Rearranging the Desk Chairs: A Large Randomized Field Experiment on the Effects of Close Contact on Interethnic Relations*

Felix Elwert, *University of Wisconsin-Madison*

Tamás Keller, *CSS-RECENS Centre for Social Sciences, Budapest, Hungary*

Andreas Kotsadam, *Ragnar Frisch Centre for Economic Research, Oslo, Norway*

Abstract. Contact theory predicts that interethnic exposure reduces anti-minority discrimination. By contrast, conflict theory predicts that interethnic exposure worsens discrimination. The scope conditions for both theories are vague; prior evidence is mostly correlational; and supportive field experiments for contact theory have largely accrued in rarified settings. This begs the question how interethnic contact affects interethnic relations in everyday situations. We test the causal effect of interethnic exposure on discrimination under quotidian conditions in a pre-registered randomized field experiment involving N=2,395 students in 39 Hungarian schools. We find that neither manipulating the closeness of interethnic exposure between students within classrooms, nor variation in ethnic composition across grade levels, affects anti-minority discrimination. This shows that the domains of contact and conflict theory are much less expansive than previously thought. Interethnic contact may not affect discrimination either way in many everyday settings.

*** Acknowledgements.** We thank Markus Brauer, Steven Durlauf, Mustafa Emirbayer, Henning Finseraas, Mosi Ifatunji, Åshild Johnsen, Julia Rohrer, audiences at Chicago, Wisconsin, Copenhagen, Cornell, and Humboldt University, and at the Symposium Social Cohesion in Ethnically Diverse Societies, Tilburg, for comments. The research was funded by grants from the Hungarian National Research, Development, and Innovation Office (NKFIH), FK-125358; a János Bolyai Research Scholarship, Hungarian Academy of Sciences, BO/00569/21/9; the ÚNKP-22-5-CORVINUS-134 New National Excellence Program, Ministry for Innovation and Technology, National Research, Development and Innovation Fund, Hungary; the Research Council of Norway, #287766; NICHD, #P2C HD047873; and a Romnes Fellowship, University of Wisconsin-Madison. Tamas Keller is also affiliated with the Centre for Economic and Regional Studies and the TÁRKI Social Research Institute, Budapest, Hungary. Direct correspondence to Felix Elwert, Department of Sociology, University of Wisconsin-Madison, 1180 Observatory Dr., Madison, WI 53706, elwert@wisc.edu.

Introduction

Whether ethnic intermingling exacerbates or diminishes ethnic tensions is a pressing question of our time (Abascal and Baldassari 2015). Social science presents sharply different predictions for the effects of interethnic exposure. On one hand, contact theory (Allport 1954) predicts that deep and cooperative interethnic exposure promotes understanding, reduces prejudice, and increases interethnic trust. On the other hand, conflict and group threat theories (Blalock 1967; Bobo 1999; Williams 1964) predict that shallow or competitive exposure spurs exclusionary attitudes on the part of the ethnic majority toward their ethnic others. Constrict theory (Putnam 2007), a variant of conflict theory, additionally posits that interethnic exposure weakens in-group solidarity among the majority group.

These contrasting theoretical predictions beg the question of scope conditions (Paluck et al. 2019). When does interethnic contact promote trust (contact theory), and when does it promote prejudice (conflict theory)? Building on Allport's (1954) classical formulation, over time, social scientists have endorsed increasingly expansive scope conditions for contact theory, reporting positive effects not only for prolonged cooperative interethnic exposure but also for indirect exposure via mass media and friends of friends (Pettigrew et al. 2011). Perhaps buoyed by this confidence, "the promotion of intergroup contact has arguably become the foremost strategy for reducing prejudice" (Paluck, Green, and Green 2019:130).

At least three reasons, however, caution against resting sweeping policy hopes on contact theory. The first concern is selection bias. The great majority of research supportive of contact theory is cross-sectional and observational, i.e., it compares individuals who, at least in part, self-select into, or out of, intergroup contact. It is reasonable to expect that individuals who choose to have contact with ethnic others are less prejudiced against them in the first place. Separating selection based on ex-ante predispositions from the causal effects of the resulting interethnic contact on attitudes and behavior is difficult (Morgan and Winship 2015). The methodological

consensus in sociology is that causal claims are most credible when they are backed by field experimental evidence that eliminates selection bias through randomization (Baldassari and Abascal 2017).

The second concern is impact. Compared to the large positive effects reported by observational studies, randomized field experiments have generally found smaller—and often much smaller—positive effects of intergroup contact (see Paluck et al. 2019 for a review). Furthermore, larger experiments report smaller effects, as do experiments that curtail specification searches by following pre-registered analysis plans (Paluck et al. 2019). These regularities suggest publication bias (Christensen, Freese, and Miguel 2019) in favor of studies that support contact theory, even among randomized experiments.[1]

The third concern is generalizability. The randomized field experiments that most strongly support contact theory were conducted not only under favorable, but arguably under rarified conditions (Dixon, Durrheim, and Tredoux 2005). Ten of the 24 randomized field experiments supporting contact theory in Paluck et al.'s (2019) comprehensive review of the field-experimental literature investigate the effects of enforced coresidence among highly selected populations of relative strangers. To wit, research demonstrates conclusively that interethnic exposure promotes inclusionary attitudes and behaviors when members of different ethnic groups are forced to live together in elite college dorms (Boisjoly et al. 2006), the military (Finseraas and Kotsadam 2017), or elite military college dorms (Carell, Hoekstra, and West 2015). Clearly, if the positive effects of interethnic contact on interethnic relations hinged on coresidence, the policy potential of contact theory would be limited (Paluck and Green 2009).

[1] Empirical evidence indicates a combination of researcher malfeasance and publication bias in favor of statistically significant findings in leading journals of sociology (Gerber and Malhotra 2008a), political science (Gerber and Malhotra 2008b), and economics (Brodeur et al. 2016).

Together, these concerns raise the question of when, and to what extent, interventions that promote intergroup contact diminish intergroup discrimination, especially when these interventions occur in mundane, and hence scalable, conditions.

In this study, we test the causal effects of interethnic contact under quotidian and scalable conditions. Rather than intervening on coresidence among former strangers in elite colleges or the military, we randomize the seating chart in public-school classrooms in Hungary, a country where ethnic tensions between minority Roma and majority non-Roma Hungarians run strong. We conduct two experiments: A vignette experiment to measure discrimination, and a field experiment that randomizes the seating chart in 186 classrooms of 39 schools for the duration of one semester. This allows us to investigate whether close interethnic contact, defined as sitting next to a deskmate belonging to the Roma ethnic minority, affects outgroup discrimination, outgroup-friendships, and ingroup cooperation among members of the non-Roma Hungarian majority.

Our study differs from previous randomized evaluations of contact theory in several ways. First, our intervention has the potential of universal scalability, since, in contrast to college or the military, almost all members of a birth cohort attend school during childhood and adolescence. Second, our intervention has a light touch. Rather than intervening in individuals' non-family living arrangements, which is typically the domain of intimate personal choice, we intervene in classroom seating charts, which are routinely set by teachers. Third, we study a well-established natural setting. Rather than inducing coresidence between strangers in a new environment (college freshmen or military recruits), we reseat 3rd through 8th grade students who have grown up in the same villages and small towns and have attended school together for at least 2 years prior to the intervention. Fourth, to our knowledge, this study is by far the largest randomized field experiment of interethnic contact, involving N=2,395 students. Fifth, unlike most field experiments on

interethnic contact (Paluck et al. 2019) our study was pre-registered and closely adheres to a pre-analysis plan (Appendix D) to curtail specification searches in pursuit of desired findings.[2]

As we elaborate below, one central problem of the literature on interethnic exposure is that the scope conditions of contact theory are vague, multidimensional, may interact with each other, and are not sharply distinguished from the scope conditions of conflict and constrict theory in many natural settings. We argue that our setting is a good *ex ante* fit for the canonical scope conditions of contact theory. To the extent, however, that our setting can also be argued to fit the scope conditions of conflict and constrict theory, our study also tests conflict and constrict theory.

Our results are unambiguous. Our randomized vignette experiment documents substantial discrimination against Roma students by non-Roma students in our sample. The probability that a non-Roma student would lend money to a classmate is reduced by 27 percent if that classmate is described as Roma. But our field experimental findings fail to lend support to *any* of the main theories of interethnic exposure. First, they disappoint the hopes raised by contact theory, as being randomly assigned to a Roma deskmate for an entire semester does not reduce the ethnic majority's discrimination against Roma students. Indeed, being randomly assigned to a Roma deskmate does not even lead to a higher probability of having a Roma friend inside or outside of the classroom. Second, in contrast to conflict theory, we do not find that exposure to a Roma deskmate leads to more discrimination against Roma children or fewer interethnic friendships. Third, in contrast to constrict theory, we find no evidence that exposure to Roma deskmates reduces intra-ethnic cooperation among non-Roma Hungarians.

[2] Pre-registration does not per se guarantee high methodological standards, but it guards against "p-hacking" (respecifying models until results meet desired levels of statistical significance [Head et al. 2015]) and "HARKing" ("hypothesizing after the results are known"), both of which would invalidate statistical inference (Christensen et al. 2019).

Our Null results are informative, i.e., they are not due to inadvertent averaging across heterogeneous effects. All estimates are centered around the Null hypothesis of no effect, and there is little indication of effect heterogeneity based on students' own characteristics, their deskmates' characteristics, or classroom characteristics. This suggests that our Null results do not present lack of evidence for an effect, but evidence for the lack of an effect.

Supplementary analyses further rule out that grade-level exposures to Roma students affect discrimination. We show this using a quasi-experimental approach that exploits plausibly random variation in the share of Roma students across grades within schools (Hoxby 2000). We find that while the number of interethnic friendships increases with the share of Roma students in the grade, the magnitude of discrimination is unaffected. Hence, these results suggest that neither desk- nor grade-level contact with the ethnic minority ameliorates (or exacerbates) out-group discrimination among the ethnic majority. In other words, we neither find support for contact theory nor for conflict theory in response to changes in interethnic exposures in a school setting.

Previous findings and scope conditions

The interdisciplinary literature on the effects of interethnic exposure in sociology, social psychology, political science, and economics divides into two main perspectives: contact theory and conflict theory. According to contact theory, close and cooperative contact, under scope conditions detailed below, can reduce prejudice and increase trust (Allport, 1954). Such exposures may diminish interethnic animosity by fostering empathy, increasing understanding, normalizing or habituating otherness, and promoting friendship. By contrast, conflict theory posits that animosity across groups is worsened with shallow, fleeting, or competitive exposure (Blalock 1967; Williams 1964). Such exposures may worsen outcomes by lacking the depth to promote understanding and empathy so that the perceived otherness of the interacting parties dominates to activate aversion

and perception of threat. Putnam's (2007) constrict theory is a variant of conflict theory arguing that interethnic exposure may also undermine in-group relations.

The canonic treatment of interest in contact, conflict, and constrict theory is the co-location of members of different ethnic groups in the same physical space. The wider (especially interventionist) literature on inter-group relations sometimes additionally evaluates bundled treatments that combine interventions on co-location with mandatory perspective-taking exercises on race or ethnic relations (e.g., Sorensen 2010; Markowicz 2009). We sidestep consideration of bundled treatments and focus on the effects of co-location.

Contact theory is supported by hundreds of observational, and mostly cross-sectional, studies (see e.g., Pettigrew and Tropp 2006; Pettigrew et al. 2011 for reviews), and also by a small number of randomized field experiments. We found 8 prior randomized field experiments involving interracial or interethnic exposures (Boisjoly et al. 2006; Burns, Corno and La Ferrara 2016; Camargo et al. 2010; Carrell et al. 2019; Green and Wong 2009; Finseraas and Kotsadam 2017; Finseraas et al. 2019; Page-Gould et al. 2008). Most field experiments that report positive effects of interethnic contacts evaluate rarified interventions that enforce prolonged coresidence of young adults in college dorms or military bootcamps (Boisjoly et al., 2006; Burns et al. 2016; Carrell et al. 2019; Finseraas and Kotsadam, 2017; Finseraas et al. 2019; Camargo et al. 2010). For example, Boisjoly et al. (2006) found that random assignment to African American roommates at a selective U.S. university increased white students' support for affirmative action. Carrell et al. (2019) found that assignment to African American roommates at the United States Air Force Academy increased white students' requests for African American roommates in subsequent years. Finseraas and Kotsadam (2017) found that random assignment to ethnic-minority roommates during boot camp improved Norwegian army recruits' opinion of immigrants' work ethic. In a similar spirit, Green and Wong (2009) find that white high-school students developed greater out-group tolerance after random assignment to racial or ethnic others during a three-week long outdoor survival course.

One prior field experiment evaluates the consequences of contact with ethnic others under less rarified conditions. Page-Gould et al. (2008) evaluate a friendship building exercise that randomly paired white and Latinx college students for three one-hour meetings. They find no evidence for an effect of interethnic contact on initiating cross-group interactions 10 days after the end of the intervention on average, but they do report statistically significant positive effects for students with high initial prejudice.

Some other experiments report positive effects of inter-group contact without prolonged coresidence albeit for groups that are arguably related to, but different from, ethnicity, such as assignment to members of different castes in Indian cricket teams (Lowe 2020), or assignment of Christians and Muslims to vocational training courses in Nigeria (Scacco and Warren 2018) or to soccer teams in Iraq (Mousa 2020). Still farther afield conceptually from interethnic exposure, Rao (2019) investigates in-group bias in Indian schools and find that wealthy students exposed to poor classmates discriminate less against poor students.[3]

Like contact theory, conflict theory is supported by numerous observational, and mostly cross-sectional, studies (e.g., Alesina and La Ferrara, 2002; Delhey and Newton, 2005; Dinesen and Sønderskov, 2015; Legewie and Scheffer 2016; Stolle, Soroka, and Johnston, 2018), and also by a small number of recent experimental or quasi-experimental studies. For example, in a quasi-experimental study exploiting demographic changes in Chicago neighborhoods following the demolition of large public housing complexes, Enos (2016) found that voter turnout and the vote share for conservative candidates decreased sharply among whites as African American neighbors

[3] Enos and and Celaya (2018; Study 2) executed a small field experiment on intergroup contact between entirely made-up groups. Recruited participants were randomly assigned to groups (identified by holding orange vs purple folders) and then randomly seated either in mixed or segregated fashion in a university waiting room. Exposure lasted 5 minutes and participants were not allowed to interact. Upon exiting the waiting room, participants in the segregated condition evinced greater out-group bias with respect to the group that had held a different-color folder.

moved away. Using an instrumental-variables approach, Hangartner et al. (2019) find that an increase in refugee arrivals on Greek islands caused a large increase in native hostility toward refugees, immigrants, and Muslims. Yet more strikingly, in one of the few randomized field experiments on the topic, Enos (2014) found that randomly placing Spanish-speaking individuals at commuter train stations in Boston significantly increased exclusionary attitudes toward immigrants among white passengers.

Constrict theory, as a variant of conflict theory, currently lacks support from high-quality evidence. Putnam (2007) based constrict theory on the observation that more ethnically diverse neighborhoods in the United States have lower levels of in-group trust among whites. However, more ethnically diverse neighborhoods are not only less white, but also poorer and less stable than more ethnically homogenous neighborhoods (Abascal and Baldassarri 2015; Meer and Tolsma 2014). It is difficult to disentangle the effect of ethnic diversity on trust from the confounding effects of poverty, residential mobility, and other, potentially unmeasured, correlates of diverse locales in observational studies. The only randomized field experiment testing constrict theory (Finseraas et al. 2019) fails to support it.

The apparently inconsistent—positive and negative—effects of interethnic contact are generally reconciled by different scope conditions for contact and conflict theory (e.g., Abascal and Baldassarri 2015; Dinesen and Sønderskov 2015; and Valdez 2014). In Allport's (1954) classical formulation of contact theory, interethnic contact will promote positive intergroup relations when (i) contact is supported by an authority and both groups (ii) share equal status and (iii) cooperate in the pursuit of common goals. Subsequent research additionally emphasized the importance of (iv) close and prolonged interactions (e.g., Finseraas and Kotsadam 2017) and (v) friendship potential among the interacting individuals (e.g., Pettigrew 1998; Laurence 2009; Stolle, Soroka, and Johnston 2008; Van Laar et al. 2005).

The scope conditions for conflict theory lack a canonic statement, but include (i) fleeting, shallow, and non-repeated exposures (Enos 2014), or (ii) exposures occurring when in-group and out-group members manifestly compete over scarce resources, social rights, and social status (e.g., Bobo 1999; Semyonov Raijman, and Gorodzeisky 2006), or (iii) exposures to out-group members that are *perceived* to pose economic or cultural threats (Blalock, 1967; Williams, 1964; Bobo 1999; McLaren, 2003; Valdez, 2014).

The scope conditions for both theories are plainly vague. For contact theory, it is unclear how close is close enough; what characterizes friendship potential; and in what sense equal status is even possible in societies where the majority oppresses a minority. Regarding conflict theory: when do individuals not compete over some resource, and when can social scientists exclude the possibility of perceived cultural threat? These problems are multiplied if each scope condition must be evaluated along multiple dimensions, e.g., cooperation vs. competition with respect to multiple goals, or status equality with respect to multiple status criteria. Furthermore, even if individual scope conditions are unambiguously met, it is unclear to what extent scope conditions can trade off against each other or interact; for example, does enthusiastic cooperation in pursuit of a single overriding common goal compensate for status inequality or perceived cultural threat?

The unavoidable conclusion is that most real-world settings do not constitute ideal-typical matches for the scope conditions of either theory. This is not to say that those settings do not exist. For example, Norwegian military bootcamps may be considered well-nigh obvious settings for contact theory (Finseraas and Kotsadam 2017), because multiethnic teams spend most of their days pursuing externally mandated group goals that are evaluated at the team level, cooperation is enforced by drill sergeants, and equal status within teams is a stated policy of the military, symbolically manifested by identical rank, pay, and uniforms. By contrast, even in college dorms, which are the main setting in which prior field experiments found support for contact theory, one

might wonder to what extent roommates should be expected *ex ante* to cooperate in the pursuit of a common goal, rather than, say, compete for space and quiet time.

