European Sociological Review, 2024, **XX**, 1–15
https://doi.org/10.1093/esr/jcae044
Advance access publication 16 November 2024
Original Article

OXFORD

Discrimination in track recommendation but not in grading: experimental evidence among primary school teachers in Hungary

Dorottya Kisfalusi[1,2,*], Zoltán Hermann[3,4] and Tamás Keller[2,3]

[1]Institute for Analytical Sociology, Linköping University, 601 74 Norrköping, Sweden
[2]Computational Social Science Research Group, HUN-REN Centre for Social Sciences, 1097 Budapest, Hungary
[3]Institute of Economics, HUN-REN Centre for Economic and Regional Studies, 1097 Budapest, Hungary
[4]Institute of Economics, Corvinus University of Budapest, 1093 Budapest, Hungary
[*]Corresponding author. Email: dorottya.kisfalusi@liu.se

This study examines discrimination in teacher assessments and track recommendations against Roma minority students in Hungary. We conducted a pre-registered randomized experiment among 413 primary school teachers. Participating teachers evaluated six mathematics or literacy and grammar tests with fictitious, randomized student names and recommended a high school track. Our results show mixed evidence for discrimination against Roma students: teachers do not discriminate in test evaluations but do so in high school track recommendations, though this latter effect is small. We find that contextual factors play a substantial role in discrimination in track recommendations: teachers who receive tests with fewer Roma than non-Roma names discriminate against Roma students, whereas teachers who receive tests with more Roma names do not. In the latter case, non-Roma students receive similarly low track recommendations as Roma students in both experimental conditions. The results are consistent with stereotype-based theories of discrimination.

Introduction

Ethnicity is one of the main student characteristics along which substantial educational inequalities occur. In most Western countries, children of some immigrant and ethnic-racial minority groups considerably lag behind majority students in terms of academic achievement and educational attainment (Heath and Brinbaum, 2007; Borgna and Contini, 2014).

Ethnic biases in teacher assessments and track recommendations might contribute to the ethnic educational penalties. In many educational systems, teacher-given grades are taken into account in within-school track placement decisions or at transition points to the next educational level. Similarly, teachers' track recommendations influence students' track placements either because they are binding or because informal recommendations influence families' track choices (Caro *et al.*, 2009; Boone and Van Houtte, 2013). Several observational studies have shown ethnic biases in teacher-given grades (Lindahl, 2007; Ouazad, 2008; Burgess and Greaves, 2013; Kiss, 2013; Botelho, Madeira and Rangel, 2015; Triventi, 2020; Kisfalusi, Janky and Takács, 2021; for a review, see Zanga and De Gioannis, 2023) and track recommendations (Caro *et al.*, 2009; Boone and Van Houtte, 2013; Geven, Batruch and van de Werfhorst, 2018; Timmermans *et al.*, 2018), conditional on students' school performance.

Observational studies, however, often suffer from the problem of omitted variable bias: unobserved differences in students' motivation, aspirations, or effort might explain the ethnic differences in teacher assessments and track recommendations (Small and Pager, 2020). Experimental studies overcome this limitation by eliminating potential unobserved ethnic differences and holding student performance across ethnic groups constant. These studies identify discrimination by randomly assigning minority and majority names to tests teachers need to evaluate. Experimental evidence was found for grading discrimination against low-caste students in India (Hanna and Linden, 2012) and against students of Turkish origin in Germany (Sprietsma, 2013) but not in the Netherlands (van Ewijk, 2011).

Received: December 2023; **revised:** August 2024; **accepted:** November 2024
© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

Furthermore, experimental evidence was found for ethnic discrimination in secondary school track recommendations in Germany and the Netherlands (van Ewijk, 2011; Sprietsma, 2013; Wenz and Hoenig, 2020). In Hungary, experimental studies were only conducted among preservice teachers but not among in-service teachers: while no discrimination was found in grading (Takács, 2018), there was discrimination against Roma students in track recommendations (Bruneau *et al.*, 2020).

