning. The darker colors indicate exam days that took place during Ramadan.

Table A8: Impact of Ramadan on education outcomes of students with a high Ramadan probability

	Secondary School Outcomes
	Final grade
	(1)
DiD	- 0.040***
	(.014)
DiD *Afternoon	0.010
	(.015)
School-Track-Subject f.e.	Yes
Treated-Subject-Afternoon-Two day f.e.	Yes
Exam year f.e.	Yes
All controls	Yes
Mean outcome before	6.13
Share treated	4.6%
Observations	4,731,843

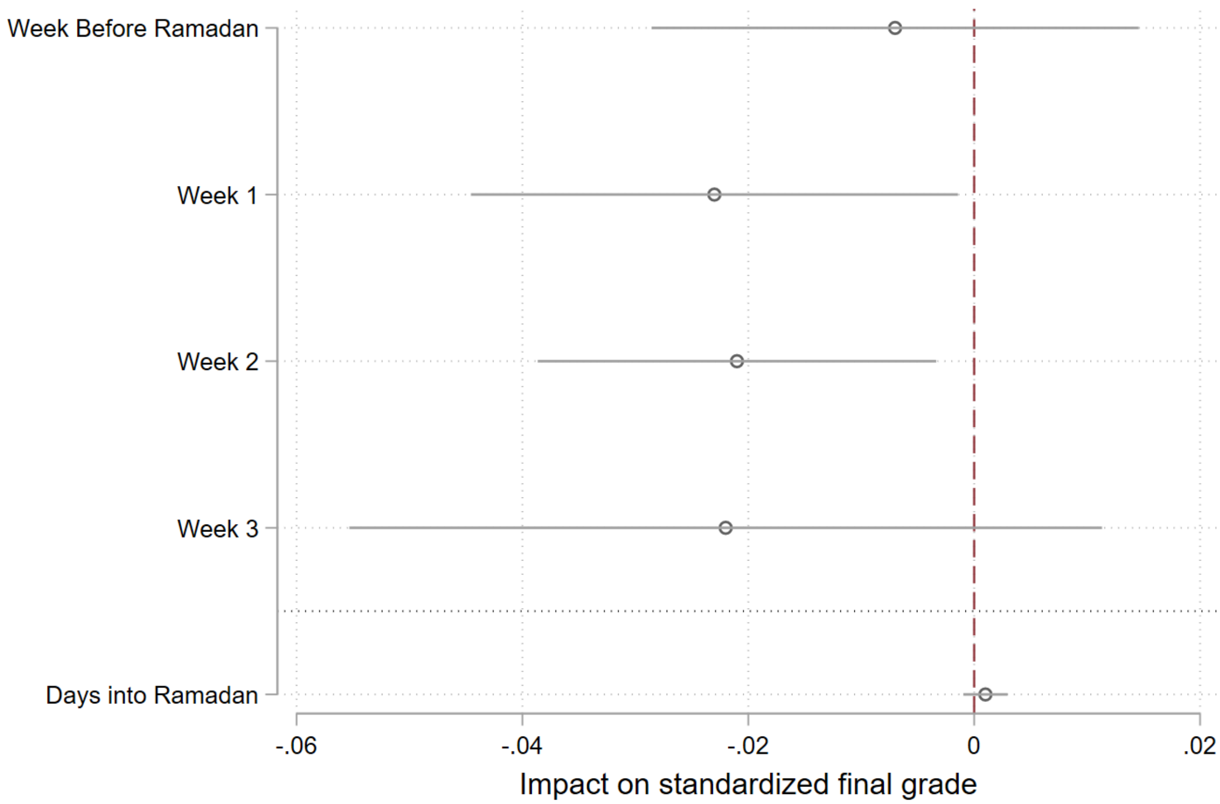
Notes: The table shows point estimates for the coefficient β and for the interaction term between β and a dummy that takes value 1 if the exam took place in the afternoon estimated using equation (3), based on the subject level data. Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. The dependent variable is the grade obtained in an exam (after retakes) standardized at subject-track-year level. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table A5. Robust standard errors are clustered at the school-track-subject-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: DiD estimates by type of subject using high Ramadan probability as treated