This ambiguity in the scope condition for contact and conflict theory triggers theoretical and practical concerns. While the field experimental evidence demonstrates that contact *can* reduce prejudice, it also demonstrates that contact *can* increase prejudice. Absent exhaustive, let alone unambiguously disjunct, scope conditions, however, it is very difficult to formulate *ex-ante* expectations about the effects of contact in new settings. This is a problem for theory, because two theories that make opposing predictions for the same situation cannot both be valid. It is a problem for policy inasmuch as promoting interethnic contact is a popular strategy for improving interethnic relations in new settings where the effects of contact have not yet been evaluated.

We take the view that the labor of refining the scope conditions of contact and conflict theory should include building an expansive evidence base that evaluates germane interventions in new settings with dependable research designs.

Empirical Setting and Theoretical Expectations

Setting

We test the effects of interethnic exposure on interethnic friendships and discrimination, and on intra-ethnic cooperation, in 186 3rd through 8th grade classrooms of 39 schools in rural Hungary. Hungary is an ethnically fairly homogenous country. Roma people constitute Hungary's largest ethnic minority, comprising 3 percent of the total population, and 12 percent of Hungarian youth.[4] Many Roma Hungarians speak the Romani language in addition to Hungarian; and many are recognizable by appearance to non-Roma Hungarians (Kertesi and Kézdi 2011b). Roma people suffer severe economic, social, and health disadvantages and are frequent targets of bullying,

[4] Estimates vary across methodologies (Morauszki and Papp 2015).

prejudice, and discrimination (Kertesi and Kezdi 2011a; Hajdu, Kertesi, and Kézdi 2017; 2018; Simonovits et al. 2018; Grow et al. 2016; Kisfalusi and Pál 2020; Kisfalusi et al. 2019).

Students in Hungary attend untracked compulsory primary schools from 1st through 8th grade, corresponding to elementary and middle school in the United States. Students attend local schools. Therefore, the ethnic composition of the student body in each school reflects the ethnic composition of the local catchment area. Students form stable classrooms that receive instruction in all core (and most other) subjects together. Furthermore, most subjects are taught in the same room (except physical education, and, depending on facilities and grade level, art, music, and the sciences). Seating charts are typically set by teachers and fixed for all subjects taught in the same classroom. The core subjects in primary school are Hungarian grammar (writing), Hungarian literature (reading), and mathematics. Grades in these subjects determine the subsequent allocation of students to tracked secondary schools, starting in 9th grade. Instruction is mostly lecture based, interspersed with group work between deskmates.

Interethnic contact in schools occurs at multiple levels. Students have limited exposure to schoolmates across grades, with whom they only share recess. Students spend more time with their classmates, with whom they share instruction in most subjects. Finally, students are most exposed to their deskmates, with whom they share the closest proximity throughout the school day, and with whom they routinely cooperate in group work, as we document below.

Scope conditions in context—ex-ante predictions

We test the causal effects of interethnic exposures between Roma and non-Roma students on interethnic friendship and discrimination against Roma at two different levels: first, at the desk level within classrooms; second at the grade level (which often consists of a single classroom) within schools.

Our *ex-ante* expectations about the fit between our setting and the scope conditions for contact and conflict theory determine our theoretical predictions for the effects of these exposures. We argue that deskmate exposures to ethnic others, even more so than grade-level exposures, best fit the scope conditions of contact theory by Allport (1954) and others, especially in the expansive interpretation of Pettigrew et al. (2011): *Support by an authority* for interethnic contact is self-evident: schools have assigned Roma and non-Roma students to the same classrooms, and teachers seat Roma and non-Roma students next to each other at the same desk. *Equal status* of students is given in the sense that Roma and non-Roma students share the same classroom as peers, have the same teacher, and are subject to the same curriculum, exams, and grading standards. Roma and non-Roma students frequently *collaborate* in shared tasks and group work *in pursuit of common goals.* For example, deskmates explain course material to each other, help each other with homework assignments during class time, play-act situations, and discuss problems, among other cooperative activities. Teachers in Hungarian primary schools report that the majority of deskmates collaborate almost every lesson (61 percent), and almost all deskmates collaborate at least once a week (95 percent).[5] Exposure is *long lasting*, in that students typically stay with their classmates from 1st through 8th grade, and seating charts are set for a whole semester. Finally, classrooms are ripe with *friendship potential*. Below, we document that non-Roma deskmates in our setting are indeed more likely to befriend each other than are non-Roma students who do not sit

[5] We conducted a survey on deskmate cooperation among homeroom teachers in all Hungarian primary schools online in February 2022 ($N = 656$ teachers from 288 schools). The survey prompted teachers for the frequency of nine collaborative activities among deskmates in their own classes. The list of activities was compiled from interviews with out-of-sample homeroom teachers. See Appendix Figure A3 for details. The findings of this national survey likely apply to our experimental sample, as reports of deskmate collaboration hardly varied across schools: neither the frequency nor the type of deskmate collaboration were predicted by respondents' school's location, size, achievement level, or share of Roma students (Appendix Table A13). This survey was not pre-registered with the original experimental protocol.

next to each other in our sample (Appendix C).[6] We confirm buy-in for aspects of these scope conditions in a national survey of homeroom teachers. Almost all teachers agree that deskmates affect each other's academic achievement, behavior, and friendship formation, so that many teachers seat weak students next to strong students.[7]

We therefore expected that sitting next to, and sharing a grade level with, Roma students will increase interethnic friendships and decrease discrimination by non-Roma students, in line with contact theory.

Like most natural settings, however, one can also easily argue that our setting strains against the scope conditions of contact theory and instead fits the scope conditions of conflict and constrict theory. For example, the degree of actual collaboration between deskmates and classmates in pursuit of common goals is likely less pronounced than, e.g., in U.S. schools, and deskmates may also compete for grades, just like the degree of cooperation may not be extensive between college roommates, who may also compete for privacy. Similarly, equal status as students within the formal framework of Hungarian education may not prevent community prejudice from inflecting interactions among students, or between students and teachers, much like randomly assigning college roommates does not prevent social hierarchies from entering dorm rooms. One

[6] Here, we specifically refer to intra-ethnic friendships among non-Roma students to avoid confusing the description of our setting with the effects of our intervention on inter-ethnic outcomes.

[7] We fielded this survey among 7th and 8th-grade homeroom teachers in all Hungarian primary schools online in late 2021 ($N = 413$ teachers from 266 schools). Responding schools were nationally representative. Between 92 and 98 percent of respondents reported that deskmates influence each other's academic achievement, behavior, and diligence. Less than 3 percent believe that deskmates do not affect each other at all. Thirty-nine percent of homeroom teachers who employ a seating chart specifically place strong and weak students together. See Appendix Figures A1 and A2. This survey was not pre-registered with the original experimental protocol.

might therefore also legitimately expect that sitting next to, and sharing a classroom with, Roma students might worsen inter- and intra-ethnic relations, in line with conflict and constrict theory.

Our pre-analysis plan registered positive expectations for the effects of interethnic exposure at the desk-level, because we judge the fit between our setting and the scope conditions of contact theory to be roughly on par with the fit in studies of college roommates, which form the backbone of field experimental support of contact theory. However, in order to honor the ambiguity of the scope conditions, we pre-registered two-sided statistical tests at all levels of analysis, which enables testing the predictions not only of contact, but also of conflict and constrict theory.

Our ability to test contact, conflict, and constrict theory in the same design means that our study tests whether our novel setting meets the scope conditions of either theory. Hence, our study generates new insights not only into the scope conditions of each theory, but also into the boundary conditions between theories.

Study design

A large methodological literature in sociology, economics, and statistics documents the severe and often unexpected challenges of identifying causal peer effects when social relations (e.g., network ties or contextual exposures) are not randomly formed by outside intervention (e.g., Sobel 2006; Shalizi and Thomas 2011; VanderWeele and An 2013; Angrist 2014; Ogburn and VanderWeele 2014; Manski 1993). Our study avoids these difficulties by randomizing deskmate relationships in the primary analysis, and by exploiting plausibly random variation in ethnic composition across grades in exploratory fixed-effects analyses.

Our study is distinguished by three main design elements. First, we execute a randomized vignette experiment to measure interethnic discrimination of non-Roma students against Roma students and intra-ethnic cooperation among the non-Roma majority. Second, we randomize the seating charts within classrooms for the duration of one semester to evaluate the effects of

exposure to Roma vs. non-Roma students at the desk level within classrooms. Third, we exploit quasi-random variation in the share of Roma students across grades within schools to evaluate the effects of grade-level exposures. Outcome data were collected by the field team in the spring of 2018, including the randomized survey-vignette experiment.[8]

Vignette experiment to measure discrimination

In order to measure ethnic discrimination, we designed a two-question randomized survey experiment (as part of a larger survey instrument) that presented students with a scenario to lend money to their classmates during a hypothetical field trip to the zoo. Vignette experiments are commonly used for eliciting attitudes that cannot be inferred from manifest behavior, or when researchers are worried about priming or social desirability bias (Atzmüller and Steiner 2010; Hainmueller, Hangartner, and Yamamoto 2015). Vignette experiments have previously been used to study discrimination (e.g., Finseraas et al. 2016; Jakobsson et al. 2016).

The first question (Question 9a) was asked of all students to introduce the scenario by eliciting students' willingness to lend money to their deskmate. It reads (in translation):

> "*Imagine that you are going to the zoo with some of your classmates. Your deskmate (whom you sat next to in Hungarian class in December) has forgotten to bring money for the entrance ticket. You have enough money for two entrance tickets. Would you lend your deskmate the money for the entrance ticket?*"[9]

[8] We pre-tested the survey instrument in one out-of-sample school in the same geographic area in the spring of 2017.

[9] We note that this first question measures neither discrimination nor the effect of deskmates on discrimination. Suppose, for example, that fewer non-Roma students sitting next to a Roma deskmate are willing to lend to their deskmate than are non-Roma students sitting next to a non-Roma deskmate. This could occur either because non-Roma students would discriminate against Roma people regardless of the ethnicity of their deskmate, or because close contact with a Roma deskmate caused discrimination in the first place. More formally, this single question does not provide all potential outcomes needed to separate discrimination from deskmate effects on discrimination. See also the discussion in Appendix C.

In order to measure discrimination, we next randomly assigned students to different ethnic prompts. Question 9b reads:

> "*Now imagine that it is not your deskmate, but a different classmate, who has forgotten to bring money with him/her. [This classmate is a Roma/gypsy.] Would you lend this [Roma/gypsy] classmate the money for the entrance ticket?*"[10]

The ethnic prompt in brackets was only presented to a random half of the students. The answer categories were "Yes," "No," and "I do not know." Since the only difference in the prompts to lend money to a classmate is whether or not the imagined classmate is identified to be of Roma ethnicity, and since this difference is randomly assigned, the difference in the responses to this question measures discrimination, i.e., the causal effect of the hypothetical classmates' ethnicity on students' readiness to lend to the classmate.[11]

Randomizing deskmates

The field experiment manipulated interethnic contact between Roma and non-Roma students by randomizing the seating chart within 3rd through 8th grade classrooms in rural Hungarian schools

[10] Although the Hungarian term for "gypsy", like the English term, carries derogatory connotations, it is a commonly used term in Hungary, and it is the only descriptor of this minority group known to many school children. By contrast, the term "Roma" is not universally known to non-Roma school children. Hence, we prompt for both "Roma" and "gypsy." We did not specify the ethnicity of the hypothetical classmate when not prompting for a Roma classmate, because we did not want to make ethnicity salient for the control group.

[11] Readers might object that prompting to lend money to an "imagined" Roma classmate may cause students to think of a real Roma classmate. If that Roma classmate, for argument's sake, is objectively less trustworthy, then unwillingness to lend to this classmate would not measure discrimination. Against this objection, we argue that, in classrooms without any Roma students, students cannot imagine a specific classmate and would hence either have to imagine what it would mean for one of their non-Roma classmates to be Roma, or to substitute an imagined individual for the classmate. Either way, respondents' answers would reveal ethnic stereotypes, so that manifest differential willingness to lend to a Roma classmate would capture discrimination (either statistical or animus). Empirically, the differential willingness to lend to a Roma classmate versus a classmate of unspecified ethnicity is virtually the same in the 123 classrooms with, and the 52 classrooms without, Roma students ($\beta_{Diff} = -0.01, p = 0.75$). This indicates that our strategy measures discrimination in all classrooms.

for the three core subjects Hungarian reading, Hungarian literature, and mathematics. Depending on grade level, these subjects account for approximately 7 to 10 hours per week, or between 25 and 45 percent of the school week.[12] Students were randomly allocated to freestanding front-facing desks seating two students each, based on class lists provided by the schools (using a random number generator). The intervention commenced at the beginning of the 2017 fall semester, and we encouraged adherence to the seating chart until the end of the semester in January 2018.

Quasi-random variation in the share of Roma students across grades

We investigate the effect of grade-level exposure to Roma students using a quasi-experimental strategy that exploits as-if random variation in the share of Roma students across grades. This analysis builds on the standard assumption that variation in student composition across grades is driven by random fluctuations in birth rates across cohorts (Hoxby 2000; Gould, Levy, Passerman 2009; Kim 2020; Lee and Lee 2020). This assumption is plausible because most schools in the sample are the only primary school in the town or settlement, essentially ruling out selection across cohorts via school choice. Additionally, our grade-level analyses control for school-level fixed effects, which control for school-based selection in the few locations where students have a choice between multiple schools. In most schools, our grade-level analysis simultaneously captures grade- and classroom-level exposure to Roma students, because each grade only contains one classroom.[13]

Data, Coding, and Balance

This section describes our data and assesses experimental balance (see Appendix B for details). A detailed pre-analysis plan was archived on March 22, 2018, prior to the receipt of endline data in

[12] To the extent that additional subjects were taught in the same room, students likely spent more time sitting next to their deskmate, as seating charts are specific to rooms rather than subjects.
[13] We also run a classroom-level analysis as a robustness check, which similarly controls for school-fixed effects and additionally controls for average classroom-level characteristics to (a) control for within-grade selection of students across classrooms, and (b) possible selection into our sample for grades in which some, but not all, classrooms participated in the study. Results are unchanged (Appendix Table A11).

June 2018. Any deviation from the pre-registered plan is noted in the text. The data collection instrument and the pre-analysis plan are reproduced in Appendix D. Replication data are available online at [to be released upon publication]. We obtained written informed consent from school principals, teachers, and parents at multiple points. Consent procedures did not reveal our interest in ethnicity.[14] The study was approved by the IRB offices at the Center for Social Sciences, Budapest and the University of Wisconsin-Madison.

Data: Based on a set of pre-registered inclusion criteria, our base sample consists of 3,184 Roma and non-Roma students across 186 classrooms of 39 schools.[15] In order to focus on the effects of exposure to a stigmatized minority group on the ethnic majority group, and to avoid certain statistical complications (Angrist 2014:106), we pre-registered to exclude the Roma students themselves as subjects from all analyses.[16] The final analysis sample includes 2,395 non-Roma students in 175 classrooms and 39 schools. *Ex-ante* power calculations were based on an anticipated sample of at least 2000 non-Roma students; hence, our sample is sufficiently large to detect meaningful effects.

Treatments: We pre-registered three main treatment variables for the primary analyses: one for the deskmate intervention, one for the vignette experiment, and one for the cross-product of the two. The first treatment variable, *RomaDeskmate*, equals 1 if a student is randomized to sit next to a Roma deskmate at the beginning of the fall semester, and 0 otherwise. Students' ethnicity was reported by classroom teachers at the beginning of the study. By using assigned rather than actual deskmates, we perform an intention-to-treat analysis that is not biased by the endogenous

[14] Consent concerned permission for randomizing the seating chart and data collection in general. We further avoided priming participants for our interest in ethnicity by asking both teachers and students any questions related to ethnicity only once in the larger baseline teacher survey and the endline student survey, respectively.
[15] These numbers differ slightly from those expected in our pre-registration document, submitted before the receipt of data.
[16] Consequently, we excluded 11 all-Roma classrooms.

seating choices made by students and teachers after randomization. Compliance with the intended seating chart was high: 86 percent of students in the analysis sample were seated in compliance with their experimental assignment during school visits of the field team in late Fall.[17]

The second treatment variable is coded from the vignette experiment conducted at endline. The variable *RomaVignette* equals 1 if the vignette (Question 9b, reported above) asks students whether they would lend money to a Roma classmate, and equals 0 if the vignette asks students to lend money to a classmate of unspecified ethnicity. The third treatment variable is the cross-product between the first two, i.e., *RomaDeskmate*RomaVignette*. This cross-product equals 1 if a student sitting next to a Roma deskmate is asked to lend money to a Roma classmate, and 0 otherwise.

For our quasi-experimental secondary (exploratory) analyses, we include two additional treatment variables: the standardized share of Roma students in the grade level in our sample, *RomaShare*; and its interaction with being prompted to lend to a Roma classmate in the vignette experiment, *RomaShare*RomaVignette*.

Outcomes: Outcome variables were collected via a 45-minute student survey at endline (see Appendix D), which elicited ego-centric network data and contained the survey experiment.[18] The two versions of the endline questionnaire containing the two versions of the survey vignette were distributed to students in random order (using a random number generator).

We pre-registered two main outcome variables: *LendToClassmate* and *RomaFriend*. The outcome variable *LendToClassmate* is based on the survey vignette experiment (Question 9b,

[17] Neither low compliance, nor the share of compliance, is correlated with the share of Roma students in the classroom. Conclusions remain the same in a pre-registered robustness check that restricts the sample to classrooms with at least 90 percent compliance (see Appendix Table A.1).
[18] The survey also asked several other questions that were pre-registered not to be analyzed in this paper.

described above) and equals 1 if the respondent answers that they would lend money to the classmate, and 0 if they would not (or did not know if they would). *RomaFriend* captures whether the student has a Roma friend among his or her best friends (inside or outside of the classroom). Survey Question 5 generically prompts:

"*Now, in general, think of your best friends, not just in the class but EVERYWHERE.*"

Question 5d subsequently prompts specifically:

"*Among your best friends, how many are Roma (gypsy)?*".