This paper analyses discrimination in teacher assessments and track recommendations against Roma students in Hungary, using a pre-registered randomized experiment. In the experiment, in-service teachers were asked to evaluate mathematics and literacy tests of fictitious students and then recommend a high school track (N=413 teachers, 2,478 teacher-test observations). Students' names on the tests were randomized (Roma male, Roma female, non-Roma male, and non-Roma female). In addition to testing the pre-registered hypotheses on whether teacher discrimination exists, we explore contextual effects on discrimination. Specifically, similar to the studies by van Ewijk (2011) and Sprietsma (2013), we examine whether the ethnic composition of the names teachers encounter during the experiment moderates discrimination.

Our contribution to the literature is 2-fold. First, we investigate discrimination in teacher assessments and track recommendations against one of Europe's largest and most disadvantaged non-immigrant minority groups, whereas most previous experimental studies have focussed on discrimination against immigrant groups. In many developed countries, immigrants are a positively selected group in education with respect to their sending countries (Engzell, 2019). Therefore, conditional on socioeconomic background and school performance, children of immigrants often have higher educational aspirations and make more ambitious educational choices than their native peers (Jonsson and Rudolphi, 2011; Salikutluk, 2016; Engzell, 2019). There is no similar evidence in the case of Roma students; a recent study found that Roma students have less ambitious secondary school choices compared with non-Roma students with identical abilities, which is only partly explained by their lower socioeconomic status (Kisfalusi, 2023). Thus, teacher attitudes, stereotypes, and discrimination might significantly differ in the case of children of immigrants and the Roma.

Second, we highlight the role of contextual effects in discrimination. By integrating stereotype-based theories of discrimination with findings from research on school segregation, we argue that the perceived characteristics of the school environment might moderate the extent of discrimination against minority students (McKown and Weinstein, 2008; Glock, Kovacs and Pit-ten Cate, 2019). By asking teachers to evaluate multiple tests and varying the ethnic composition of the names, we show that the composition of the names moderates the effect of the individual names on track recommendations. This has important methodological implications: discrimination experiments are sensitive to the experimental design. The composition of the fictitious names used in the experiment might influence participants' discriminatory tendencies.

The remainder of the paper is organized as follows: in the next sections, we review the relevant theories of discrimination, explain why the ethnic context might play a role in discrimination, and introduce the Hungarian educational context. Then, we present the experimental design, the pre-registered analytical plan, and the steps of the exploratory analysis. In the Results section, we first present the confirmatory analysis of the pre-registered hypotheses; then, we conduct an exploratory analysis of the role of the ethnic context. The final section contains a discussion of the findings and limitations.

Theories of discrimination

In line with the definition of discrimination provided by Blank *et al.* (2004) and Pager and Shepherd (2008), we define discrimination in teacher assessments as the differential treatment of students on the basis of a social category (Bygren, 2020). Economic, sociological, and psychological theories identify two main sources of discrimination: group-related preferences or prejudices, and group-related prior beliefs or stereotypes.

Taste-based theories of discrimination assume that individuals treat social groups differentially based on their preferences for their in-group (Turner, 1975; Tajfel and Turner, 1979; Tajfel, 1982) or prejudices towards the out-group (Blumer, 1958; Blalock, 1967; Quillian, 1995). The most prominent economic model of taste-based discrimination was provided by Becker (1957), who suggested that individuals with inherent preferences against an out-group would be willing to pay a cost to avoid interactions with members of this group.