	Secondary School Outcomes		
	Memory-based Final (1)	Languages Final (2)	Math Final (3)
Ramadan Exam Years * High Ramadan Proba	- 0.042*** (.014)	- 0.024*** (.010)	- 0.089*** (.025)
School-Subtrack f.e.	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes
All controls	Yes	Yes	Yes
Mean exam grade	6.04	6.19	6.21
Share treated	4.1%	4.5%	4.2%
Observations	867,603	1,930,649	702,153
No. students	583,126	729,274	659,989

Notes: The table shows point estimates for the coefficient β estimated using equation (3), for three groups of exams: languages (reading comprehension tests in Dutch, English, German and French), more memory-based exams (history, geography and biology), and math. The dependent variable is the grade obtained in an exam (after retakes) standardized at subject-track-year level. We include school-track-exam subject and treatment type-exam subject-afternoon-two day exam fixed effects. Robust standard errors are clustered at the school-subtrack-year level in parentheses. All controls are included, as in Table A5. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A13: Effect on final grade by Ramadan week



Notes: This figure plots the point estimates and 99% confidence intervals for the coefficients estimated using equation (5) and equation (6). The coefficients $Week$ 1, $Week$ 2 and $Week$ 3 measure the impact of Ramadan exposure during the first, second or third week of Ramadan on the standardized result of the central exam, using data at subject level. $Week\ Before\ Ramdan$ refers to the week before the beginning of Ramadan (which is observed only in 2018). $Days\ into\ Ramadan$ captures the marginal effect of taking exams after an additional day of Ramadan exposure, as measured by the interaction term β_2 in equation (6). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. All controls are included, as in Table A5. We include school-track-exam subject and treatment type-exam subject-afternoon-two day exam fixed effects. Robust standard errors are clustered at the school-track-subject-year level in parentheses. The horizontal dashed line marks the zero or no effect.

Table A10: DiD estimates by family size using high Ramadan probability as treated

	Secondary School Outcomes			
	Small family Final	Big family Final	Small family Pass	Big family Pass
	(1)	(2)	(3)	(4)
Ramadan Exam Years	-0.054***	- 0.065***	- 0.010**	- 0.012**
*** High Ramadan Proba**	(.013)	(.013)	(.005)	(.005)
School-Subtrack f.e.	Yes	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes	Yes
Mean exam grade	6.12	6.07	87.2%	86.3%
Share treated	4.5%	5.0%	4.5%	5.0%
All controls	Yes	Yes	Yes	Yes
Observations	423,221	352,488	423,261	352,513

Notes: The table shows point estimates for the coefficient β estimated using equation (2). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. We split our sample between students with a small (below-median) family and students with a big (above-median) family. Family size is measured as extended family size, including siblings, parents, grandparents, aunts, uncles and cousins. Final is the average of the central exam grades standardized with mean zero and standard deviation one by track and final exam year. Pass is a dummy indicating that a student passed the final exams and graduated from secondary education. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table A5. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Impact of Ramadan on Longer-Term Outcomes

	Longer Term Outcomes		
	Grade retention	Adult education	Drop out
	(1)	(2)	(3)
Ramadan Exam Years	0.007***	0.004	0.004
***High Ramadan Proba**	(.003)	(.002)	(.004)
School-Subtrack f.e.	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes
All controls	Yes	Yes	Yes
Share treated	4.70%	4.70%	4.70%
Mean outcome	5.60%	4.94%	5.01%
Observations	776,284	776,284	776,284

Notes: The table shows point estimates for the coefficient β estimated using equation (2). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. Grade retention is a dummy the takes value one if the student repeats the final year in the year after the final exam. Adult education is a dummy that takes value one if the student goes into adult education in the year after the final exam. Drop out is a dummy that takes value one if the student is not registered for any education in the year after the final exam. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table A5. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Impact of Ramadan on Longer-Term Outcomes by Age