The outcome variable *RomaFriend* equals 1 if the individual has at least one Roma best friend, and 0 otherwise. Henceforth, we refer to "friends" and "best friends" interchangeably. Additionally, we pre-registered several secondary outcome variables for exploratory analyses, including the number of Roma friends within the classroom (which we measured using name generators, employing Smith's [2002] "network approach"), and whether students liked their deskmate. Analyses for these secondary outcomes are reported in Appendix C.

Covariates: We collected baseline covariates from homeroom-teacher reports, including students' age (coded in 0.1 years from birthdates), gender, and baseline (spring-semester 2017) grades in five core subjects (Hungarian literature, Hungarian grammar, mathematics, diligence, and behavior), coded on a five-point integer scale, where 1 is worst and 5 is best.[19] We filled in missing baseline grades from students' retrospective self-reports at endline.[20] We coded missing values on other covariates as zero (affecting less than 5 percent of values) and included dummy variables controlling for missing status in order to retain observations.[21] For analyzing effect heterogeneity

[19] The lowest grade in diligence and behavior is 2.
[20] About 3 percent of grades per subject were filled from student self-reports. This un-pre-registered choice did not affect conclusions.
[21] Appendix Table A.2 shows that the results are very similar if instead we exclude all observations with missing values on the control variables.

across groups defined by baseline variables, we used students' baseline GPA, defined as the average of the five subject-specific grades; an indicator for student's and deskmate's gender concordance; and the share of Roma students in the classroom.

Table 1 presents descriptive statistics. In our sample of 2,395 non-Roma Hungarian students, 291 (12 percent) have a Roma deskmate (treated) and 2104 (88 percent) had a non-Roma deskmate (control). The sample is evenly split by gender. The average age is 11.8 years at baseline. Thirty percent of non-Roma Hungarian students in the sample have at least one Roma friend inside or outside of the classroom. This share naturally differs substantially across treated and control students because this descriptive comparison does not account for the substantial difference in the share of Roma students across classrooms; our subsequent analysis remedies all classroom-level imbalances by controlling for pre-registered classroom fixed effects. We also see that 57 percent of students are willing to lend money to a classmate in the vignette experiment.

(Table 1 here)

There is a limited amount of missingness on the primary outcomes, as only 2102 (88 percent) and 2124 (89 percent) out of 2395 students in the analytic sample answered the survey questions about lending and friendship, respectively. The main reasons for missingness were student absences and lacking parental consent for the endline survey. Importantly, this attrition is non-differential, i.e., unrelated to treatment status: the p-values for *RomaDeskmate* are above 0.9 in regressions of attrition on *RomaDeskmate* and classroom fixed effects.

Balance checks: The purpose of randomizing treatments is to eliminate selection bias by creating comparable ("balanced") treatment and control groups. Table 2 tests balance on observed covariates. Columns 1-7 test balance between students who were randomized to have a Roma vs. non-Roma deskmate by regressing *RomaDeskmate* on each of the baseline covariates, and on classroom fixed effects to account for experimental design. Column 8 includes all baseline

covariates simultaneously. Associations between the covariates and having a Roma deskmate would indicate a failure of randomization. We see that there are no apparent differences across the treated and control groups. The F-test for whether the control variables jointly predict having a Roma deskmate has a p-value of 0.25, indicating that the randomization achieved balance on baseline covariates. Column 9 reports an analogous test to assess balance in the vignette experiment, similarly indicating balance on baseline covariates ($p = 0.83$).

(Table 2 here)

Empirical strategy

Experimental analysis: desk-level exposure

To analyze the causal effects of being prompted to lend money to a Roma vs. a non-Roma classmate on non-Roma students' willingness to lend money (discrimination), and the effect of sitting next to a Roma deskmate on outgroup discrimination and in-group cooperation, we estimate the following regression on the sample of non-Roma students:

$$LendToClassmate_{ic} = \beta\ RomaDeskmate_{ic} + \gamma\ RomaVignette_{ic}$$
$$+\delta\ RomaDeskmate_{ic} * RomaVignette_{ic} + \boldsymbol{\omega}'\boldsymbol{X}_{ic} + \boldsymbol{\xi}_c + \varepsilon_{ic} \qquad (1)$$

where *i* indexes individuals, *c* indexes classrooms, $RomaDeskmate_{ic} = 1$ if the student was randomly assigned to sit next to a Roma deskmate, $RomaVignette_{ic} = 1$ if the student was randomly prompted to lend money to a Roma classmate in the vignette experiment, $\boldsymbol{X}_{ic}$ is the vector of student's baseline covariates (described above), $\boldsymbol{\xi}_c$ is a vector of classroom fixed effects (subsuming the intercept) to control for experimental design (randomization within classrooms), and ε_{ic} is an individual error term. Since both deskmates and the vignette are randomly assigned, the coefficients, $\beta, \gamma,$ and δ, identify causal effects.

The coefficient for $RomaVignette$, γ, identifies discrimination against Roma students by non-Roma students, i.e., the differential willingness of non-Roma students (without a Roma deskmate) to lend money to a hypothetical Roma classmate. Since the vignette was assigned randomly, this coefficient has a causal interpretation; and since deskmates were assigned randomly, the estimate is representative of all non-Roma students, had they been seated next to a fellow non-Roma student.

Contact and conflict theory are tested by the coefficient on the interaction between deskmate assignment and the vignette prompt, δ, using a difference-in-difference logic: δ gives the differential effect of interethnic contact (sitting next to a Roma deskmate vs. not) on lending money to a classmate when that classmate is identified as Roma vs. not. A positive and statistically significant estimate, $\delta > 0$, would support the contact hypothesis that sitting next to a Roma deskmate diminishes discrimination toward Roma, whereas a statistically significant negative estimate, $\delta < 0$, would support the conflict hypothesis that sitting next to a Roma deskmate increases discrimination.

We cautiously interpret the coefficients for $RomaDeskmate$, β, as a test of constrict theory. To sharpen this interpretation, we divide the sample into 33 classrooms that are majority Roma and 142 classrooms that are majority non-Roma, and we rely on the auxiliary hypothesis that, in classes that are majority non-Roma, being prompted to lend to a classmate of unspecified ethnicity de facto prompts students to lend money to a non-Roma classmate. Hence, in majority-non-Roma classrooms, the coefficient for $RomaDeskmate$ tests whether having a Roma deskmate affects non-Roma students' willingness to lend money to a presumed non-Roma classmate, i.e., whether it affects in-group trust. According to constrict theory, exposure to ethnic others should lead to lower in-group trust, $\beta < 0$.

Our second main outcome is having any Roma friends (inside or outside of the classroom, coded from question 5d, described above). We estimate the following linear-probability model to identify the effect of random assignment to sit next to a Roma deskmate on the probability of having a Roma friend:

$$RomaFriend_{ic} = \zeta\ RomaDeskmate_{ic} + \theta\ \boldsymbol{X}_{ic} + \boldsymbol{\xi}_c + \nu_{ic} \qquad (2)$$

where indices and variables are defined as before, and ν_{ic} is the individual-level error term. A positive and statistically significant estimate $\zeta > 0$ would support the contact hypothesis that sitting next to a Roma student increases the student's chance of naming a Roma (inside or outside of the classroom) among their best friends. By contrast, $\zeta < 0$ would lend support to conflict theory that sitting next to a Roma diminishes the chance of interethnic friendship.

We present all results with and without the baseline controls, $\boldsymbol{X}_{ic}$. Our primary experimental specification is without controls. Control variables are included only as they may shrink standard errors and increase power (although, in our case, they did not). To avoid distortions from functional form restrictions (e.g., falsely assuming linear effects of the covariates), we add all control variables as series of indicator variables (Athey and Imbens 2017). We report heteroscedasticity-robust Huber-White standard errors for all models. Standard errors do not need to be clustered at any level, as randomization occurred at the individual level (Abadie et al. 2017).

To guard against the possibility that our main specifications average over sociologically interesting effect heterogeneity across groups, we further pre-registered a series of exploratory specifications that interact the treatment variables with all covariates of students and deskmates, and with the share of Roma students in the classroom or grade level.

Quasi-experimental evaluation: Grade-level exposure

We explore the causal effect of grade-level variation in exposure to Roma students using the following fixed-effects specification,

$$LendToClassmate_{igs} = \eta\ RomaShare_{gs} + \kappa\ RomaVignette_{igs} +$$

$$\lambda\ RomaShare_{gs} * RomaVignette_{igs} + \alpha\ \boldsymbol{X}_{igs} + \boldsymbol{\chi_s} + \epsilon_{igs}, \qquad (3)$$

where *i*, *g*, and *s* index non-Roma students, grade levels, and schools, respectively, and $\boldsymbol{\chi_s}$ is a vector of school-level fixed effects (subsuming the intercept). The interpretation is analogous to the desk-level analysis. Specifically, λ identifies the causal effect of a change in the share of Roma students at the grade level on anti-Roma discrimination under the fixed-effects assumption. We estimate this regression with and without covariates, $\boldsymbol{X}_{igs}$. Standard errors are clustered at the grade level, because exposure is at the grade level (Abadie et al. 2017).

Results

Desk-level exposure effects

Table 3 presents our primary field-experimental results from equation (1). Column 1 shows our main pre-specified regression for the causal effects of having a Roma deskmate, and of being prompted to lend to an unnamed Roma classmate, on lending to that classmate, with classroom fixed effects but without other covariates. We start by demonstrating the presence of anti-Roma discrimination using our vignette experiment. We see that the coefficient for $RomaVignette$ is negative, substantively large, and statistically significant, implying that non-exposed students (those who are not sitting next to a Roma student) are considerably less willing to lend to a Roma classmate than to a classmate of unspecified ethnicity. Being randomly prompted to lend to a Roma classmate lowers the probability that a non-Roma student is willing to lend money by 18 percentage points, a 27 percent decline from the mean willingness of 67 percent to lend to a

classmate in the control group (non-Roma vignette and non-Roma deskmate). We conclude that non-Roma students strongly discriminate against Roma students.[22]

(Table 3 here)

Next, we estimate whether sitting next to a Roma student affects non-Roma students' differential willingness to lend to a Roma student, as captured by the coefficient on the $RomaDeskmate * RomaVignette$ interaction. The coefficient is close to zero, and not statistically significant ($\delta = -0.03, p = 0.61$). This indicates that having a Roma deskmate did not meaningfully exacerbate, or diminish, anti-Roma discrimination and hence fails to support both contact and conflict theory.

The coefficient on $RomaDeskmate$ estimates the effect of being assigned to a Roma deskmate on willingness to lend money to a classmate who was not specifically identified as Roma. This effect is close to zero and not statistically significant ($\beta = -0.03, p = 0.57$). In appendix Table A.10, we show that the coefficient is similarly substantively small and not statistically significant in majority non-Roma classrooms. By the logic explained above, our experiment thus fails to support the constrict hypothesis that exposure to the ethnic outgroup reduces cooperation within the ethnic majority group. Adding baseline covariates in Column 2 does not change the results of Column 1.

Column 3 of Table 3 presents estimates for the effect of having a Roma deskmate on non-Roma students' probability of counting at least one Roma person among their best friends (inside or outside the classroom). The estimate is very close to zero and is not statistically significant ($\zeta = -0.03, p = 0.34$). Hence, we find no support for the notion that sitting next to a Roma student affects interethnic friendships for non-Roma Hungarian students. Adding baseline covariates in

[22] Anti-Roma discrimination also aligns with additional—descriptive—measures of ethnic antipathy in our study. Non-Roma students who sat next to a Roma (rather than another non-Roma) student were less inclined to like their deskmate and to lend to their deskmate. We discuss these results, which should be interpreted descriptively rather than causally, in Appendix C.

column 4 does not change this result.[23] Importantly, however, sitting next to each other did increase the probability of being friends among non-Roma students (see Appendix C and Rohrer et al. 2021). In addition to confirming the conventional finding that propinquity affects friendship in general (Segal 1974; Back, Schmukle, and Egloff 2008) the finding that close exposure promotes friendship within the ethnic majority is important as it demonstrates that our deskmate treatment was not globally inert for all outcomes: Sitting next to each other evidently conferred new information to students and had demonstrable effects on some outcomes, but it specifically did not affect interethnic discrimination or interethnic friendship.

We extensively explored the possibility that the effects of sitting next to a Roma student on anti-Roma discrimination or interethnic friendships varied by students' and deskmates' baseline characteristics (age, gender, and baseline grades). However, we found little indication of substantively meaningful or statistically significant effect heterogeneity with respect to any specific group (appendix Tables A3-A9). [24] We also found no evidence that the effect of micro-level

[23] Additional pre-registered secondary analyses, reported in Appendix C, further support this conclusion. There, we found no evidence that sitting next to a Roma deskmate affected non-Roma students' (a) *number* of Roma friends inside or outside of the classroom, or (b) the probability of having a Roma student among one's five best friends inside the classroom.

[24] Appendix Tables A.3 and A.4 show that there is no indication of heterogenous treatment effects based on students' own age, gender, or baseline grades. Appendix Tables A.5 and A.6 show the absence of heterogenous effects with respect to deskmates' characteristics. Appendix Table A.7 shows that there are no statistically significant treatment effects for same-sex deskmates or sex discordant deskmates. We further pre-specified that we would investigate effect heterogeneity by whether or not the student was willing to lend money to his or her deskmate. In Appendix Table A.8 we show that students who are willing to lend to their deskmate are also more willing to lend to a Roma classmate, but the treatment effect of exposure to a Roma deskmate is not statistically significantly different between those willing to lend to their deskmates and those who are not. Table A.9 shows that the effects of deskmate contact did not differ across grade levels. Tables A.10 shows that the effects of deskmate contact did not differ across minority vs majority Roma classrooms. As we found no indication of effect heterogeneity in any of these analyses, we did not proceed with the machine learning techniques that we had pre-specified to validate the search for heterogenous treatment effects.

exposure to Roma deskmates varies with the meso-level share of Roma classmates (appendix Tables A.10 and A.12).

Grade-level exposure effects

Table 4 shows the results of our quasi-experimental fixed-effects analysis for the causal effect of the share of Roma students in the grade level on anti-Roma discrimination and interethnic friendships. Like the primary experimental analysis above, being prompted to lend to a Roma classmate reduces non-Roma students' willingness to lend money by 19 percentage points ($p < 0.01$, Column 1), demonstrating strong ethnic discrimination. Increasing the share of Roma students in the grade level by one standard deviation (about 17 percent of students) does not meaningfully diminish this discrimination: the coefficient on the interaction $RomaShare * RomaVignette$ is substantively small, precisely estimated, and not statistically significant ($\lambda = 0.02, p = 0.24$). Results are essentially unchanged when additionally controlling for individual-level covariates (Column 2).

Column 3 shows that increasing the share of Roma students in the grade level by one standard deviation increases the probability of having at least one Roma friend (inside or outside of the classroom) by around 13 percentage points. This effect is statistically significant ($p < 0.01$), with or without covariates (Column 4).

In sum, while increasing the share of Roma students in the grade promotes friendship formation, it does not affect the degree of discrimination in lending against Roma students. Results are the same when analyzing classroom-level exposure to Roma students (Appendix Tables A.11 and A.12). Therefore, neither deskmate- nor grade-level exposure to ethnic others affect out-group discrimination positively or negatively, thus failing to support either contact or conflict theory.

Discussion and Conclusion

Theoretical predictions for the causal effects of interethnic contact on cooperation and discrimination are ambiguous. Contact theory posits that close, prolonged, and collaborative interethnic exposure under conditions of institutionally supported equal status will reduce prejudice and discrimination against ethnic others. By contrast, conflict theory suggests that shallow or competitive exposure to ethnic others may increase outgroup antipathy and discrimination, and constrict theory predicts that interethnic exposure will weaken ingroup cooperation.

In a sea of correlational evidence, a small number of well-identified experimental studies previously found evidence both for contact and for conflict theory. The field-experimental evidence for the contact theory of interethnic exposure, however, has almost exclusively accrued in rarified settings (Dixon et al. 2005; Paluck et al. 2019), begging the question of when the positive effects of interethnic contact turn negative, and of whether promoting interethnic contact holds promise for promoting desired social change on a broad scale.

We executed a well-powered, pre-registered, randomized field experiment that tests whether prolonged deskmate-level exposure to Roma students in Hungarian schools affects discrimination and interethnic friendship. Our setting *a priori* best fits the received scope conditions of contact theory. Hence, our experiment tests the generalizability of contact theory to quotidian settings, and, conversely, also the reach of conflict theory.

Empirically, we find that being seated next to a Roma deskmate for the duration of one whole semester does not affect non-Roma students' anti-Roma discrimination or the probability of having a Roma friend in either direction. These Null findings hold on average and within all subgroups of the study population. We further leveraged a quasi-experimental fixed-effects strategy to test whether interethnic contact at the grade level affects anti-Roma discrimination.

While we find that a larger share of Roma students in the grade promotes interethnic friendships, changes in grade-level ethnic composition still do not affect anti-Roma discrimination.[25]

Our findings thus disappoint the hope of contact theory that an easy intervention of increasing spatial proximity through deskmate assignments, or even varying the ethnic composition of schools, will ameliorate ethnic discrimination in settings that a priori fit the received scope conditions of contact theory. At the same time, our results also assuage the fears of conflict theory that increased exposure to ethnic others will lead to more discrimination. Finally, our study also rejects Putnam's constrict theory that exposure to ethnic minorities weakens cooperation within the majority group.

Why does interethnic contact fail to reduce discrimination in our setting, contrary to the prediction of contact theory? Despite strong priors that our setting fits the received scope conditions for contact theory, perhaps it lacks in the particulars?