The other main line of discrimination theories emphasizes the role of imperfect individual information in discrimination and assumes that people, consciously or unconsciously, rely on their group-related prior beliefs, experiences, and stereotypes in social interactions (Quillian, 2006; Lorenz, 2021). These stereotypes may be accurate or inaccurate (Jussim *et al.*, 2009). Social psychological theories suggest that individuals automatically categorize others based on group-specific stereotypes first. Then, as more information becomes available and they are motivated to update their category-based opinion, they integrate individuating

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

information into their person perception (Fiske and Neuberg, 1990; Fiske, 1998). Economic theories of statistical discrimination assume, too, that individuals rely on their prior beliefs about the performance of social groups until information about individual characteristics becomes available (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977). While early work in this literature assumes that decision makers use the correctly perceived average characteristics of the target group, and thus, they rely on information that is correct on average (Phelps, 1972; Arrow, 1973), more recent work on statistical discrimination recognizes that information or prior beliefs about group characteristics might be inaccurate (England and Lewin, 1989; Bohren *et al.*, 2023). We will refer to these social psychological and economic theories collectively as stereotype-based theories of discrimination. In this context, the term 'stereotype' denotes a belief held about a social group that may or may not be accurate on average.

In randomized experiments such as ours, teachers only know the names of the (fictitious) students whose tests they need to evaluate but have no additional knowledge about the students. The quality of the test, however, provides information on students' abilities. In this setting, taste-based theories of discrimination suggest that teachers, who generally belong to the majority ethnic group in the Hungarian context, favour majority students and disfavour minority students in both test evaluations and track recommendations. Although stereotypes might also influence test evaluations if, for instance, teachers assume that minority students cheated during the test, stereotype-based discrimination is more likely to occur in track recommendations. When recommending a secondary school track, teachers usually take into account other factors besides students' abilities (Boone and Van Houtte, 2013). Information about these factors is lacking in the experimental situation. Hence, teachers might substitute the lacking individual information with prior beliefs about group characteristics.

Our pre-registered hypotheses concentrate on whether *there is* discrimination against Roma students in test evaluations and track recommendations. We hypothesize that Roma students receive lower test scores on the same test than non-Roma students (*Hypothesis 1*) and that Roma students receive lower track recommendations than non-Roma students based on the same test results (*Hypothesis 2*). In the discussion, we also address the potential underlying mechanisms that might explain our findings.

The role of the ethnic context in discrimination

In an exploratory analysis, we examine whether the ethnic composition of the names teachers encounter in the experiment moderates discrimination against Roma students. We argue that in discrimination experiments where participants are presented with multiple fictitious names, the composition of the names (i.e., the share of minority names) might carry important information for teachers over and above the individual names (i.e., whether the name is a minority name or not). While individual names indicate students' ethnic group belongings, the ethnic composition of the names might provide information on and activate stereotypes about students' school environment (McKown and Weinstein, 2008; Glock, Kovacs and Pit-ten Cate, 2019; Keller, 2024). Specifically, a higher number of ethnic-sounding names might trigger perceptions of segregated schools.[1]

Research on school segregation from both sides of the Atlantic shows that schools with a high proportion of minority students usually also enrol lower-status majority students than schools with a low proportion of minority students (Brunello and Rocco, 2013; Billings, Deming and Rockoff, 2014). This phenomenon is also observed in Hungary. Since standardized achievement tests demonstrate a considerable disparity in performance by ethnicity and social status, this also means that segregated schools exhibit a lower average performance than schools attended by high-status majority students (Hermann and Kisfalusi, 2023).

As we highlighted earlier, stereotype-based theories of discrimination assume that teachers substitute the lacking individual information with prior beliefs about group characteristics. If a higher number of minority names indeed trigger perceptions of segregated schools, then ethnicity is not the only information teachers might rely on in the experiment. Besides their prior beliefs about minority and majority students' academic achievement, teachers might also rely on their prior knowledge about the achievement level of students attending segregated schools and combine these two pieces of information in their decision-making. How these different stereotypes are combined determines whether we can expect a moderation effect to occur (Nicolas, de la Fuente and Fiske, 2017).

One possibility is that negative stereotypes about minority students and negative stereotypes about students attending segregated schools simply add up. This implies that tests with Roma names as well as tests with non-Roma names receive lower evaluations and track recommendations when the share of minority students is higher compared with the case when this share is lower. However, the ethnic gap is the same in the two conditions, with minority students receiving lower evaluations.