	Longer Term Outcomes					
	Younger than 18			**Older than 18**		
	Grade retention	Adult education	Drop out	Grade retention	Adult education	Drop out
	(1)	(2)	(3)	(4)	(5)	(6)
Ramadan Exam Years	0.006**	0.004*	0.001	0.008	0.000	0.017**
***High Ramadan Proba**	(.003)	(.002)	(.004)	(.005)	(.008)	(.008)
School-Subtrack f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Exam year f.e.	Yes	Yes	Yes	Yes	Yes	Yes
All controls	Yes	Yes	Yes	Yes	Yes	Yes
Share treated	5.32%	5.32%	5.32%	3.34%	3.34%	3.34%
Mean outcome	5.86%	2.18%	4.26%	4.68%	14.5%	7.59%
Observations	536,927	536,927	536,927	238,111	238,111	238,111

Notes: The table shows point estimates for the coefficient β estimated using equation (2). Treated students in this figure are students with a Moroccan or Turkish migration background that have a high Ramadan probability, as predicted by our machine learning model. We split our sample into students that are younger or older than 18 years old at the 1st of September after the final exam. Grade retention is a dummy the takes value one if the student repeats the final year in the year after the final exam. Adult education is a dummy that takes value one if the student goes into adult education in the year after the final exam. Drop out is a dummy that takes value one if the student does not register for any education in the year after the final exam. Mean outcome before is the average outcome for treated students in the pre-treatment years 2014 to 2017. All controls are included, as in Table A5. Robust standard errors are clustered at the school-track-subtrack-year level in parentheses. $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Data Appendix

B.1 Construction of sample

We use administrative data from Statistics Netherlands which contains information on all individuals who are registered in a municipality by 1995.[35] We start with the registry of exam takers (*EXAMVOTAB*), select all exam takers between 2014 and 2019 (where exam year is the school year plus one, because students start the final school year in September of i.e. 2013 and take the exam in May 2014) and drop entries for which the final grade is missing,[36] which gives us a sample of 983,837 individuals (of whom some take more than one exam, so we have a total of 1,048,400 observations). We match them to the track and sub-track they are in using the education reference book (*OPLEIDINGSNRREFV26*).

B.2 Construction of variables

We match the exam takers in our sample to their demographic characteristics (country of birth, country of origin, country of birth of parents, gender, birth date and migration generation) in the registry of persons (*GBAPERSOONTAB*). The parent-child registry (*KINDOUDERTAB*) is used to link the exam takers in our sample to their parents such that we can determine outcomes for the parents. We match parents to their demographic characteristics (country of birth, country of origin, birth date and country of birth of parents) in the registry of persons. We also use parents to identify the rest of the (extended) family to use in a family fixed effects analysis. We first use the identifying number of the mother to identify an exam taker's siblings, then we use it to identify the grandparents. We use the grandparents to identify aunts and uncles and then link those aunts and uncles to their children to identify cousins. We also do this for the father to identify the father's side of the

[35] The administrative data from Statistics Netherlands is available at a remote-access facility after signing a confidentiality agreement.

[36] The number of missing observations in the exam years 2016 and 2019 is unusually high, but these entries refer to students that were already pre-registered in the exam grade dataset in the year before the final year for policy reasons, and not to students in the final year taking the exams.

family. Due to the problem of overlapping families, the family fixed effects are based on the mother of the father or, if this is not available, the mother of the mother, or, if this is not available the mother (only siblings) or father (only siblings). Treated students suffer from more limitations than control students, because in order to identify the extended family we need to identify the grandparents, which are often not available in the administrative data for students with a migraton background. Therefore, while for students in our control group we can identify for 95% the extended family (and compare the exam taker not only with siblings but also with cousins), we can only do this for 77% of the students in our treatment group.

Country of origin is defined as the country of birth if a person is born abroad or, if born in the Netherlands, the country of birth of the mother, or, if the mother is also born in the Netherlands, the country of birth of the father. We construct a new variable that also considers third generation migrants, by replacing the country of origin with the country of origin of the mother if the country of origin of the exam taker is the Netherlands, or the country of origin of the father if the country of origin of the mother is the Netherlands. We classify exam takers as Muslim if this new country of origin is a Muslim majority country (see Table A13).[37]

We use birth date from the registry of persons to define the age at exam (using the 1st of May of the exam year). We add information from the registry with the information on the individual's obtained test score at the standardized test at the end of primary education (*CITOTAB*). This registry is only available from 2006 to 2019, so given that exam takers took the CITO test 4 to 6 years before their final exam the exam takers to the CITOTAB 2006-2019. We can match 75.4% of the exam takers to their obtained CITO test score. This leaves us with a sample of 841,347 observations (789,259 unique students).