For example, perhaps sharing a desk for the duration of one semester (5 months) is insufficiently long? Indeed, exposure to ethnic others in the college roommate studies that are the mainstay of field-experiment support for contact theory lasts longer (8-9 months). On the other hand, exposure in experiments in the Norwegian military, which similarly find positive effects of interethnic contact, ran shorter (2 months). Additionally, since classroom composition is highly stable, our grade-level analysis of 3rd through 8th graders investigates exposures that have lasted for

[25] The grade-level analysis also addresses the possible concern about our desk-level analysis that sitting next to a deskmate may confer no additional signal to students who already share a classroom, as the validity of the grade-level analysis would not be affected by cross-desk, within-classroom, spillovers. More importantly, the concern that our deskmate intervention may be inert is also empirically refuted for the desk-level analysis: Sitting next to each other does confer a signal beyond sharing the same classroom, as it affects friendship formation among non-Roma students (Appendix C and Rohrer et al. 2021). This demonstrates that our desk-level analysis does not suffer a generic lack of power due to possible cross-desk, within-classroom, spillovers.

multiple years. This suggests that length of exposure alone is not a decisive factor in explaining our Null results.

Or, perhaps, sharing a desk or a grade with ethnic others is insufficiently intense or collaborative? Indeed, the strongest evidence for contact theory derives from field experiments that enforce coresidence in college dorm rooms, or that enforce both coresidence and collaboration in military squadrons. The intensity of such interventions may be hard to replicate in everyday life (Paluck and Green 2009), our evidence for frequent collaboration between deskmates notwithstanding.

Or perhaps institutional support for equal status is insufficient in Hungarian schools, compared to the outspoken anti-racist ethos in elite American colleges and the structured environment of the military that is augmented by comparatively lower levels of community discrimination in the U.S. and Norway compared to Hungary (cf. Simonovits et al.2018)? That is, perhaps the positive effects of interethnic contact presuppose relatively low levels of ethnic inequality in the first place?

But if the scope conditions for contact theory really need to be as stark as sharing rooms, or if contact only has positive effects in settings where (relatively more) equal status is already achieved in the wider community, then there is little hope for improving interethnic relations by promoting mere interethnic exposure in settings where such improvement is needed most. For policy, this raises the specter that contact interventions in scalable, everyday settings, such as rearranging the desk chairs in classrooms, may only amount to cosmetic changes, much like rearranging the deck chairs won't affect the course of a ship.

For theory, our results plainly show that the scope conditions for contact theory are less expansive than previously thought. Recall, for example, that Pettigrew et al. (2011) argued that even vicarious contact with ethnic others through friends of friends improves ethnic relations. By

contrast, we find no effect of close and collaborative contact that lasted several months. Our findings indicate that the uncharted territory between the canonic scope conditions for contact theory and conflict theory is potentially vast. Perhaps, as-yet unknown additional scope conditions might better delineate the domains of contact and conflict theory. The current evidence base, however, appears insufficient for inferring such scope conditions, as the number of potential scope conditions (including possible non-linearities and interactions between existing scope conditions) exceeds the number of available studies. [26] Lacking data, we advise modesty when extrapolating claims about the effects of close interethnic contact from existing studies to new settings and urge the systematic empirical evaluation of additional scope conditions as a topic for future research.

Broadly, our findings thus continue the disconcerting trend that well-identified studies are often less supportive of popular social theories than are observational studies (Paluck et al. 2019). This suggests that sociologists should run more randomized field experiments in order to refine and disambiguate between competing social theories to inform concrete policies for improving interethnic relations.

References

[26] Take, for example, age as a prima-facie plausible new scope condition for contact theory. On one hand, successful field experiments on interethnic contact have mostly studied young adults, whereas we study pre-teens and adolescents (mean age 12) up to 8th grade. Hence, it is possible that interethnic contact generically does not affect ethnic discrimination in our age group. On the other hand, (a) we find no evidence that effects in our study vary across age (grade level), (b) Green and Wong (2009) report effects among high-school students, (c) racial bias is known to be more malleable in childhood than in adulthood (e.g., Gonzales, Dunlop, and Baron 2017), and (d) Rao (2019) finds positive effects of inter-group contact among even younger children at elite primary schools in India (albeit for socio-economic rather than ethnic groups). Hence, if age were a scope condition for contact theory, its effect would have to be distinctly non-linear or interact with other scope conditions.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" Technical Report, *National Bureau of Economic Research*.

Abascal, Maria, and Delia Baldassarri. 2015. "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." *American Journal of Sociology* 121 (3): 722–782.

Alesina, Alberto, and Eliana La Ferrara. 2002. "Who Trusts Others?" *Journal of Public Economics* 85 (2): 207–234.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Reading: Addison-Wesley.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98–108.

Athey, Susan, and Guido W. Imbens. 2017. "The Econometrics of Randomized Experiments," Pp. 73–140 in Banarjee, Abhijit, and Esther Duflo, *Handbook of Economic Field Experiments*, Vol. 1, Elsevier, Oxford University Press, Oxford.

Atzmüller, Christiane, and Peter M. Steiner. 2010. "Experimental Vignette Studies in Survey Research." *European Journal of Research Methods for the Behavioral and Social Sciences* 6(3): 128–138.

Back MD, Schmukle SC, Egloff B. "Becoming friends by chance." *Psychological Science* 19(5):439–40.

Baldassarri, Delia, and Maria Abascal. 2017. "Field Experiments Across the Social Sciences." *Annual Review of Sociology* 43 (1): 41-73.

Blalock, Hubert M.. 1967. *Toward a Theory of Minority-group Relations*. New York: Wiley.

Bobo, Lawrence D. 1999. "Prejudice as Group Position: Microfoundations of a Sociological Approach to Racism and Race Relations." *Journal of Social Issues* 55 (3), 445–472.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96 (5): 1890–1905.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8 (1): 1-32.

Burns, Justine, Lucia Corno, and Eliana La Ferrara. 2016. "Interaction, Prejudice and Performance. Evidence from South Africa." Technical Report, Working Paper Bocconi University

Camargo, Braz, Ralph Stinebrickner, and Todd Stinebrickner. 2010. "Interracial friendships in college." *Journal of Labor Economics* 28 (4): 861-892.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2019. "The Impact of College Diversity on Behavior Toward Minorities." *American Economic Journal: Economic Policy* 11 (4): 159–82.

Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to do Open Science*, University of California Press.

Dahl, Gordon B., Andreas Kotsadam, and Dan-Olof Rooth. 2018. "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams." Technical Report, Institute for the Study of Labor (IZA).

Delhey, Jan, and Kenneth Newton. 2005. "Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?" *European Sociological Review* 21 (4): 311–327.

Dinesen, Peter Thisted, and Kim Mannemar Sønderskov. 2015. "Ethnic Diversity and Social Trust: Evidence from the Micro-context." *American Sociological Review* 80 (3): 550–573.

Dixon, John, Kevin Durrheim, and Colin Tredoux. 2005. "Beyond the Optimal Contact Strategy: A Reality Check for the Contact Hypothesis." *American Psychologist* 60 (7): 697-711.

Enos, Ryan D. 2014. "Causal Effect of Intergroup Contact on Exclusionary Attitudes." *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704.

Enos, Ryan D. 2016. “What the Demolition of Public Housing Teaches us About the Impact of Racial Threat on Political Behavior.” *American Journal of Political Science* 60 (1): 123–142.

Enos, R. D., & Celaya, C. (2018). The effect of segregation on intergroup relations. Journal of Experimental Political Science, 5(1), 26-38.

Finseraas, Henning, and Andreas Kotsadam. 2017. “Does Personal Contact with Ethnic Minorities Affect Anti-immigrant Sentiments? Evidence from a Field Experiment.” *European Journal of Political Research* 56 (3): 703 – 722.

Finseraas, Henning, Åshild A. Johnsen, Andreas Kotsadam, and Gaute Torsvik. 2016. “Exposure to Female Colleagues Breaks the Glass Ceiling: Evidence from a Combined Vignette and Field Experiment.” *European Economic Review* 90: 363–374.

Finseraas, Henning, Torbjørn Hanson, Åshild A Johnsen, Andreas Kotsadam, and Gaute Torsvik. 2019. “Trust, Ethnic Diversity, and Personal Contact: A Field Experiment.” *Journal of Public Economics* 173: 72-84.

Gerber, Alan S., and Neil Malhotra. 2008a. “Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results?” *Sociological Methods & Research* 37 (1): 3–30.

Gerber, Alan, and Neil Malhotra. 2008b. "Do Statistical Reporting Standards Affect What is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science* 3 (3): 313-326.

Gonzalez, A. M., Dunlop, W. L., & Baron, A. S. (2017). Malleability of implicit associations across development. Developmental Science,20(6)

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2009. “Does Immigration Affect the Long-term Educational Outcomes of Natives? Quasi-experimental Evidence.” *The Economic Journal* 119 (540): 1243–1269.

Green, Donald P., and Janelle S. Wong. 2009. “Tolerance and the Contact Hypothesis: A Field Experiment,” Pp. 228-246 in Borgida, Eugene, Christopher M. Federico, and John L. Sullivan (eds.), *The Political Psychology of Democratic Citizenship*. Oxford University Press, Oxford, UK.

Grow, Andre, Karoly Takacs, and Judit Pal. 2016. “Status Characteristics and Ability Attributions in Hungarian School Classes: An Exponential Random Graph Approach.” *Social Psychology Quarterly* 79 (2): 156–167.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. “Validating Vignette and Conjoint Survey Experiments Against Real-world Behavior.” *Proceedings of the National Academy of Sciences* 112 (8): 2395–2400.

Hajdu, Tamás, Gábor Kertesi, and Gábor Kézdi. 2017. “Health Differences at Birth Between Roma and Non-Roma Children in Hungary-Long-Run Trends and Decompositions.” Technical Report, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences.

Hajdu, Tamás, Gábor Kertesi, and Gábor Kézdi. 2018. "Interethnic Friendship and Hostility Between Roma and non-Roma Students in Hungary: The role of exposure and academic achievement." *The BE Journal of Economic Analysis & Policy,* 19 (1): 1-17.

Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos, and Dimitrios Xefteris. 2019. “Does Exposure to the Refugee Crisis Make Natives More Hostile?” *American Political Science Review* 113 (2): 442–455.

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. 2015. “The Extent and Consequences of P-hacking in Science.” *PLoS Biology* 13 (3), e1002106.

Hoxby, Caroline. 2000. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” Technical Report, *National Bureau of Economic Research*.

Jakobsson, Niklas., Kotsadam, Andreas., Syse, Astrid., and Øien, Henning. 2016. “Gender Bias in Public Long-term Care? A Survey Experiment Among Care Managers”. *Journal of Economic Behavior & Organization* 131: 126-138.

Kende, Anna, Linda Tropp, and Nora Anna Lantos. 2017. “Testing a Contact Intervention Based on Intergroup Friendship Between Roma and non-Roma Hungarians: Reducing Bias Through Institutional Support in a Non-supportive Societal Context.” *Journal of Applied Social Psychology* 47 (1): 47–55.

Kertesi, Gábor, and Gábor Kézdi. 2011a. “Roma Employment in Hungary After the Post-communist Transition.” *Economics of Transition* 19 (3): 563–610.

Kertesi, Gábor, and Gábor Kézdi, 2011b. “The Roma/non-Roma Test Score Gap in Hungary.” *American Economic Review* 101 (3): 519–25.

Kim, Jinho. 2020. “The Effect of Classmates’ Maternal College Attainment on Volunteering in Young Adulthood.” *Social Science Quarterly* 101(6): 2289–2311.

Kisfalusi, Dorottya, Béla Janky, and Károly Takács. 2019. "Double Standards or Social Identity? The Role of Gender and Ethnicity in Ability Perceptions in the Classroom." *The Journal of Early Adolescence* 39 (5): 745-780.

Kisfalusi, Dorottya, and Judit Pál. 2020. "Bullying and Victimization among Majority and Minority Students: The Effects of Self-and Peer-Reported Ethnicity." *Social Networks* 60: 48-60.

Laurence, James. 2009. “The Effect of Ethnic Diversity and Community Disadvantage on Social Cohesion: A Multi-level Analysis of Social Capital and Interethnic Relations in UK Communities.” *European Sociological Review* 27 (1): 70–89.

Lee, Dohoon, and Byungkyu Lee. 2020. "The Role of Multilayered Peer Groups in Adolescent Depression: A Distributional Approach." *American Journal of Sociology* 125 (6): 1513-1558.

Legewie, Joscha, and Merlin Schaeffer. 2016. "Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict." *American Journal of Sociology* 122 (1),:125–161.

Lowe, Matt, 2020. "Types of Contact: A field Experiment on Collaborative and Adversarial Caste integration." *CESifo Working Paper*, Number 8089,

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic studies* 60 (3): 531–542.

Markowicz, Jocelyn A. 2009. *Intergroup Contact Experience in Dialogues on Race Groups: Does Empathy and an Informational Identity Style Help Explain Prejudice Reduction?* (Doctoral dissertation). The Pennsylvania State University, State College, PA.

McLaren, Lauren M. 2003. "Anti-immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants." *Social Forces* 81 (3): 909–936.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107 (3): 679–716.

Morauszki, András, and Attila Papp Z. 2015. "Ethnic Revival? The Methodology of the 2011 Census and the Nationalities of Hungary." Minority Studies: Demography, Minority Education, Ethnopolitics 18: 141-160.

Mousa, Salma. 2020. "Building Social Cohesion Between Christians and Muslims Through Soccer in Post-ISIS Iraq." *Science* 369 (6505): 866-870.

Ogburn, Elizabeth L., and Tyler J. VanderWeele. 2014. "Causal Diagrams for Interference." *Statistical Science* 29 (4): 559-578.

Page-Gould, Elizabeth, Rodolfo Mendoza-Denton, and Linda R. Tropp. 2008. "With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship." *Journal of Personality and Social Psychology* 95 (5): 1080-1094.

Paluck, Elizabeth Levy, and Donald P. Green. 2009. "Prejudice reduction: What Works? A Review and Assessment of Research and Practice." *Annual Review of Psychology* 60: 339–367.

Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. 2019. "The Contact Hypothesis Re-evaluated." *Behavioural Public Policy* 3 (2): 129–158.

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49 (1): 65–85.

Pettigrew, Thomas F., and Linda R. Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90 (5): 751-783.

Pettigrew, Thomas F., Linda R. Tropp, Ulrich Wagner, and Oliver Christ. 2011. "Recent Advances in Intergroup Contact Theory." *International Journal of Intercultural Relations* 35 (3): 271–280.

Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-first Century." *Scandinavian Political Studies* 30 (2): 137–174.

Rao, Gautam. 2019. "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools." *American Economic Review* 109 (3): 774-809.

Rohrer, Julia, Tamas Keller, and Felix Elwert. 2021. "Proximity Can Induce Diverse Friendships: A Large Randomized Classroom Experiment". *PLoS ONE 16(8): e0255097.* *https://doi.org/10.1371/journal.pone.0255097*

Scacco, Alexandra, and Shana S. Warren, 2018. “Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria.” *American Political Science Review* 112 (3): 654–677.

Segal, Mady W. 1974. "Alphabet and Attraction: An Unobtrusive Measure of the Effect of Propinquity in a Field Setting." *Journal of Personality and Social Psychology* 30(5): 6454-57.

Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2006. “The Rise of Anti-foreigner Sentiment in European Societies, 1988-2000.” *American Sociological Review* 71 (3): 426–449.

Shalizi, Cosma R., and Thomas, Andrew C. 2011. “Homophily and Contagion are Generically Confounded in Observational Social Network Studies.” *Sociological Methods and Research* 40: 211–239.

Simonovits, Gábor, and Gábor Kézdi. 2016. “Economic Hardship Triggers Identification with Disadvantaged Minorities.” *The Journal of Politics* 78 (3): 882–892.

Simonovits, Gábor, Gábor Kézdi, and Peter Kardos. 2018. “Seeing the World Through the Other’s Eye: An Online Intervention Reducing Ethnic Prejudice.” *American Political Science Review* 112 (1): 186–193.

Smith, Tom W. (2002). “Measuring inter-racial friendships.” *Social Science Research* 31(4): 576-593.

Sobel, Michael E. 2006. "What do Randomized Studies of Housing Mobility Demonstrate? Causal inference in the face of interference." *Journal of the American Statistical Association* 101 (476): 1398-1407.

Sorensen, Nicholas A. 2010. *The Road to Empathy: Dialogic Pathways for Engaging Diversity and Improving Intergroup Relations*, (Doctoral thesis). University of Michigan, Ann Arbor, MI.

Stolle, Dietlind, Stuart Soroka, and Richard Johnston. 2008. “When does Diversity Erode Trust? Neighborhood Diversity, Interpersonal Trust and the Mediating Effect of Social Interactions.” *Political Studies* 56 (1): 57–75.

Valdez, Sarah. 2014. “Visibility and Votes: A Spatial Analysis of Anti-immigrant Voting in Sweden.” *Migration Studies* 2 (2): 162–188.

van der Meer, Tom, and Jochem Tolsma. 2014. “Ethnic Diversity and its Effects on Social Cohesion.” *Annual Review of Sociology* 40: 459–478.

Van Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. “The Effect of University Roommate Contact on Ethnic Attitudes and Behavior.” *Journal of Experimental Social Psychology* 41 (4):329–345.

VanderWeele, Tyler J., and Weihua An. 2013. “Social Networks and Causal Inference.” Pp. 353-374 in Morgan, Stephen L. (ed.), *Handbook of Causal Analysis for Social Research*. Springer: Dodrecht.

Williams, Robin Murphy. 1964. *Strangers Next Door: Ethnic Relations in American Communities*, Englewood Cliffs, NJ: Prentice-Hall.