Another possibility is suggested by studies which have demonstrated that students with similar academic abilities are graded less favourably in a class with a higher

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

average achievement level than in a class with a lower average achievement level (Marsh, 1987; Südkamp and Möller, 2009). If this contrast effect occurs, and teachers associate the lower (higher) share of minority names with a higher (lower) average achievement level, then both tests with Roma and non-Roma names are expected to receive higher evaluations and track recommendations when the share of minority students is higher, and therefore, the assumed achievement level is lower. But, again, the ethnic gap should be the same in the two conditions, with minority students receiving lower evaluations.

However, if the stereotypes associated with one ethnic group are more pervasive than those associated with the other group, then the ethnic composition of the names may moderate the effect of the individual names. This is because the additional contextual information may be less influential in the evaluation of the ethnic group that is subject to more robust stereotypes. For instance, if negative stereotypes about minority students are more pervasive than positive stereotypes about majority students, then contextual effects might not influence the evaluation of minority students. Consequently, they will receive lower evaluations regardless of their school context. Conversely, majority names might receive lower (in the case of additive stereotypes) or higher (in the case of a contrast effect) evaluations or recommendations when they appear together with many minority names because, due to less robust stereotypes, the contextual information affects their evaluations.

Institutional background

In Hungary, primary schools provide education in grades 1–8 (ISCED 1 and ISCED 2 levels, from age 6 to 14), whereas upper secondary education encompasses grades 9–12/13. Compulsory education lasts until the age of 16. Thus, after grade 8, students are obliged to choose from three different secondary school tracks. On the one extreme, grammar school is the academic and college-bound secondary track (*gimnázium*) that provides the high school final examination (*érettségi*). On the other extreme, vocational schools (*szakközép iskola*) prepare students for manual professions and trades without providing direct access to tertiary education. Finally, there is a mixed track (*technikum*) that provides the high school final examination and also gives a vocational diploma. Forty-five per cent of Hungarian secondary school students are enrolled in the academic track, 38 per cent are enrolled in the mixed track, and the remaining 17 per cent attend vocational schools (KSH, 2022).

Admission to secondary education is merit-based and depends on students' achievement. On the one hand, students' end-of-semester grades in the core subjects, including Hungarian grammar and literature and mathematics, are taken into account. On the other hand, students participate in a centrally organized admission test in mathematics and Hungarian. The most competitive secondary schools also require an oral exam that the individual secondary schools organize for the applicants.

Although Hungary belongs to those countries where primary school teachers' track recommendations are not binding, teachers' informal recommendations influence students' school choices even in these educational systems (Caro *et al.*, 2009; Boone and Van Houtte, 2013). This is because advice given by primary school teachers is an important clue for parents about which track is most suitable for their children (Boone and Van Houtte, 2013). Empirical studies from Hungary show that teachers communicate their track recommendations to the students, and families take into account these recommendations (Suhajda, 2017). Track recommendations might be especially important for minority students and students from lower socioeconomic backgrounds who are less informed about the requirements of the different tracks than students with highly educated parents (Borgna *et al.*, 2022; Keller, Takács and Elwert, 2022).

Social context

The Roma minority is one of the largest ethnic minorities in Europe (O'Nions, 2016). In Hungary, their share is estimated to be around 5–6 per cent of the total population and around 12–14 per cent of school-age children (Kemény and Janky, 2006). Roma people experience substantial economic and social exclusion (Kertesi and Kézdi, 2011) as well as residential and school segregation (Kemény and Janky, 2006; Kertesi and Kézdi, 2012). As a result, Roma students have far lower average test scores than non-Roma students and are more likely to attend segregated schools providing low-quality education (Havas and Liskó, 2005; Kertesi and Kézdi, 2011; Hajdu, Kertesi and Kézdi, 2019). Furthermore, Roma students are more likely to drop out of secondary school and less likely to attend the academic or the mixed track than non-Roma students. Therefore, they are less likely to obtain the final exam and continue their studies at the tertiary level (Kertesi and Kézdi, 2009; Hajdu, Kertesi and Kézdi, 2014).