We aggregate the new country of origin variable at the school level to create a variable with the proportion of Moroccans in each school and track. We then divide this variable in

[37](We are excluding Indonesia, even though this is a Muslim majority country, as a very selected group of (mostly non$_{M}uslim)migrantscamefromIndonesiatotheNetherlandsarenotMuslim$.

Table A13: Muslin majority countries (source: worldpopulationreview.com).

Country	Percentage of Muslims
Morocco	99
Somalia	100
Afghanistan	100
Tunisia	99
Iran	99
Mauritania	100
Yemen	99
Iraq	97
Maldives	100
Comoros	98
Niger	99
Turkey	100
Algeria	99
Azerbaijan	97
Jordan	97
Uzbekistan	88
Djibouti	94
Libya	97
Senegal	96
Pakistan	97
Tajikistan	98
Gambia	96
Egypt	90
Mali	94
Turkmenistan	93
Syria	87
Saudi Arabia	99
Bangladesh	89
Sudan	100
Kyrgyzstan	90
Indonesia	87
Oman	86
Guinea	89
Albania	57
Sierra Leone	77
Brunei	79
United Arab Emirates	76
Kazakhstan	70
Kuwait	75
Bahrain	74
Malaysia	61
Qatar	65
Burkina Faso	63
Lebanon	61
Chad	52
Bosnia and Herzegovina	51
Nigeria	50

four quantiles.

We then match the parents to data on labour income (*SECMWERKNDGAMNDBEDRAG-BUS*), unemployment benefits (*SECMWERKLMNDBEDRAGBUS*), income from profits

(*SECMZLFMNDBEDRAGBUS*), and illness benefits (*SECMZIEKTEAOMNDBEDRAG-BUS*). We aggregate the income data to yearly income data and only keep data on income in the year before the exam was taken. We generate a dummy variable that equals one if one of the parents receives unemployment benefits and do the same for illness benefits. We sum all sources of income to create a variable on total parental income and divide this in year-specific deciles.

In our analysis we make use of different spatial data: we exploit several characteristics of the students' neighbourhood of residence and we compute (approximate) distances between the students' residences and the closest mosque and primary school. In order to link students with their neighbourhood of residence[38], in the year in which the final exam is taken, we use the dataset called *GBAADRESOBJECTBUS*. This dataset provides two codes (*soortobjectnummer* and *rinobjectnummer*) that uniquely identify the neighbourhood of residence of each individuals that is registered in the Dutch population registration system. More specifically, we need to match these two codes with an additional code that identifies a neighbourhood (the variable *BU2021*, available in the dataset called *VSLGWTAB*). This neighbourhood code can be then used to obtain the geospatial data of the neighbourhood (from the *gemeentewijkbuurt* dataset) or its socio-demograpic characteristics (from the *wijk_en_buurtstatistieken* dataset).

We obtained the mosques' geolocation and characteristics from the website *moskeewijzer.nl* in November 2021[39]. Importantly, for each mosque, we know whether it is a Moroccan or Turkish mosque as reported by the website. In order to use these data in our analysis, we asked CBS to encrypt the address of the mosques so that each could get a *soortobjectnummer* and *rinobjectnummer*. We then matched again these codes with the corresponding neighbourhood identifier (*BU2021*). Finally, we computed the distance between the centroids

[38] Spacial data at neighbourhood level (*buurt* in Dutch) represent the smallest geographical unit we could obtain from CBS.

[39] The dataset used in our analysis contains the existing mosques in the Netherlands in the year 2021. As we do not know when a mosque was built, we use all the existing mosques in 2021 to compute the students' distance to the closest mosque. This introduces some measurement error as the map of mosques location has most likely changed between 2021 and 2014 (the first year considered in our analysis).

of the students neighbourhood of residence and the closest neighbourhood with a mosque. In doing so, we considered only Turkish mosques for students with a Turkish migration background and only Moroccans mosque for students with a Moroccan backgound.

B.3 Construction of outcome variables

We use two different datasets on outcomes in secondary education: a registry of the exam takers and exam outcomes (*EXAMVOTAB*) and a registry of the exam outcomes at course level (*EXAMVOVAKTAB*).