Table 1: Descriptive statistics

	Total		Treated Roma deskmate		Control Non-Roma deskmate	
Main dependent variables						
Roma friend	0.30	(0.46)	0.52	(0.50)	0.27	(0.44)
Lend to classmate	0.57	(0.49)	0.57	(0.50)	0.57	(0.49)
Main exposure variable						
Roma deskmate	0.12	(0.33)	1.00	(0.00)	0.00	(0.00)
Control variables						
Age (in years)	11.83	(1.82)	11.97	(1.81)	11.81	(1.82)
Girl	0.48	(0.50)	0.43	(0.50)	0.48	(0.50)
Grade mathematics	3.83	(1.01)	3.68	(1.00)	3.85	(1.01)
Grade grammar	3.96	(1.00)	3.79	(1.11)	3.98	(0.98)
Grade literature	3.83	(1.08)	3.67	(1.13)	3.85	(1.07)
Grade diligence	4.15	(0.88)	4.03	(0.92)	4.17	(0.88)
Grade behavior	4.41	(0.76)	4.34	(0.82)	4.42	(0.75)
Variables for heterogeneity analyses						
GPA	3.87	(0.94)	3.71	(0.99)	3.89	(0.93)
Deskmate of same sex	0.50	(0.50)	0.46	(0.50)	0.50	(0.50)
Share of Roma in class	0.12	(0.17)	0.38	(0.24)	0.08	(0.13)
N	2395		291		2104	

Notes: All samples consist of non-Roma students that have a non-missing deskmate. Column 2 is restricted to students that have a Roma deskmate. Column 3 is restricted to students that do not have a Roma deskmate.

Table 2: Balance test

	(1) Roma deskmate (RD)	(2) RD	(3) RD	(4) RD	(5) RD	(6) RD	(7) RD	(8) RD	(9) Roma vignette
Age	0.0012							0.00029	-0.00057
	(0.0011)							(0.0012)	(0.0024)
Girl		-0.0012						-0.00054	-0.012
		(0.012)						(0.013)	(0.025)
Mathematics		-0.0071						-0.00056	0.0070
			(0.0065)					(0.011)	(0.021)
Grammar				-0.0082				-0.0052	0.0019
				(0.0066)				(0.012)	(0.022)
Literature				-0.0041				0.0066	-0.023
					(0.0063)			(0.010)	(0.020)
Diligence						-0.010		-0.019	-0.018
						(0.0075)		(0.014)	(0.028)
Behavior							-.00051	0.0093	0.036*
							(0.008)	(0.011)	(0.022)
Mean outcome for controls con	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.50
N	2360	2395	2274	2274	2277	2279	2279	2238	2101
R-squared	0.36	0.37	0.35	0.35	0.35	0.35	0.35	0.35	0.08
Classroom fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions control for classroom fixed effects. The F-statistic for whether the control variables jointly predict deskmate status in column 8 is 1.09 (p-value = 0.245). The F-statistic for whether the control variables jointly predict receipt of the Roma vignette in column 9 is 0.87 (p-value = 0.825). Heteroscedasticity robust standard errors in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 3: Effects of Roma deskmate on lending to a classmate and on having a Roma friend among one's best friends

	1	2	3	4
	Lend to classmate	Lend to classmate	Roma friend	Roma friend
Roma deskmate	-0.030	-0.027	-0.034	-0.036
	(0.053)	(0.056)	(0.036)	(0.037)
Roma vignette	-0.18***	-0.17***		
	(0.023)	(0.023)		
Roma deskmate * Roma vignette	-0.034	-0.052		
	(0.067)	(0.070)		
Mean outcome for controls	0.67	0.67	0.27	0.27
No. of observations	2102	2102	2124	2124
R-squared	0.22	0.26	0.34	0.38
Classroom fixed effects	Yes	Yes	Yes	Yes
Baseline Controls	No	Yes	No	Yes

Notes: All regressions control for classroom fixed effects. The baseline controls in columns 2 and 4 are indicator variables for gender, age in 0.1-year brackets, and baseline grades in mathematics, grammar, literature, diligence, and behavior (all with one dummy for each level of the grade, which range from 1-5). Heteroscedasticity robust standard errors in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 4: Effects of grade-level exposure to Roma students on lending to a classmate and on having a Roma friend among one's best friends

	1	2	3	4
	Lend	Lend	Friend	Friend
Share Roma in grade	0.057	0.054	0.13***	0.12***
	(0.035)	(0.034)	(0.035)	(0.032)
Roma vignette (RV)	-0.19***	-0.19***		
	(0.023)	(0.024)		
RV * Share Roma	0.023	0.015		
	(0.020)	(0.020)		
Mean outcome at average share	0.57	0.57	0.30	0.30
No. of observations	2102	2102	2124	2124
R-squared	0.08	0.13	0.20	0.27
School fixed effects	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes

Robust standard errors clustered at the grade level are presented in parentheses.
*, **, *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

ONLINE APPENDIX

A Tables and figures discussed in the text

Table A.1: Effects of Roma deskmate when restricting the sample to classrooms with over 90 percent compliance and correlations between compliance and exposure.

	(1)	(2)	(3)	(4)
	High compliance sample		Full analytic sample	
	Roma friend	Lend to Classmate	Low compliance	Compliance share
Roma Deskmate	0.0070	-0.10		
	(0.049)	(0.074)		
Roma vignette		-0.15***		
		(0.031)		
Roma Deskmate*Roma vignette		-0.0034		
		(0.088)		
Share of Roma in class			0.089	-0.035
			(0.18)	(0.072)
Mean dep. var in C group	0.23	0.66	0.48	0.85
No. of observations	1120	1117	2395	2395
R-squared	0.33	0.22	0.00	0.00
Class F.E.	Yes	Yes	No	No

Notes: The sample in columns 1 and 2 are restricted to classrooms with over 90 percent compliance. The sample in columns 3 and 4 is the pre-registered sample we use throughout the paper. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.2: Effects of Roma deskmate in regressions with controls without including indicator variables for missing values

	(1)	(2)
	Roma friend	Lend to classmate
Roma deskmate	-0.019	-0.016
	(0.037)	(0.057)
Roma vignette		-0.17***
		(0.024)
Roma deskmate * Roma vignette		-0.043
		(0.071)
Mean outcome for controls	0.26	0.67
No. of observations	2047	2021
R-squared	0.38	0.26
Class fixed effects	Yes	Yes
Baseline Controls	No	Yes
Baseline*Treatment	No	No

Notes: All regressions control for classroom fixed effects. The baseline controls are indicator variables for gender, age in 0.1 year brackets, baseline grades in mathematics, grammar, literature, diligence, and behavior (all with one dummy for each level of the grade, which range from 1-5). We do not include missing values as indicator variables in these regressions. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.3: Heterogeneous effects of Roma deskmates on Roma friend based on students' own baseline characteristics

	(1) Friend	(2) Friend	(3) Friend	(4) Friend	(5) Friend	(6) Friend	(7) Friend	(8) Friend
Roma deskmate (RD)	-0.035	-0.031	-0.024	-0.027	-0.021	-0.016	-0.014	-0.025
	(0.036)	(0.048)	(0.037)	(0.037)	(0.037)	(0.037)	(0.036)	(0.037)
Roma deskmate*Age	0.052							
	(0.033)							
Roma deskmate*Girl		-0.0062						
		(0.061)						
Roma deskmate*Mathematics			0.013					
			(0.034)					
Roma deskmate*Grammar				0.0052				
				(0.029)				
Roma deskmate*Literature					0.015			
					(0.032)			
Roma deskmate*Diligence						0.042		
						(0.032)		
Roma deskmate*Behavior							0.0097	
							(0.033)	
Roma deskmate*GPA								0.013
								(0.031)
Age	-0.096***							
	(0.035)							
Girl		-0.046**						
		(0.019)						
Mathematics			-0.057***					
			(0.011)					
Grammar				-0.055***				
				(0.011)				
Literature					-0.049***			
					(0.011)			
Diligence						-0.066***		
						(0.011)		
Behavior							-0.052***	
							(0.011)	
GPA								-0.060***
								(0.011)
Mean outcome for controls	0.27	0.27	0.26	0.26	0.26	0.26	0.26	0.26
No. of observations	2124	2124	2046	2047	2048	2049	2049	2048
R-squared	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions control for classroom fixed effects. All continuous variables are standardized to have mean zero and standard deviation 1 for ease of interpretation. Girl is kept as a dummy variable. Robust standard errors are presented in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

Table A.4: Heterogeneous effects of Roma deskmates on Lend to classmate based on students' own baseline characteristics

	(1) Lend	(2) Lend	(3) Lend	(4) Lend	(5) Lend	(6) Lend	(7) Lend	(8) Lend
Roma deskmate (RD)	-0.027	0.027	-0.032	-0.031	-0.030	-0.029	-0.031	-0.029
	(0.053)	(0.070)	(0.053)	(0.054)	(0.053)	(0.054)	(0.054)	(0.054)
Roma deskmate*Age	0.088*							
	(0.052)							
Roma deskmate*Girl		-0.13						
		(0.095)						
Roma deskmate*Mathematics			-0.026					
			(0.047)					
Roma deskmate*Grammar				-0.0022				
				(0.044)				
Roma deskmate*Literature					0.070			
					(0.046)			
Roma deskmate*Diligence						-0.025		
						(0.046)		
Roma deskmate*Behavior							-0.068	
							(0.052)	
Roma deskmate*GPA								0.018
								(0.046)
Roma vignette (RV)	-0.18***	-0.20***	-0.18***	-0.18***	-0.18***	-0.18***	-0.18***	-0.18***
	(0.023)	(0.031)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Roma deskmate * Roma vignette	-0.032	-0.071	-0.027	-0.026	-0.024	-0.017	-0.013	-0.027
	(0.067)	(0.089)	(0.069)	(0.068)	(0.069)	(0.069)	(0.069)	(0.069)
RV*Age	0.073***							
	(0.023)							
RV*Girl		0.042						
		(0.044)						
RV*Mathematics			0.025					
			(0.023)					
RV*Grammar				0.011				
				(0.024)				
RV*Literature					0.041*			
					(0.023)			
RV*Diligence						0.016		
						(0.024)		
RV*Behavior							0.00095	
							(0.024)	
RV*GPA								0.030
								(0.024)
RD*RV*Age	-0.078							
	(0.069)							
RD*RV*Girl		0.093						
		(0.13)						
RD*RV*Mathematics			-0.049					
			(0.065)					
RD*RV*Grammar				-0.059				
				(0.059)				
RD*RV*Literature					-0.12**			
					(0.061)			
RD*RV*Diligence						0.0067		
						(0.063)		
RD*RV*Behavior							0.072	
							(0.067)	
RD*RV*GPA								-0.086
								(0.062)
Age	-0.093**							
	(0.038)							
Girl		-0.0050						
		(0.031)						
Mathematics			-0.00048					
			(0.017)					
Grammar				0.014				
				(0.017)				
Literature					-0.015			
					(0.017)			
Diligence						0.0082		
						(0.017)		
Behavior							0.0000035	
							(0.017)	
GPA								-0.00082
								(0.017)
Mean outcome for controls	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
No. of observations	2102	2102	2020	2022	2024	2025	2025	2024
R-squared	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions control for classroom fixed effects. All continuous variables are standardized to have mean zero and standard deviation 1 for ease of interpretation. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.5: Heterogeneous effects of Roma deskmates on Roma friend based on deskmates' baseline characteristics

	(1) Friend	(2) Friend	(3) Friend	(4) Friend	(5) Friend	(6) Friend	(7) Friend	(8) Friend
Roma deskmate (RD)	-0.035	-0.043	-0.028	-0.052	-0.014	0.00062	-0.014	-0.030
	(0.036)	(0.047)	(0.050)	(0.048)	(0.047)	(0.048)	(0.045)	(0.051)
Roma deskmate*Age	-0.051							
	(0.034)							
Roma deskmate*Girl		0.019						
		(0.059)						
Roma deskmate*Mathematics			0.0088					
			(0.036)					
Roma deskmate*Grammar				-0.028				
				(0.035)				
Roma deskmate*Literature					0.022			
					(0.034)			
Roma deskmate*Diligence						0.043		
						(0.033)		
Roma deskmate*Behavior							0.028	
							(0.031)	
Roma deskmate*GPA								-0.0025
								(0.037)
Age	0.095***							
	(0.034)							
Girl		-0.017						
		(0.018)						
Mathematics			0.0071					
			(0.011)					
Grammar				0.0071				
				(0.011)				
Literature					-0.0010			
					(0.012)			
Diligence						-0.0045		
						(0.011)		
Behavior							-0.0016	
							(0.011)	
GPA								0.0057
								(0.012)
Mean outcome for controls	0.27	0.27	0.26	0.26	0.26	0.26	0.26	0.26
No. of observations	2124	2124	2015	2015	2019	2020	2020	2020
R-squared	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions control for classroom fixed effects. All continuous variables are standardized to have mean zero and standard deviation 1 for ease of interpretation. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.6: Heterogeneous effects of Roma deskmates on Lend to classmate based on deskmates' baseline characteristics

	(1) Lend	(2) Lend	(3) Lend	(4) Lend	(5) Lend	(6) Lend	(7) Lend	(8) Lend
Roma deskmate (RD)	-0.027	-0.099	0.0065	-0.015	0.016	0.0096	-0.0082	0.016
	(0.053)	(0.074)	(0.064)	(0.069)	(0.073)	(0.068)	(0.064)	(0.071)
Roma deskmate*Age	-0.087*							
	(0.052)							
Roma deskmate*Girl		0.14						
		(0.093)						
Roma deskmate*Mathematics			0.021					
			(0.057)					
Roma deskmate*Grammar				0.0013				
				(0.055)				
Roma deskmate*Literature					0.059			
					(0.054)			
Roma deskmate*Diligence						0.030		
						(0.051)		
Roma deskmate*Behavior							0.0086	
							(0.046)	
Roma deskmate*GPA								0.038
								(0.056)
Roma vignette (RV)	-0.18***	-0.17***	-0.18***	-0.18***	-0.18***	-0.18***	-0.18***	-0.18***
	(0.023)	(0.031)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Roma deskmate * Roma vignette	-0.032	-0.011	0.029	0.051	0.016	0.035	-0.0056	0.041
	(0.067)	(0.093)	(0.091)	(0.096)	(0.093)	(0.088)	(0.082)	(0.098)
RV*Age	-0.073***							
	(0.023)							
RV*Girl		-0.011						
		(0.045)						
RV*Mathematics			-0.017					
			(0.024)					
RV*Grammar				-0.035				
				(0.024)				
RV*Literature					0.017			
					(0.024)			
RV*Diligence						-0.0070		
						(0.025)		
RV*Behavior							-0.031	
							(0.025)	
RV*GPA								-0.013
								(0.025)
RD*RV*Age	0.077							
	(0.069)							
RD*RV*Girl		-0.045						
		(0.13)						
RD*RV*Mathematics			0.044					
			(0.077)					
RD*RV*Grammar				0.084				
				(0.073)				
RD*RV*Literature					0.012			
					(0.070)			
RD*RV*Diligence						0.058		
						(0.067)		
RD*RV*Behavior							0.037	
							(0.062)	
RD*RV*GPA								0.055
								(0.076)
Age	0.092**							
	(0.038)							
Girl		-0.023						
		(0.031)						
Mathematics			0.019					
			(0.017)					
Grammar				0.017				
				(0.017)				
Literature					-0.014			
					(0.017)			
Diligence						0.0087		
						(0.017)		
Behavior							0.014	
							(0.017)	
GPA								0.0073
								(0.018)
Mean outcome for controls	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
No. of observations	2102	2102	1990	1990	1994	1996	1996	1995
R-squared	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions control for classroom fixed effects. All continuous variables are standardized to have mean zero and standard deviation 1 for ease of interpretation. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.7: Effects of Roma deskmate depending on sex concordance of deskmate

	(1) Roma friend	(2) Roma friend	(3) Lend to classmate	(4) Lend to classmate
Roma deskmate	-0.043	-0.063	-0.24***	0.100
	(0.068)	(0.049)	(0.087)	(0.076)
Roma vignette			-0.19***	-0.19***
			(0.033)	(0.034)
Roma deskmate * Roma vignette			0.14	-0.18*
			(0.11)	(0.094)
Mean outcome for controls	0.27	0.26	0.68	0.66
No. of observations	1056	1068	1045	1057
R-squared	0.40	0.40	0.30	0.28
Class fixed effects	Yes	Yes	Yes	Yes
Deskmate sex	Same sex	Different sex	Same sex	Different sex

Notes: All regressions control for classroom fixed effects. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.8: Effects of Roma deskmate on lending to Roma interacted with lending to deskmate

	(1) Lend to Roma
Roma deskmate	-0.12
	(0.079)
Lend to Deskmate	0.31***
	(0.045)
Roma deskmate*Lend to Deskmate	0.087
	(0.096)
Mean outcome for controls	0.48
No. of observations	978
R-squared	0.33
Class fixed effects	Yes

Notes: All regressions control for classroom fixed effects. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.9: Effects of Roma deskmate and different grade level exposure

	(1) Lend	(2) Lend	(3) Lend	(4) Friend	(5) Friend	(6) Friend
Roma deskmate (RD)	0.024	0.0096	0.0058	0.027	-0.0044	-0.012
	(0.070)	(0.070)	(0.069)	(0.046)	(0.045)	(0.045)
Share Roma in grade	0.014	0.061	0.046	0.21***	0.14***	0.12***
	(0.025)	(0.040)	(0.035)	(0.019)	(0.037)	(0.033)
Roma deskmate*share Roma	-0.037	-0.0050	0.0078	-0.070**	-0.026	-0.014
	(0.043)	(0.042)	(0.041)	(0.031)	(0.030)	(0.030)
Roma vignette (RV)	-0.19***	-0.19***	-0.18***			
	(0.025)	(0.023)	(0.024)			
Roma deskmate * Roma vignette	-0.14	-0.13	-0.14			
	(0.096)	(0.090)	(0.090)			
RV*Share Roma	0.022	0.019	0.022			
	(0.028)	(0.026)	(0.026)			
RD*RV*Share Roma	0.094*	0.058	0.051			
	(0.054)	(0.052)	(0.051)			
Average grade GPA			-0.12***			-0.18***
			(0.044)			(0.031)
Mean outcome for controls	0.26	0.26	0.26	0.27	0.27	0.27
No. of observations	2102	2102	2102	2124	2124	2124
R-squared	0.04	0.08	0.09	0.15	0.21	0.22
School F.E.	No	Yes	Yes	No	Yes	Yes
Average grade GPA	No	No	Yes	No	No	Yes

Notes: Robust standard errors clustered at the grade level are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.10: Effects of Roma deskmate in majority Roma and non Roma classrooms.