Experimental design

Study overview

We conducted a large-scale online experiment between 18 June and 6 September 2021. Our target population consisted of grammar and literature and mathematics teachers in the fifth to eighth grades (ISCED 2 level).

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

We contacted every Hungarian primary school with an invitation letter sent to the schools' publicly available e-mail addresses. The invitation letter was addressed to the school principal. Principals were asked to forward the e-mail to the grammar and mathematics teachers in the school. Principals and teachers were told that the study focuses on differences in teachers' grading practices. The number of teachers allowed to participate in the experiment was limited to five mathematics and five grammar teachers per school. Teachers received 10,000 HUF (equivalent to 34.4 USD at the exchange rate from June 2021) for participation.

Overall, 193 grammar and 220 mathematics teachers participated in the online experiment.[2] Respondents were on average 49.9 years old (SD = 8.9) and had an average teaching experience of 21.9 years (SD = 10.5). The vast majority of the teachers were female in both subjects (grammar: 92.8 per cent; mathematics: 86.4 per cent). The age and gender distribution of participants are similar to those of the primary school teacher population in Hungary (more than 85 per cent of primary school teachers are female, and almost 50 per cent are above the age of 50, see Hajdu *et al.*, 2022: pp. 76–79). Supplementary Table S1 shows that the distribution of primary schools participating in the experiment represents the Hungarian primary schools well according to location, provider, and the share of socioeconomically disadvantaged students. Supplementary Table S2 shows the number of respondents per school.

Teachers' tasks in the experiment

Mathematics teachers were asked to correct six different solutions for the same mathematics test, whereas grammar teachers were asked to correct six different solutions for the same literacy test. The different solutions varied in quality. The fictitious student names on the tests were randomized. Around half of the teachers received the six tests with two Roma (one male and one female) and four non-Roma names (two male and two female), and the other half of the teachers received four Roma (two male and two female) and two non-Roma names (one male and one female). A detailed description of the experimental procedure and the selection of names can be found in the Supplementary Material Appendix A.

Teachers had four tasks in the experiment:

1. They first corrected and evaluated the tests on a 30-point scale. Teachers were not provided with a solution because we wanted to avoid influencing teachers' grading practices.
2. Teachers assigned a grade to the test on the 5-grade scale used in the Hungarian educational system (1 = fail, 2 = pass, 3 = satisfactory, 4 = good, 5 = excellent). Schools (and often teachers within schools) apply different rules to translate test scores to grades. Therefore, it is possible that discrimination does not occur in assigning test scores, but in translating the test scores into grades.
3. Teachers recommended a secondary school track to the students whose tests they evaluated (1—vocational track, 2—mixed track, 3—academic track). Before the track recommendations, we communicated to teachers fictitious, randomized test results that students received in the other school subject (for details, see the pre-analysis plan). We added this random noise since high school admission depends on both mathematics and literacy and grammar tests. Therefore, it is more natural to give a recommendation based on both test results.
4. After the experiment, teachers were asked to fill out a questionnaire focussing on background information. They also had the possibility to provide qualitative feedback.

Research ethics

The study was reviewed and approved by the IRB office at the HUN-REN Centre for Social Sciences. The IRB's decision was issued on 17 June 2021. Our data collection process corresponded to the recent General Data Protection Regulation of the European Union. All participants gave their informed consent prior to their participation in the research, and adequate steps were taken to protect participants' confidentiality.

Statistical analyses

Research transparency

Our statistical analysis follows the detailed pre-registration we submitted to the RCT Registry of the American Economic Association before beginning the fieldwork: https://doi.org/10.1257/rct.7838-2.0. Any deviations from the original pre-analysis plan are indicated in the paper.