We start with the exam outcomes at the general level. First, we drop exam takers for which the exam outcomes are missing, while the exam taker did not withdrew from the exam. We only keep the students that passed or failed the final exams or that withdrew from the final exam. We standardize the average grade from the first central exams, the average grade from the central exams including the first and second retake (if applicable and grade improved) and the average grade from the continuous school exam by track with mean zero and standard deviation one. We generate a dummy that takes value one if the student passed the exams (and graduated).

For the data at course level (*EXAMVOVAKTAB*) we consider only the exam takers that are included in our analysis of *EXAMVOTAB*. The dataset contains 39 different exam subjects. We manually classified them into 29 subjects. Those exams that we could not classify into one of these subjects are dropped. The final list of exam subjects included in our analysis is the following: Arabic, Arts, Biology, Chemistry, Dutch, Economics, English, French, Friesian, Geography, German, Greek, History, Latin, Management, Math (VMBO level, A, B, C), Music, Philosophy, Physics, Physics and Chemistry (I and II), Russian, Social Studies, Spanish, Technical Design, Turkish. This means that some practical exams and uncommon exams (both mainly in the pre-vocational track) are dropped from the sample. We merge these data with the date and time the exams are held, as reported by the website *www.examenblad.nl*.

B.4 SIM Survey

The *Survey integratie minderheden* (SIM) is a survey conducted by the *Sociaal en Cultureel Planbureau* (SCP) in collaboration with Statistics Netherlands. The survey includes questions on education, labour, social life, cultural integration, religion, health and sports. The interviews were conducted face-to-face, by telephone and online in 2006, 2011, 2015 and 2020.[40] The sample of the survey consists of respondents of 15 years or older of the four important (first and second generation) immigration groups in the Netherlands: Turkey, Morocco, the (former) Dutch Antilles, and Suriname (and in 2015 also Poland and Somalia), and of a control group of people without a migration background. Sample selection consisted of two steps: municipalities were randomly selected, within these municipalities the sample size was determined by municipality size, and people were selected randomly from the different immigration groups in the municipality.

In 2006, the data collection took place from March 2006 to December 2006. The sample consisted of 11,520 people, of whom 2,162 had a Turkish migration background and 2,390 a Moroccan migration background. The response rate was 53% (1,032 Moroccans and 1,132 Turks). In 2011, the data collection took place from November 2010 to June 2011. The sample consisted of 14,178 people, of whom 2,544 had a Turkish migration background and 2,826 a Moroccan migration background. The response rate was 48% (1,395 Moroccans and 1,347 Turks). In 2015, the data collection took place from January 2015 to July 2015. The sample consisted of 15,028 people, of whom 1,858 had a Turkish migration background and 2,082 a Moroccan migration background. The response rate was 45% (920 Moroccans and 951 Turks). In 2020, the data collection took place from September 2020 to Januari 2021.[41] The sample consisted of 15,498, of whom 2020 had a Turkish migration background

[40] The interviews were conducted between March and December in 2006, between January to April and November to December in 2011, January to June in 2015 and March 2020 to January 2021. Due to measures against COVID-19 during Ramadan in 2020, mosques were closed and Muslims did not share the iftar meal with friends and neighbours.

[41] The data collection was delayed by the Covid-19 pandemic and took longer than planned because of lock-downs.

and 2,114 a Moroccan migration background. The response rate was 35.4% for individuals with a Turkish migration background (696 respondents) and 27.5% for individuals with a Moroccan migration background (570 respondents).

We linked the SIM data to the administrative data from Statistics Netherlands. Some respondents are included in more than one survey wave and for those respondents we only keep the first observation. We are left with 5,400 respondents in total of a (first or second) generation Moroccan or Turkish migration background. In total, we have 2,670 respondents with a Moroccan migration background and 2,730 respondents with a Turkish migration background.

B.5 Additional Regressions

In this section we provide details for the regressions that we estimate in our heterogeneity analyses.