	(1) Roma friend	(2) Roma friend	(3) Lend to classmate	(4) Lend to classmate
Roma deskmate	-0.093	-0.028	0.036	-0.00079
	(0.11)	(0.039)	(0.15)	(0.057)
Roma vignette			0.091	-0.18***
			(0.15)	(0.023)
Roma deskmate * Roma vignette			-0.030	-0.11
			(0.21)	(0.074)
Mean outcome for controls	0.86	0.25	0.55	0.67
No. of observations	103	2021	113	1989
R-squared	0.39	0.29	0.44	0.21
Class fixed effects	Yes	Yes	Yes	Yes
Sample	Majority Roma	Majority non-Roma	Majority Roma	Majority non-Roma

Notes: All regressions control for classroom fixed effects. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.11: Classroom exposure

	(1) Lend	(2) Lend	(3) Friend	(4) Friend
Share Roma in class	0.035	0.038	0.12***	0.13***
	(0.029)	(0.030)	(0.028)	(0.028)
Roma vignette (RV)	-0.19***	-0.19***		
	(0.023)	(0.025)		
RV*Share Roma	0.023	0.014		
	(0.019)	(0.020)		
Mean dep. var at average share	0.57	0.57	0.30	0.30
No. of observations	2102	2102	2124	2124
R-squared	0.09	0.13	0.23	0.28
School F.E.	Yes	Yes	Yes	Yes
Average class GPA	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes

Notes: Robust standard errors clustered at the class level are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.12: Interactions between deskmate intervention and the share of Roma students in the class

	(1) Lend	(2) Lend	(3) Friend	(4) Friend
Roma Deskmate (RD)	0.011	-0.0021	-0.041	-0.052
	(0.084)	(0.078)	(0.050)	(0.051)
Roma vignette (RV)	-0.17***	-0.17***		
	(0.028)	(0.026)		
Roma Deskmate*Roma vignette	-0.15	-0.13		
	(0.11)	(0.098)		
Roma Deskmate*share Roma	-0.0076	0.0069	0.0097	0.022
	(0.058)	(0.053)	(0.037)	(0.038)
Roma Vignette*Share Roma	0.020	0.029		
	(0.034)	(0.033)		
RD*RV*Share Roma	0.062	0.044		
	(0.065)	(0.060)		
Mean dep. var in C group	0.57	0.57	0.27	0.27
No. of observations	2102	2102	2124	2124
R-squared	0.26	0.22	0.34	0.38
Class F.E.	Yes	Yes	Yes	Yes
Baseline Controls	No	Yes	No	Yes

Notes: Robust standard errors clustered at the class level are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.13: School-level Correlates of Deskmate Cooperation Activities

VARIABLES	(1) work together	(2) correct each other's work	(3) explain exercise/course material	(4) discuss issues/problems	(5) help each other to learn	(6) play out situations	(7) do homework together	(8) correct each other's homework	(9) develop social skills
Type of settlement (ref. Budapest)	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Village	-0.088	-0.056	0.142	-0.059	0.189	0.118	-0.188	-0.270	-0.233
	(0.210)	(0.247)	(0.125)	(0.201)	(0.142)	(0.182)	(0.246)	(0.232)	(0.196)
Town	0.071	0.157	0.369***	-0.041	0.304**	0.357**	0.137	-0.035	-0.062
	(0.212)	(0.230)	(0.135)	(0.205)	(0.154)	(0.179)	(0.258)	(0.236)	(0.198)
County seat	0.003	0.007	0.284	-0.206	0.118	0.008	-0.020	0.062	-0.346*
	(0.217)	(0.241)	(0.185)	(0.218)	(0.188)	(0.206)	(0.263)	(0.248)	(0.195)
Share of Roma	0.001	0.001	-0.003	-0.000	-0.002	0.000	-0.005+	-0.005	-0.000
	(0.002)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Math score	-0.002**	0.000	-0.001	-0.001	0.000	-0.000	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Reading score	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
School size (ref: small school)	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Middle sized school	-0.082	0.115	-0.135	-0.052	-0.133	-0.023	-0.048	0.060	-0.076
	(0.092)	(0.147)	(0.116)	(0.115)	(0.104)	(0.111)	(0.147)	(0.148)	(0.108)
Large sized school	-0.134	-0.109	-0.186	0.104	0.011	-0.140	-0.017	-0.322*	0.010
	(0.113)	(0.171)	(0.140)	(0.125)	(0.131)	(0.134)	(0.183)	(0.184)	(0.136)
Constant	2.826***	1.509	2.449*	2.534**	1.460	2.122**	2.477*	2.583*	2.527*
	(0.955)	(1.511)	(1.258)	(1.258)	(1.027)	(1.056)	(1.263)	(1.363)	(0.997)
Observations	584	584	584	584	584	584	584	584	584
R-squared	**0.017**	**0.012**	**0.021**	**0.009**	**0.014**	**0.018**	**0.039**	**0.032**	**0.013**

Notes: Deskmate collaboration data from a national survey of homeroom teachers in Hungarian primary schools; collected February 2022, N=656. Each activity measured on a Likert scale, ranging from 1="almost every lesson to 5="never". School-level covariates from the Hungarian National Assessment of Basic Competencies (NABC) dataset, 2018. For 72 responses, complete school-level covariates were unavailable, resulting in N=584. Standard errors clustered at the school level in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

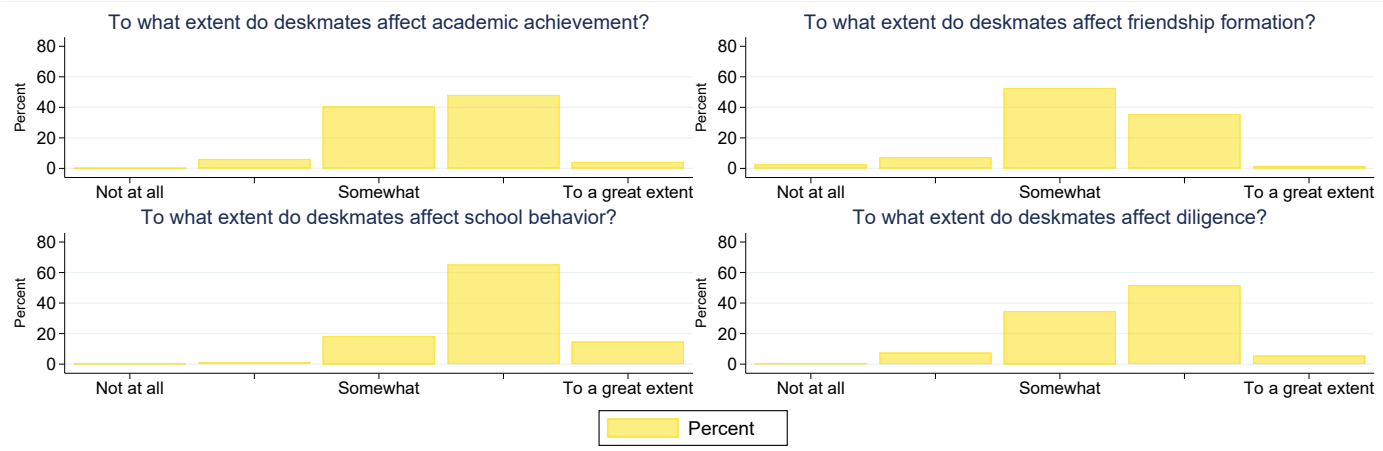

Figure A.1: Teacher perceptions of deskmate influence

Notes: Teacher perceptions from a survey among 7th and 8th-grade homeroom teachers in all Hungarian primary schools in 2021 (N=413). Responding schools were nationally representative.

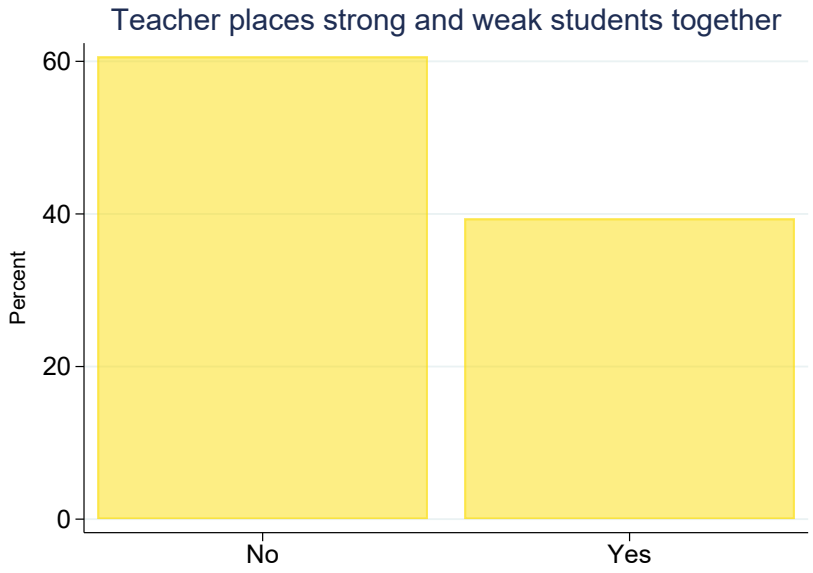

Figure A.2: Deskmate allocation

Notes: Teacher perceptions from a survey among 7th and 8th-grade homeroom teachers in all Hungarian primary schools in 2021 (N=413). Responding schools were nationally representative.

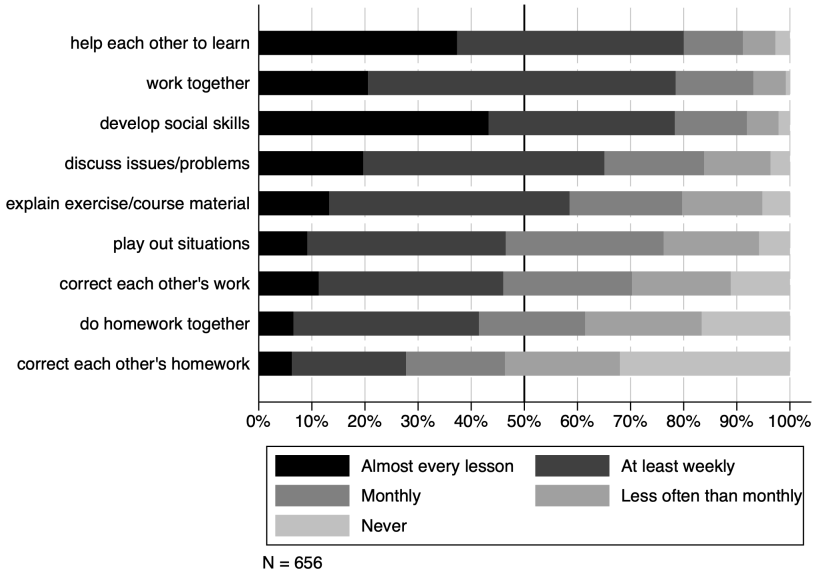

Figure A.3: Deskmate collaboration

Notes: Deskmate collaboration data from a national survey of homeroom teachers in all Hungarian primary schools; collected February 2022, N=656. Each activity measured on a Likert scale, ranging from 1=”almost every lesson to 5=”never”.

B Additional details on the data collection and experimental design

We recruited schools from 7 contiguous counties of central Hungary (excluding the capital city Budapest). In the spring of 2017, we contacted all primary schools in these counties via the heads of the local school districts to elicit information about room layouts and seating practices. By the end of the summer vacation, we had obtained initial participation agreements with 55 schools in which most 3rd-8th grade classrooms were anticipated to meet a set of inclusion criteria. These criteria were: 1) Principals and teachers would implement our randomized seating chart in three subjects: Hungarian literature, Hungarian grammar, and mathematics. 2) All students in a classroom would receive instruction in these subjects together (e.g., no ability grouping). 3) Classroom layout would comprise free-standing desks that seat two students. In these classrooms, the students were expected to sit next to their deskmates for around 20 hours each week.

The intervention assigned students to free-standing two-person desks via unconstrained random partitioning within each classroom. We based the randomization on the class rosters from the preceeding spring semester. Shortly before the start of the fall semester, we submitted the randomized seating charts to teachers for implementation. Teachers were instructed to use the charts for the duration of the fall semester, until January 2018. To account for changes to class rosters during the summer via exits and entries, we instructed teachers to fill seats vacated by exiting students with entering students from left to right, front to back, in alphabetic order of entering students' surnames. Since, (i) in expectation, students enter and exit classrooms for the same reason (repeating grades and residential moves); and (ii) student surnames are reasonably orthogonal to student grades, this replacement rule preserves randomization. We permitted teachers to reseat students where necessary, but we

asked teachers to preserve the desk-mate composition by moving both deskmates of a desk together whenever possible. We measured compliance through teacher reports of the actual seating chart for September 15, 2017. The field team again recorded the actual seating chart during school visits between October and December 2017 and verified classroom layouts through classroom photographs.

Schools and classrooms that did not meet our pre-registered inclusion criteria were dropped from the study. We excluded 133 entire classrooms for the following reasons: Withdrawal from the study (25); Less than 10 students at baseline (8); split (e.g. ability-grouped) classrooms (10); desk layouts other than free-standing two-person desks (30); Unreliable baseline reporting (7); Failure to implement the seating chart (40); Failure to report student ethnicity (13). We also pre-registered to exclude students who were randomly assigned to sit alone at a desk at baseline or who had missing outcomes.

C Additional results on secondary outcomes

In supplementary analyses, we investigate the effect of sitting next to a Roma student on non-Roma students number of Roma friends inside or outside the classroom [Question 5d]; whether students count at least one Roma student among their 5 best friends in the classroom [using name generators, Question 4]; whether the deskmate is among the student's 5 best friends [using name generators, Question 4].

Results show that sitting next to a Roma student has no detectable effect (near-zero point estimates, not statistically significant) on the number of Roma friends or having a Roma best friend in the classroom (columns 1 and 2 of Table A.14).

In the last column, the fact that 31 percent of non-Roma students count their non-Roma deskmate among their best friends testifies to the positive effect of propinquity on friendship formation, since the average class size is 20 students, so that only 25 percent of students

should count their deskmate among their best friends if friendship was unrelated to propinquity. The probability of counting a deskmate among one's 5 best friends is only 0.31-0.13=18 percent if the deskmate is a Roma student, showing no positive effect of propinquity on interethnic friendship. Rohrer et al. (2021) provide a full dyadic analysis of within- and between-group friendships in the present experiment.

The finding that non-Roma students are less likely to befriend their deskmate when they are assigned a Roma rather than a non-Roma deskmate reveals ethnic antipathy on the part of the non-Roma students. This finding does not, however, mean that being seated next to a Roma deskmate has caused this ethnic antipathy in the first place. Hence, it does not imply that seating non-Roma students next to Roma students makes non-Roma students less favorably inclined to Roma students in general, or even toward the particular Roma student who is their deskmate. This reading is consistent with our main specification, which does test whether sitting next to a Roma student increases (or decreases) discrimination, and finds no evidence that it does.

Table A.15 shows descriptive results for the association between sitting next to a Roma student and (a) lending to a deskmate, and (b) liking to sit next to one's deskmate. These results are intended to capture antipathy and distrust towards Roma deskmates. If the deskmate was Roma, non-Roma students were less likely to espouse a willingness to lend to their deskmate and were less likely to like sitting next to their deskmate. Note that these results are not suited to detect any effects of exposure to a Roma deskmate on students' attitudes toward Roma individuals in general, as all treated individuals have a Roma deskmate and all non-treated individuals have a non-Roma deskmate. The results show, however, that there is substantial antipathy and distrust towards Roma students, even among those exposed to Roma deskmates.

Table A.14: Effects of Roma deskmate on other outcomes

	(1) Number of Roma friends	(2) Roma among best friends	(3) Deskmate among best friends
Roma deskmate	-0.042	0.0063	-0.13***
	(0.19)	(0.033)	(0.036)
Mean outcome for controls	0.71	0.17	0.31
No. of observations	2124	2395	2170
R-squared	0.37	0.49	0.14
Class fixed effects	Yes	Yes	Yes

Notes: The friendship measure in column 1 includes friends both inside and outside of the classroom (question 5d), whereas the measure in columns 2 and 3 refers to friends within the classroom (from a name generator, question 4). All regressions control for classroom fixed effects. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table A.15: Effects of Roma deskmate on other outcomes

	(1) Lend to deskmate	(2) Liked sitting next to deskmate
Roma deskmate	-0.073**	-0.085**
	(0.034)	(0.041)
Mean outcome for controls	0.84	0.59
No. of observations	1989	2028
R-squared	0.13	0.16
Class fixed effects	Yes	Yes

Notes: All regressions control for classroom fixed effects. Robust standard errors are presented in parentheses. *, **, *** denote statistical significance at the 10%, 5% and 1% level, respectively.

D Pre-Analysis Plan and Survey Instrument for Outcomes (in English Translation)

The pre-analysis plan was archived at https://www.socialscienceregistry.org/ on March 22 2018, prior to the receipt of outcome data.

"Close contact, Trust, and Interethnic Friendship - A large scale field experiment of Roma desk mates in Hungarian schools." (c) By Felix Elwert, Tamas Keller, and Andreas Kotsadam. AEA RCT Registry. March 22 2018. https://doi.org/10.1257/rct.2795 All rights reserved. Used with permission.