We archived the data, analytic scripts, and an anonymized version of the pre-analysis plan on the project's page on the Open Science Framework: https://osf.io/743df/

Data and questionnaires have also been reposited at the Research Documentation Centre of the HUN-REN Centre for Social Sciences: https://doi.org/10.17203/KDK558.

Main variables and coding

We pre-registered two primary outcome variables. First, the *test score* captures the total points teachers assigned to the test, which ranges from 0 to 30 in the case of both subjects. Figure 1 presents the empirical distribution of test scores in the experiment. The

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

figure shows substantial variation in test scores in both subjects.

Second, the *recommended secondary school track* was recoded into a dummy variable, which equals 1 if the teacher recommended a secondary track that ends with the high school final exam (academic or mixed track), 0 otherwise.[3] Hypotheses are tested for these two outcome variables.

Our pre-registered secondary outcome variable is the *grade* assigned to the test (integer between 1 and 5).

The treatment variable *Roma* equals 1 if the name of the fictitious test-writer student is a Roma name; otherwise, Roma = 0.

We explore the effect of the ethnic context on discrimination with the variable *fewer Roma names* (*fR*), which equals 1 if teachers received a test package with two Roma and four non-Roma names and 0 if teachers received a test package with four Roma and two non-Roma names.

Empirical model

In our primary, confirmatory analysis, we pre-registered the following linear regression model with teacher and test fixed effects to test our hypotheses:

$$Y_{t,e,j} = \beta_0 + \beta_1 \times Roma_{t,e,j} + \beta_2 \times male_{t,e,j} + \varphi_e + \theta_t + \epsilon_{t,e,j} \quad (1)$$

where $Y_{t,e}$ is the outcome variable (H1: test score, H2: track recommendation). The index t indicates teacher, the index e indicates test, and the index j indicates the name written on the test. The variable *Roma* indicates (=1) if the name of the fictitious test-writer student was a common Roma name; otherwise, Roma = 0. The variable *male* indicates (=1) if the fictitious test-writer student was male; otherwise, *male* = 0. φ_e represents test fixed effects and θ_t represents teacher fixed effects. β_1 is the coefficient of interest and represents the difference in teachers' evaluation regarding the Roma and non-Roma students' tests/track recommendations. The coefficient has a causal interpretation because test packages were randomly assigned to teachers. Note that in the case of track recommendations, the estimated model is a linear probability model, as the outcome is binary. We calculate standard errors clustered at the teacher level.[4] In our primary analysis, we use one-sided t-tests to test our hypotheses.

In order to gain power and since teacher and test fixed effects are not needed for the identification as test packages were randomly assigned to teachers, we repeated the analyses without teacher and test fixed effects as a (not pre-registered) robustness check.

Our secondary, exploratory analysis consists of various tests.[5] First, we substitute $Y_{t,e}$ with our secondary outcome, which is the grade assigned to the test. Second, to reveal potential heterogeneities in the treatment effect, we investigate the interaction between gender and ethnicity, as well as between test quality and ethnicity, by including the appropriate interaction terms in the model. Third, we restrict our main analysis to teachers who might teach Roma students in