In section 6.2, we replicate our main results, using data at the subject level. To do so, we estimate the following regression:

$$
\begin{aligned}
y_{i,t,z} = {} & \alpha_{p,z} + \theta_{i,z} + \eta_{i,z} + \beta \, Treated_i * Ramadan\ Year_t + \zeta \, Treated_i \\
& + Exam\ Num_{i,t,z} + \delta X_{i,t} + \epsilon_{i,t,z}
\end{aligned}
\tag{3}
$$

$y_{i,t,z}$ is the grade obtained by student i, in year t, in subject z, standardized at the subject-track-year level. $Exam\ Num_{i,t,z}$ represents fixed effects for the number of exams that student i in year t has already completed while taking exam z. $\alpha_{p,z}$ are school-track-exam subject fixed effects. $\theta_{i,z}$ are treatment type-exam subject-afternoon- two day exam fixed effects. The controls in the equation are the same as in the previous specifications. Standard errors are clustered on school-subtrack-exam year-exam subject level.

In section 7.1 we estimate the impact of Ramadan on exam performance, based on the timing of the exams and how many exams a student takes in a day. To do so, we employ the following regression:

$$
\begin{aligned}
y_{i,t,z} = {} & \alpha_{p,z} + \theta_{i,z} + \delta X_{i,t} + \gamma_t + \mathit{Exam\ Num}_{i,t,z} + \beta_1 DiD_{i,t} + \beta_2 Afternoon_z + \beta_3 \mathit{Two\ Exam}_z \\
& + \beta_4 DiD_{i,t} * Afternoon_z + \beta_5 DiD_{i,t} * \mathit{Two\ Exam}_{i,z} + \beta_6 Afternoon_z * \mathit{Two\ Exam}_{i,z} \\
& + \beta_7 DiD_{i,t} * \mathit{Two\ Exam}_z * Afternoon_z + \zeta\, \mathit{Treated}_i + \epsilon_{i,t,z}
\end{aligned} \tag{4}
$$

$y_{i,t,z}$ is the grade obtained by student i, in year t, in subject z, standardized at the subject-track-year level. DiD_{it} is equal to the interaction between $Treated_i$ and $Ramadan\ Year_t$. $Afternoon_z$ is a dummy equal one if exam z is in the afternoon. $Two\ Exam_{i,z}$ is a dummy equal one if individual i is taking exam z in the day same day of another exam. $Treated_i$ is replaced by individual fixed effects in some specifications. All the other variables are the same as in equation(1).

In section 7.1 we estimate the impact of Ramadan on exam performance, depending on the week of Ramadan in which the exam is held. We estimate the following regression:

$$
\begin{aligned}
y_{i,t,z} = {} & \alpha_{p,s,z} + \theta_{i,z} + \omega_{i,z} + \delta X_{i,t} + \gamma_t + \mathit{Exam\ Num}_{i,t,z} + \sum_{t=2014}^{2017} \beta_t\, \mathit{Treated}_i * \mathit{Year}_t + \zeta\, \mathit{Treated}_i \\
& + \sum_{n=0}^{3} Week_{z,n} + \phi_n Week_{z,n} * \mathit{Treated}_i + \epsilon_{i,t,z}
\end{aligned} \tag{5}
$$

$y_{i,t,z}$ is the grade obtained by student i, in year t, in subject z, standardized at the subject-track-year level. $Week_{z,n}$ is a dummy taking value on if exam z is held in week n of the Ramadan month. $Week_{z,0}$ is equal to one if exam z is held in 2018 in the week before the beginning of the Ramadan month. The coefficients of interests are ϕ_1, ϕ_2, ϕ_3, which captures the impact of Ramadan exposure during the first, second or third week of the Ramadan month.

To test the possibility that Ramadan exposure affects students differently, depending on whether the exam is taken at the beginning or at the end of the Ramadan month, we also implement the following regression:

$$y_{i,p,s,t} = \alpha_{p,s} + \beta_1 \, Treated_i * Ramadan\ \ Year_t + \beta_2 Ramadan\ \ Days * Treated_i * Ramadan\ \ Year_t$$
$$+ \delta X_{i,t} + \gamma_t + \epsilon_{p,s,t} \tag{6}$$

In the equation above *Ramadan Days* is a variable measuring how many days after the beginning of Ramadan an exam is taken. All other variables are the same as in equation (1). The coefficient β_2 measures the marginal effect of taking exams after an additional day of Ramadan exposure.