Pre-analysis plan for: Close contact, Trust, and Interethnic Friendship - A large scale field experiment of Roma desk mates in Hungarian schools*

Felix Elwert[1], Tamás Keller[2], and Andreas Kotsadam[3]

[1]University of Wisconsin-Madison
[2]Research Center for Educational and Network Studies, Hungarian Academy of Sciences, Center for Social Sciences and TÁRKI Social Research Institute
[3]Ragnar Frisch Centre for Economic Research

Abstract

We are conducting a large scale field experiment where we randomly assign desk mates in Hungarian schools. By comparing people with Roma desk mates to people with non-Roma desk mates we investigate whether close personal contact to Roma increases inter-ethnic friendship and trust. In this plan we pre-register some key decisions to follow once we receive the data.

*The research has been funded by Tamás Keller's grant from the National Research, Development and Innovation Office (NKFIH), Grant number: FK 125358 and Vilas Associate Award from the University of Wisconsin-Madison.

1 Introduction

The Roma is one of the largest and poorest ethnic minority groups in Europe. In Hungary, the Roma population is estimated to constitute around 6 percent of the total population and 10 to 12 percent of the young adolescent population (Kertesi and Kézdi, 2011b). The Roma lag behind the general population in terms of health (Hajdu, Kertesi, and Kezdi, 2017a), education (Kertesi and Kézdi, 2011b), and employment (Kertesi and Kézdi, 2011a) and prejuice against the Roma is widespread (Hajdu, Kertesi, and Kezdi, 2017b). Using a large scale field experiment where we randomly assign desk mates in Hungarian schools we investigate whether close personal contact to Roma increases inter-ethnic friendship and trust.

Whether exposure reduces prejudice is an important question and previous evidence is mixed. Several empirical studies find patterns that shallow exposure is correlated with more prejudice and less trust (Alesina and La Ferrara, 2002; Delhey and Newton, 2005; Dinesen and Sønderskov, 2015; Stolle, Soroka, and Johnston, 2008). Putnam, (2007) has even proposed a constrict theory, arguing that ethnic diversity may not only lead to less trust between the majority and minority groups, it may also undermine trust within the majority group. A major limitation of these studies is the inability to control for selection issues and reverse causality (Baldassarri and Abascal, 2017). There are studies of close personal contact arguing that contact under some conditions reduce prejudice and increases trust (Allport, 1954) and well identified studies using random assignment of peers have found such effects (Boisjoly et al., 2006; Burns, Corno, and La Ferrara, 2016; Carrell, Hoekstra, and West, 2015; Finseraas and Kotsadam, 2017a; Finseraas et al., 2016). Kende, Tropp, and Lantos, (2017) randomly assign 61 non-Roma Hungarians to face-to-face interaction with a Roma person and found reduced prejudice for those exposed. The key condition for exposure to reduce prejudice has been argued to be friendship potential (Laurence, 2009; Pettigrew,

1998; Stolle, Soroka, and Johnston, 2008).

Similar people are more likely to form social ties (McPherson, Smith-Lovin, and Cook, 2001). This phenomenon, often described as *social homophily*, is consistent with a general preference for similarity and has been documented within several fields of science (Byrne, 1961, 1971). The tendency of lower probability of friendships across ethnic groups, inbreeding homophily, has been widely documented (see e.g. Jackson, (2014) and McPherson, Smith-Lovin, and Cook, (2001), also in Hungary with respect to Roma (Hajdu, Kertesi, and Kezdi, 2017b).

Homophily generally comes in two distinct forms that are hard to disentangle: Choice homophily and induced homophily (McPherson, Smith-Lovin, and Cook, 2001). The former arises as a function of individual preferences for similarity while the latter is purely a function of the opportunities people have to come into contact with each other. Exposure leads to greater opportunities for choice homophily but the degree to which exposure is causing friendship is uncertain. In previous studies it seems as if the level of analysis of the exposure is crucial and neighborhoods do not seem to be close enough, and may even lead to increased animosity, dorm rooms and army teams teems seem to be close and repetitive enough. It is an open question whether classrooms and desk mates fall in the positive contact realm or the negative conflict realm.

2 The field experiment and sample

We execute a large-scale randomized field experiment in 182 classrooms of 38 Hungarian primary schools (after exclusions) containing 3539 students. The intervention consists of randomizing the seating chart within each classroom at the beginning of the fall semester, 2017, and encouraging adherence until the end of the semester in January 2018. Endline outcomes data are collected and will become available to the research team in May 2018.

In the spring of 2017, we contacted all primary schools in 7 contiguous counties of central

Hungary via the heads of the local school districts to elicit information about room layouts and seating practices. By the end of the summer vacation, we obtained initial participation agreements with 55 schools in which most 3rd-8th grade classrooms were anticipated to meet a set of inclusion criteria. These criteria were: 1) Principals and teachers would implement our randomized seating chart in three subjects: Hungarian literature, Hungarian grammar, and mathematics. 2) All students in a classroom would receive instruction in these subjects together (e.g., no ability grouping). 3) Classroom layout would comprise free-standing desks that seat two students.

The intervention assigned students to free-standing two-person desks via unconstrained random partitioning within each classroom. We based the randomization on the class rosters from the spring semester. Shortly before the start of the fall semester, we submitted the randomized seating charts to teachers and teachers were instructed to use the charts for the duration of the fall semester until January 2018. To account for changes to class rosters during the summer via exits and entries, we instructed teachers to fill seats vacated by exiting students with entering students from left to right, front to back, in alphabetic order of entering students' surnames. Since, (i) in expectation, students enter and exit classrooms for the same reason (repeating grades and residential moves); and (ii) student surnames are reasonably orthogonal to student grades, this replacement rule preserves randomization.

While teachers were expressly permitted to reseat students if they have to, we asked to preserve the desk-mate composition wherever possible. We measured compliance through teacher reports of the actual seating chart for September 15, 2017. The field team again recorded the actual seating chart during school visits between October and December 2017 and verified classroom layouts through classroom photographs.

Schools and classrooms that do not meet our conditions are dropped from the study. To date, we have dropped 133 classrooms for the following reasons: Withdrawal from the study (25); Less than 10 students at baseline (8); split classrooms (10); Not free-standing desks

that seat two students (30); Unreliable baseline reporting (7); Did not implement the seating chart (40); Does not include information on Roma ethnicity (13). Based on these school- and classroom-level exclusions, we anticipate an analysis sample of 3539 students across 182 classrooms of 38 schools.

Decision rules for dropping future observations: We will exclude students who are randomly assigned to sit alone at a desk at baseline and who have missing values on our outcomes.

Decision rules for dropping variables: If 95 percent or more of the sample answers the same value on a variable we define this as limited variation. We will drop variables with limited variation from the analysis.

Missing values: If we have missing values on variables we will code the variables as zero and include dummy variables controlling for missing status so that we do not loose observations. If more than 30 percent of the respondents do not answer a particular question, it will no longer be seen as a main outcome variable.

3 Data and coding of main variables

We collect baseline variables via teacher reports. Outcome variables are collected via a student survey at endline. In particular, we field a 45-minute two-part in-class survey (see appendix). The first part (20 minutes) consists of a student questionnaire that elicits self-reported grades for the spring and fall semester 2017, academic self-concept, and several attitudinal measures. The second part of the endline survey consists of a reading comprehension test that is not used in this paper. Since the endline questionnaire contains a survey experiment with two vignettes, we randomly sort questionnaires, using a random number generator. Data collection will conclude in April of 2018. The research team will receive outcomes data in May, 2018.

Treatment variables: We define our (exposure) treatment variable, *Treatment* as equal

to 1 if a person is assigned a desk mate that is Roma and zero otherwise. We also have a treatment variable in the survey experiment that we call *Roma vignette*, which is equal to one if the vignette in question 9b includes the bold text saying that the classmate to lend money is Roma (see below).

Primary and secondary outcome variables: We have 2 primary outcome variables: *Roma friend* and *Lend to Classmate*.

Roma friend captures whether an individual has a Roma friend among his or her best friends. The variable is from survey question 5d. Survey question 5 prompts: "Now in general think of your best friends, not just in the class but EVERYWHERE.", and option d is "Among your best friends, how many are Roma (gypsy)?". We code the variable as 1 if the individual has at least one Roma friend and zero otherwise.

The variable *Lend to Classmate* is based on a survey experiment where students were presented with a scenario where they could lend money to a classmate, survey question 9b. The survey question 9b builds on question 9a, which reads: "Imagine that you are going to the zoo with some of your classmates. Your desk mate (who you sat next to in Hungarian class in December) has forgotten to bring money for the entrance ticket. You have enough money for two entrance tickets. Would you lend your desk partner the money for the entrance ticket?". Question 9b then reads: "Now imagine that it is not your desk mate, but a different classmate who has forgotten to bring money with him/her. **This classmate is a Roma/Gypsy.** Would you lend this **Roma/Gypsy** classmate the money for the entrance ticket?" The bold text is only presented to a random half of the students. The answer categories are Yes, No, I do not know. We will recode the variable to be 1 for Yes and zero otherwise.

We have several secondary outcome variables. These variables will not necessarily be analyzed as extensively nor by themselves be seen as confirmatory. Of special interest among these are the variables *Lend to Roma* and *Lend to non-Roma*. These variables take the

same values as the main variable, *Lend to Classmate*, but they are only defined for different samples. *Lend to Roma* is only defined for individuals receiving the bold text in the vignette and *Lend to non-Roma* is only defined when the vignette excludes the bold text. For all classrooms that are majority non-Roma, a random classmate will be more likely to be someone from the in-group for non-Roma respondents. Hence, if we restrict the sample to majority non-Roma, the variables can be used to test whether close exposure to a Roma desk mate affects both in out-group and in-group trust.

In addition to investigating the probability of having a Roma friend we will also investigate effects on the *Number of Roma friends*, which just counts the number of Roma friends in question 5d. We expect that we will get similar results with both variables.

Control variables: We only include control variables that are collected at baseline or stable over time. The variables we include are age (in 0.1 years), gender and spring 2017 grades in five core subjects (Hungarian literature, Hungarian grammar, mathematics, diligence, and behavior). These variables are obtained from the classroom teacher.

Other variables: There are a set of questions that will be used for supplementary analyses. Survey question 9a will be used to create a variable, *Lend to Desk mate*. We will also create other variables such as *Desk mate among best friends* (to see if desk mate relations in general are characterized by friendship potential) and *Liked sitting next to desk mate*.

Heterogeneity: The possibilities for heterogeneous treatment effects are endless. Both characteristics of the exposed and the exposer are likely to matter. It is likely that people that are similar to each other in other aspects have a higher likelihood of transmitting or changing attitudes of the desk mate. With so many options, the heterogeneity analysis will necessarily be seen as explorative. We here outline some of the aspects we will explore. We will interact students' own baseline GPA with Treatment in a model including GPA as well. We will control for and interact a variable for whether the desk mates are of the same sex or of different sex. At the contextual level there are also many possible moderators and we

will investigate the moderating role of *Share of Roma in class*.

4 Empirical strategy

Identifying peer effects is difficult as people self-select into networks and since outcomes are affected by correlated effects (Manski, 1993). With random variation in peer contact we get around most of the challenges associated with identifying network effects. In order to focus on the effect of exposure to a stigmatized minority group on the attitudes and preferences of the majority population, we exclude the Roma students themselves from the regressions.

We first estimate the following regression to identify the treatment effect on the probability to have a Roma friend:

(1) $Romafriend_{ict2} = \beta Treated_{ict1} + \alpha Class_{ct1} + \gamma X_{ict1} + \epsilon_{ict2}$,

where i indexes individuals, c classes, and t is time (either baseline 1 or follow up 2). $Treated_{ict1}$ is a dummy equal to 1 if this person is assigned a Roma desk mate, X_{ict1} is a set of individual level control variables either measured at baseline or reflecting stable characteristics (described in section 3), and the error term, ϵ_{ict2}. We will present results with and without the baseline controls but the main specification is without controls. We use robust standard errors in all estimations. The standard errors do not need to be clustered at any level as the randomization is at the individual level (see Abadie et al., (2017)). The class fixed effects are included as the randomization was conducted within classes.

The vector of individual level control variables is included as they may increase power. To make the models fully saturated, we partition the covariate space and add these control variables as indicator variables rather than using their multi-valued codings and we also interact them with treatment (Athey and Imbens, 2017). We create an indicator for missing values in the controls and include the missing indicator in the regressions in order not to lose observations.

The same specification is also run for *Number of Roma friends* as the outcome variable.

For our second main outcome variable we estimate the following regression:

(2) Lend to Classmate$_{ict2}$ = $\beta Treated_{ict1} + \theta$ Roma vignette$_{ict2}$ + $\delta Treated_{ict1}$*Roma vignette$_{ict2}$ + $\alpha Class_{ct1} + \gamma X_{ict1} + \epsilon_{ict2}$,

where we add the variables *Roma vignette*, which equals one if the bold text in the vignette is included, and the interaction between *Roma vignette* and *Treatment*.

We also run the same specification without the interaction term and without *Roma vignette* separately for *Lend to Roma* and *Lend to non-Roma* and for a sample restricted to non-Roma majority. In the analysis of lending to Roma we will also investigate whether it makes a difference whether or not the person lent to his or her desk mate.

To explore heterogeneity we will first interact the treatment variable with the baseline control variables (Gender and baseline grades). We will also test whether the effect is different in classes with relatively many and relatively few Roma by interacting treatment with *Share of Roma in class*. The standard errors will then be clustered at the class level.

We will also use machine learning techniques to automate the search for heterogenous treatment effects. There are many different types of machine learning algorithms and we have have decided to use classification and regression trees (R package *causalTree*, (Athey and Imbens, 2016)); and random forests (R package *grf*, (Wager and Athey, 2017)). As this field is moving rapidly, however, it is possible that there will be other techniques that are relevant for us once we start analyzing the data.

Balance tests: To test for balance we will regress our main treatment variable on the control variables described above both individually and together, while controlling for class fixed effects. We will judge whether the randomization worked by conducting an F-test of whether the control variables jointly predict treatment status.

5 Hypotheses

In the literature on interethnic exposure there are, broadly speaking, two perspectives on the effects of diversity. One perspective argues that diversity leads to negative outcomes. Several empirical studies find patterns that are consistent with what is denoted conflict theory; diversity is associated with e.g. less trust (Alesina and La Ferrara, 2002; Delhey and Newton, 2005; Dinesen and Sønderskov, 2015; Stolle, Soroka, and Johnston, 2008). The other perspective is more positive. Contact theory (Allport, 1954) suggests that personal contact with members of out-groups can reduce prejudice and misperceptions, and thereby increase trust. There is ample evidence from well identified studies using random assignment, either of students (e.g. Boisjoly et al., 2006; Burns, Corno, and La Ferrara, 2016) or within the military (Carrell, Hoekstra, and West, 2015; Finseraas and Kotsadam, 2017b; Finseraas et al., 2016), showing that personal contact reduces prejudice and strengthens cooperation (Goette, Huffman, and Meier, 2006).

According to contact theory, the positive effects of personal contact are expected to apply when certain criteria are met (Allport, 1954). The contact should take place in a context with equal status, shared common goals, be cooperative, and take place under some form of authority (Pettigrew, 1998). Finally, the setting should have friendship potential, which increases the probability of affective ties and willingness to learn about out-group members (Van Laar et al., 2005). In fact, several authors argue that friendship potential is the most essential condition (Laurence, 2009; Pettigrew, 1998; Stolle, Soroka, and Johnston, 2008). Contact theory has received support in several field experiments with randomly assigned contact (e.g. Boisjoly et al., (2006), Burns, Corno, and La Ferrara, (2016), Carrell, Hoekstra, and West, (2015), Finseraas and Kotsadam, (2017b), and Finseraas et al., (2016)) but most of the evidence is based on correlational patterns (see Brown and Hewstone, (2005) and Pettigrew et al., (2011), and Paluck, Green, and Green, (2017) for reviews).

Our first main hypothesis is that being randomly assigned to a Roma desk mate increases the probability of having a Roma as one of the best friends. The reasons for this are that induced homophily is larger and that contact theory is likely to operate at this very personal level. The hypothesis will be seen as confirmed if β is positive and statistically significant in equation 1.

Our second main hypothesis is that being randomly assigned to a Roma desk mate increases the relative probability of wanting to lend money to a Roma classmate rather than a random classmate. The reasons for this is that contact theory is likely to operate at this level and that we expect friendship ties (as in the hypothesis above). The hypothesis will be seen as confirmed if δ is positive and statistically significant in equation 2.

The second hypothesis is thereby tested by a difference in difference model. The estimate gives us the differential effect of contact on sending money to a classmate when being given the Roma vignette. The coefficients for *Treatment* and for *Roma vignette* will also be interesting to investigate. In particular in classes that are majority non-Roma, as the coefficient for *Treatment* will then show if having a Roma desk mate affects in group trust. Following constrict theory, exposure to ethnic diversity will lead to lower trust towards the in-group as well. However, trust may increase also to the in-group by being exposed to people that were mistakenly thought of as less trustworthy before contact (see Finseraas et al. 2018 for a similar reasoning). The coefficient for *Roma vignette* will tell us the difference in willingness to lend to a Roma classmate for individuals not exposed to a Roma desk mate. As outlined in the empirical strategy, we will also investigate these aspects by running separate regressions of *Lend to Roma* and *Lend to non-Roma* on *Treatment* for a sample of non-Roma majority.

As conflict and contact theories envisages different types of interactions between the majority and minority individuals, they may both be correct at the same time. Many contributions highlight this fact (e.g. Abascal and Baldassarri, (2015), Dinesen and Sønderskov, (2015), and Valdez, (2014)) and already Allport (1954) argued that shallow exposure may

increase rather than decrease antipathy towards minorities. Furthermore, a series of contributions argue that contact may diminish or even reverse the negative effects of exposure (Laurence, 2009; McLaren, 2003; Schneider, 2008; Stolle, Soroka, and Johnston, 2008; Uslaner, 2012). The argument is that the threatening aspects of exposure are mitigated by contact or that social interactions changes the very conception of whom is considered to be in the in-group (McLaren, 2003; Stolle, Soroka, and Johnston, 2008). The empirical evidence for these claims is exclusively based on correlations whereby individuals self-select into having contact with or being friends with minorities.

We use our data to contrast and combine the conflict and contact perspective on ethnic diversity, by studying treatment heterogeneity according to previous exposure to diversity. We do not have random variation in the exposure to Roma at other levels of analyses but we will explore whether close personal exposure has a different impact in classes with more or less Roma. As this analysis is explorative we remain agnostic as to the direction of the heterogeneity in the effect. We also expect that there may be heterogeneity in the effects based on whom is exposed and based on qualities of the Roma child the person is exposed to. We will investigate this exploratively and we think that grades and gender may be important moderators. In an observational study, Hajdu, Kertesi, and Kezdi, (2017b) find that academically high achieving Roma students have more interethnic friendships. Other heterogeneity analyses are outlined in section 3.

6 Power calculation

In testing our different hypotheses we are restricting the sample to non-Roma individuals. For the test of the difference in difference model we will furthermore base the power calculation on half of the sample as only a random half is assigned the Roma version of the vignette.

We also adjust the p-values for the fact that we are testing two hypotheses. We follow the

recommendations of Fink, McConnell, and Vollmer, (2014) and use a method developed by Benjamini and Hochberg, (1995) and Benjamini and Yekutieli, (2001) to minimize the false non-discovery rate (see also Almeida, (2012) and Finseraas and Kotsadam, (2017b) for pre-analysis plans with the same decision rules for correction of p-values). The main advantage of the method is that it is limiting the risk of false discoveries while only adjusting the critical values based on other true hypotheses. The false discovery rate method developed by Benjamini and Hochberg (1995) implies that the m p-values of the i hypotheses are ordered from low to high and that the critical value of the p-value is then p(i) = a*i/m. In our case, with 2 hypotheses and a significance level (a) of 0.05, the critical p-value would be 0.025 for the one with the lowest p-value (0.05* 1/2, which is the same as a Bonferroni correction). For the second hypothesis, the critical p-value is 0.05 (0.05*2/2).

Conservatively, we expect to have a sample of at least 2000 non-Roma individuals in our samples. We calculate power using the program optimal design and if we use the most conservative p-value of 0.025 we have a minimum detectable effect (MDE) of 0.14 for the Roma friend hypothesis. For the second hypothesis, wanting to lend money to a Roma classmate, we only have half as many people in each cell since it is based on an interaction term. Our calculated MDE for this hypothesis with the most conservative p-value of 0.025 is 0.2. We therefore think that our study is well powered to detect relatively small effects.

7 IRB approval and consent

This study was reviewed and approved by the IRB offices at the Hungarian Academy of Science (data collection and analysis); and at the University of Wisconsin-Madison (data analysis). We obtained consent at multiple points. First, we asked school administrators and teachers to consent to participate in the study. Second, we had the teachers ask the parents to consent to data collection about their children.

8 Archive

The pre-analysis plan is archived before any endline data is received. We archive it at the registry for randomized controlled trials in economics held by The American Economic Association: https://www.socialscienceregistry.org/ on March 22 2018. We will receive the endline data in May 2018.

References

Abadie, Alberto et al. (2017). *When Should You Adjust Standard Errors for Clustering?* Tech. rep. National Bureau of Economic Research.

Abascal, Maria and Delia Baldassarri (2015). "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." In: *American Journal of Sociology* 121.3, pp. 722–782.

Alesina, Alberto and Eliana La Ferrara (2002). "Who Trusts Others?" In: *Journal of Public Economics* 85.2, pp. 207–234.

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

Almeida, Rita et al. (2012). *The Impact of Vocational Training for the Unemployed in Turkey: Pre-Analysis Plan*. Pre-analysis plan posted at povertyactionlab.org.

Athey, Susan and Guido Imbens (2016). "Recursive partitioning for heterogeneous causal effects." In: *Proceedings of the National Academy of Sciences* 113.27, pp. 7353–7360.

Athey, Susan and Guido W Imbens (2017). "The Econometrics of Randomized Experimentsa." In: *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 73–140.

Baldassarri, Delia and Maria Abascal (2017). "Field Experiments Across the Social Sciences." In: *Annual Review of Sociology* 43.1.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.

Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency." In: *Annals of statistics*, pp. 1165–1188.

Boisjoly, Johanne et al. (2006). "Empathy or Antipathy? The Impact of Diversity." In: *American Economic Review* 96.5, pp. 1890–1905.

Brown, Rupert and Miles Hewstone (2005). "An integrative theory of intergroup contact." In: *Advances in experimental social psychology* 37, pp. 255–343.

Burns, Justine, Lucia Corno, and Eliana La Ferrara (2016). *Interaction, prejudice and performance. Evidence from South Africa*. Tech. rep. Working Paper.

Byrne, Donn Erwin (1961). "Interpersonal attraction and attitude similarity." In: *The Journal of Abnormal and Social Psychology* 62.3, pp. 713–715.

— (1971). *The attraction paradigm*. New York: Academic Press.

Carrell, Scott E., Mark Hoekstra, and James E. West (2015). *The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences*. NBER Working Paper No. 20940. http://www.nber.org/papers/w20940.

Delhey, Jan and Kenneth Newton (2005). "Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?" In: *European Sociological Review* 21.4, pp. 311–327.

Dinesen, Peter Thisted and Kim Mannemar Sønderskov (2015). "Ethnic diversity and social trust: Evidence from the micro-context." In: *American Sociological Review* 80.3, pp. 550–573.

Fink, Günther, Margaret McConnell, and Sebastian Vollmer (2014). "Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures." In: *Journal of Development Effectiveness* 6.1, pp. 44–57.

Finseraas, Henning and Andreas Kotsadam (2017a). "Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment." In: *European Journal of Political Research* 56.3, pp. 703 –722.

— (2017b). "Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment." In: *European Journal of Political Research* 56.3, pp. 703–722.

Finseraas, Henning et al. (2016). "Exposure to female colleagues breaks the glass ceiling: Evidence from a combined vignette and field experiment." In: *European Economic Review* 90, pp. 363–374.

Goette, Lorenz, David Huffman, and Stephan Meier (2006). "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." In: *The American Economic Review* 96.2, pp. 212–216.

Hajdu, Tamas, Gabor Kertesi, and Gabor Kezdi (2017a). *Health Differences at Birth between Roma and Non-Roma Children in Hungary-Long-Run Trends and Decompositions*. Tech. rep. Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences.

— (2017b). "Inter-Ethnic Friendship and Hostility between Roma and Non-Roma Students in Hungary." In:

Jackson, Matthew O. (2014). "Networks in the understanding of economic behaviors." In: *The Journal of Economic Perspectives* 28.4, pp. 3–22.

Kende, Anna, Linda Tropp, and Nóra Anna Lantos (2017). "Testing a contact intervention based on intergroup friendship between Roma and non-Roma Hungarians: reducing bias through institutional support in a non-supportive societal context." In: *Journal of Applied Social Psychology* 47.1, pp. 47–55.

Kertesi, Gábor and Gábor Kézdi (2011a). "Roma employment in Hungary after the post-communist transition." In: *Economics of Transition* 19.3, pp. 563–610.

— (2011b). "The Roma/non-Roma test score gap in Hungary." In: *American Economic Review* 101.3, pp. 519–25.

Laurence, James (2009). "The effect of ethnic diversity and community disadvantage on social cohesion: A multi-level analysis of social capital and interethnic relations in UK communities." In: *European Sociological Review* 27.1, pp. 70–89.

McLaren, Lauren M (2003). "Anti-immigrant prejudice in Europe: Contact, threat perception, and preferences for the exclusion of migrants." In: *Social forces* 81.3, pp. 909–936.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks." In: *Annual review of sociology*, pp. 415–444.

Paluck, Elizabeth Levy, Seth Green, and Donald P Green (2017). "The Contact Hypothesis Revisited." In: Available at SSRN: https://ssrn.com/abstract=2973474.

Pettigrew, Thomas F (1998). "Intergroup contact theory." In: *Annual review of psychology* 49.1, pp. 65–85.

Pettigrew, Thomas F et al. (2011). "Recent advances in intergroup contact theory." In: *International Journal of Intercultural Relations* 35.3, pp. 271–280.

Putnam, Robert D. (2007). "E pluribus unum: Diversity and community in the twenty-first century." In: *Scandinavian Political Studies* 30.2, pp. 137–174.

Schneider, Silke L (2008). "Anti-immigrant attitudes in Europe: Outgroup size and perceived ethnic threat." In: *European Sociological Review* 24.1, pp. 53–67.

Stolle, Dietlind, Stuart Soroka, and Richard Johnston (2008). "When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions." In: *Political Studies* 56.1, pp. 57–75.

Uslaner, Eric M (2012). *Segregation and mistrust: Diversity, isolation, and social cohesion*. Cambridge University Press.

Valdez, Sarah (2014). "Visibility and votes: A spatial analysis of anti-immigrant voting in Sweden." In: *Migration Studies* 2.2, pp. 162–188.

Van Laar, Colette et al. (2005). "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." In: *Journal of Experimental Social Psychology* 41.4, pp. 329–345.

Wager, Stefan and Susan Athey (2017). "Estimation and inference of heterogeneous treatment effects using random forests." In: *Journal of the American Statistical Association* Forthcoming.

DeskPartner Research Survey Instrument. (c) 2018 by Tamas Keller and Felix Elwert. All rights reserved. Used with permission.

Write your name! Do not use your nickname!

Family Name	Christian/Given Name(s) (write all your given Names)

Which Grade/Class are you in? (e.g: 3/a)

☐/☐

When were you born?

Year: ☐☐☐☐ Month: ☐☐ Day: ☐☐

3rd Year

MTA Social Sciences Research
„Lendület" RECENS Research Group
„Lendület" RECENS Kutatócsoport

DESK PARTNER RESEARCH

2018.

YOUR TEACHER will complete these tables.

School's name or official stamp

. . . . . .

Comprehension

A

Student ID Number (see separate sheet)

☐☐☐☐☐☐

Test date

Month: ☐☐ Day: ☐☐

Test Start Time

Hour: ☐☐ Min.: ☐☐

Test End Time

Hour: ☐☐ Min.: ☐☐

The research is supported by the National Research, Development, and Innovation Office in the framework of the Youth Research (FK) 125358 competition.

General Information about the Exercises

Please read the following information carefully, and then start answering the questions in the notebook!

The test notebook consists of two parts.

In Part 1, we ask questions about you, or rather we are interested in your opinions. Here it is important for us that we get to know what you think.

In the test notebook's second part you will find comprehension exercises. Please read the assignments carefully, and answer the questions to the best of your knowledge!

Start doing the exercises from the beginning of the notebook! (i.e. start at the beginning?)

Always indicate your answer to the question by shading the corresponding circle. As shown in the image below.

Please make sure that you **only mark one answer for each question!**

If you have already marked an answer, but then change your mind, clearly cross out the first mark or put an X over it, and then shade in the answer you think is correct in the way shown below!

Good luck (with the work)!

Part 1

STUDENT QUESTIONNAIRE

padtars_3_Szovegertes_A.indd 3 2018. 01. 24. 14:27:59

1. What grades did you receive at the END OF LAST SEMESTER in the following? Think of the report card you received this January.

Mark the appropriate number in each row! Only shade in one circle!

	1	2	3	4	5	I don't remember
a) Behaviour	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
b) Diligence	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
c) Hungarian Language	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
d) Literature	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
e) Mathematics	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

2. What grades did you get at the END OF LAST SCHOOL YEAR in the following? Think of the report card you received last summer in June.

Mark the appropriate number in each row! Only shade in one circle!

	1	2	3	4	5	I don't remember
a) Behaviour	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
b) Diligence	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
c) Hungarian Language	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
d) Literature	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
e) Mathematics	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

3. How much do you like the following subjects?

Mark the appropriate number in each row! Only shade in one circle!

	Really Don't like	Don't Like	Neutral	Like	Really Like	Don't know
a) Hungarian Language	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
b) Literature	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6
c) Mathematics	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6

4. Please think of your best friends in your class. In the table below, write down who your 5 best friends are in the class

If you have fewer than 5 friends in your class, then write fewer names in the table. Be sure to write your friends' full names into the table, in other words both their family names and their Christian/given names. Do not use your friends' nicknames! Ask for your teacher's help if you don't know your friends' family names!

	Family Name	Christian/given name (write in all Christian/given names, do not use nicknames!)
1.		
2.		
3.		
4.		
5.		

5. Now in general think of your best friends, not just in the class but EVERYWHERE.

Write in the appropriate number in each row of the table!

	Please write in the appropriate number to the question!
a) In total how many best friends do you have?	
b) Among your best friends, how many are boys?	
c) Among your best friends, how many are girls?	
d) Among your best friends, how many are roma (gypsy)?	

6. Now think of that desk partner who you sat next to in December in Hungarian class. Write down the full name of this desk partner!

If you did not have a desk partner in December in Hungarian class, please shade in this circle, and do not fill in the table!

○ 1

Family Name	Christian/Given Name (Write in all given names, do not use nicknames!)

7. How much did you like sitting next to your desk partner?

Mark the corresponding number! Only shade in one circle!

Really Did not like	**Did not like**	**Neutral**	Liked	**Really liked**	**Don't know**	**Did not have** A desk partner in December
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

padtars_3_Szovegertes_A.indd 5 2018. 01. 24. 14:28:00

8. Think of Hungarian language, literature and mathematics . The following questions relate to how good you think you are in these subjects.

In each row mark the number you consider to be true! Only shade in one circle in each row!

Let's start with HUNGARIAN LANGUAGE!

In your opinion how good are you at Hungarian language?

I am very bad at Hungarian			**I am average at Hungarian**			**I am very good at Hungarian**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your classmates how good are you at Hungarian language?

In the class I am among the worst at Hungarian			**In the class I am average at Hungarian**			**In the class I am among the best at Hungarian**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your other subjects how good are you at Hungarian language?

I am much worse at Hungarian than at other subjects			**I am as good at Hungarian as at the other subjects**			**I am much better at Hungarian than at other subjects**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Now think of LITERATURE!

In your opinion how good are you at literature?

I am very bad at literature			**I am average at literature**			**I am Very good at literature**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your classmates how good are you at literature?

In the class I am among the worst at literature			**In the class I am average at literature**			**In the class I am among the best at literature**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your other subjects how good are you at literature?

I am much worse at literature than at other subjects			**I am as good at literature as at other subjects**			**I am much better at literature than at Other subjects**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

padtars_3_Szovegertes_A.indd 6 2018. 01. 24. 14:28:00

Finally, think of MATHEMATICS!

In your opinion how good are you at mathematics?

I am very bad at mathematics			**I am average at mathematics**			**I am very good at mathematics**	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your classmates how good are you at mathematics?

In the class I am among the worst at mathematics			**In the class I am average at mathematics**			**In the class I am among the best** at mathematics	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

Compared to your other subjects how good are you at mathematics?

I am much worse at mathematics than at other subjects			**I am as good at mathematics as I am at the other subjects**			**I am much better at Mathematics than at** Other subjects	**I don't know**
○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8

9. Imagine that you are going to the zoo with some of your classmates. Your desk partner (who you sat next to in Hungarian class in December) has forgotten to bring money for the entrance ticket. You have enough money for two entrance tickets. Would you lend your desk partner the money for the entrance ticket?

*Shade just **one** circle in!*

a) Yes ○ 1

b) No ○ 2

c) I don't know ○ 3

d) I didn't have a desk partner in Hungarian class December ○ 4

Now imagine that it is not your desk partner, but a different class mate who has forgotten to bring money with him/her. This classmate is a Roma/Gypsy [This sentence is missing in Version B]. Would you lend this Roma/Gypsy [Roma/Gypsy omitted from Version B]classmate the money for the entrance ticket? [i.e. Version B makes no mention of Roma/Gypsy otherwise it is the same]

*Shade just **one** circle in!*

a) Yes ○ 1

b) No ○ 2

c) I don't know ○ 3

padtars_3_Szovegertes_A.indd 7 2018. 01. 24. 14:28:00

10. Now think about how good the boys and how good the girls are at Hungarian language, literature, and mathematics. In your opinion when it comes to Hungarian language, to literature and to mathematics, are the boys better, or are the girls better, or are they equally good?

In each row mark the corresponding number that you consider to be true!

	THE BOYS are much better than the girls	THE BOYS are somewhat better than the girls	The boys and the girls are EQUALLY good	THE GIRLS are somewhat better than the boys	THE GIRLS are much better than the boys
a) Hungarian	○ 1	○ 2	○ 3	○ 4	○ 5
b) Literature	○ 1	○ 2	○ 3	○ 4	○ 5
c) Mathematics	○ 1	○ 2	○ 3	○ 4	○ 5

11. Now think of the classmate of yours whom you consider to be the cleverest. Is this classmate a boy or a girl?

*Shade just **one** circle in!*

a) Boy ○ 1

b) Girl ○ 2

c) I can't say who is the cleverest ○ 3

12. Now think of an assignment that a group of children must solve/do together. What do you think, which group would be able to do this assignment better?

Shade just **one** *circle in!*

a) A group only of boys ○ 1

b) A group only of girls ○ 2

c) A group with both boys and girls in it ○ 3

d) I don't know ○ 4

<u>The following questions (13 & 14) are administered in Grades 6-8 only</u>

Q13 for 6th and 7th grade [not translated yet]

[Q13 in 8th grade]

13. Please indicate whether or not you applied to grammar school in February 2018! If you applied to several high schools were any of these grammar schools?

Only shade one answer!

a) Yes 1

b) No 2

c) I don't remember 3

[Q14 in 6th-8th grade]

14. Regardless of whether you did or did not apply to grammar school, do you think you would/will be accepted?

*0 means that they would definitely not accept you. 10 means that they would definitely accept you. You can also use numbers between 0 and 10 where the larger the number you circle the more certain you are that they will/would accept you. Only shade **one** answer!*

Definitely will not Accept me										**Definitely Will accept me**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Please continue on to the comprehension exercises!

padtars_3_Szovegertes_A.indd 9 2018. 01. 24. 14:28:00

padtars_3_Szovegertes_A.indd 16 2018. 01. 24. 14:28:03