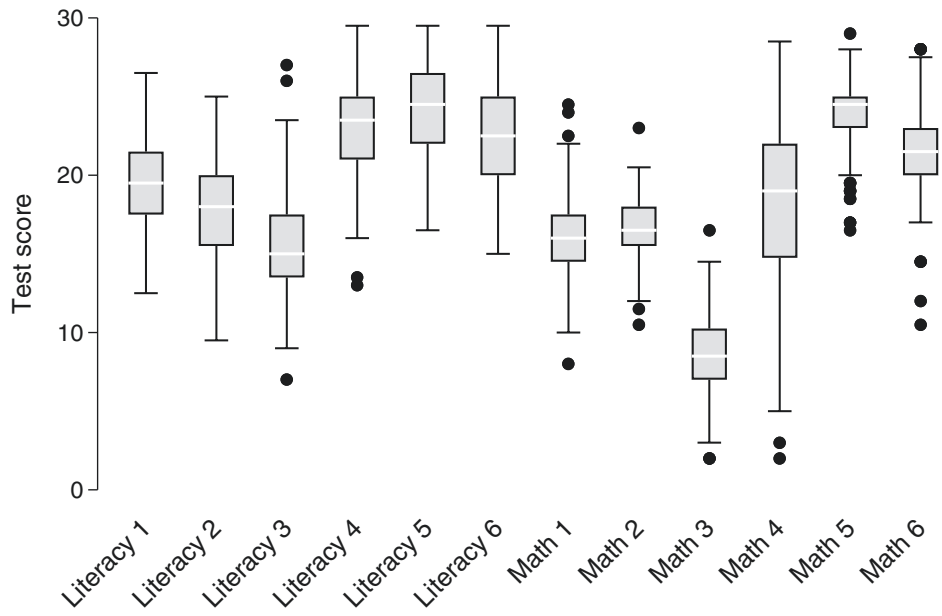

Figure 1 Distribution of scores by test

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

Table 1 Raw differences in the dependent variables by the ethnicity and gender of the fictitious names

	Non-Roma		Roma		Female		Male		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Test score	18.90	5.28	18.73	5.40	18.58	5.35	19.05	5.32	18.82	5.34
Grade	3.09	0.99	3.08	1.03	3.06	1.01	3.11	1.01	3.09	1.01
Recommendation	0.70	0.46	0.66	0.47	0.67	0.47	0.70	0.46	0.68	0.47
N	1234		1244		1239		1239		2478	

Notes: Test score ranges from 0 to 30. Grade ranges from 1 to 5. Recommendation equals 1 if the teacher recommended a secondary track that ends with a high school final exam, 0 otherwise. The difference is significant between Roma and non-Roma students in the case of track recommendations, and between female and male students in the case of test scores (at the $P < 0.05$ level).

their school and, thus, might have the opportunity to discriminate against Roma students in their everyday teaching practices. As Wenz and Hoenig (2020) argue based on Heckman (1998), concentrating on teachers who have actual contact with Roma students provides a more accurate estimation of discrimination that happens in the everyday school context. For this analysis, we restrict the sample to teachers who are employed in schools with at least a 1 per cent share of Roma students (representing 83 per cent of the total sample).[6]

Finally, we explore how the perceived ethnic context plays a role in discrimination. We estimate EQ2 to investigate whether the number of Roma names teachers encountered in the experiment moderates the effect of ethnicity on the outcomes:

$$\begin{aligned} Y_{t,e,j} = \beta_0 + \ & \beta_1 \times Roma_{t,e,j} + \beta_2 \times male_{t,e,j} \\ & + \beta_3 \times fR_t + \beta_4 \times fR_t \times Roma_{t,e,j} \\ & + \varphi_e + \epsilon_{t,e,j} \end{aligned} \quad (2)$$

where the coefficient β_3 is a dummy variable indicating the share of Roma names in the six tests (1 = fewer Roma than non-Roma names, 0 = more Roma than non-Roma names). As teachers received a random set of tests, β_3 has a causal interpretation. The coefficient β_4 captures the difference in the treatment effects in test packages with fewer and more Roma names.[7] Test fixed effects are always included in these specifications because our randomization procedure focussed on balancing the ethnicity of the individual names but not the ethnic composition of the names across the different tests. Since our sample was not large enough to completely balance the share of Roma names across the different tests (see Supplementary Table S3), and thus, across different test qualities, test fixed effects are needed to control for test quality in this exploratory analysis.

Results

Main results

Raw differences in the dependent variables by the ethnicity and gender of the fictitious names are presented in Table 1. Tests with non-Roma names received higher track recommendations on average than tests with Roma names (0.70 vs. 0.66, $P = 0.03$), whereas tests with male names received higher test scores on average than tests with female names (19.05 vs. 18.58, $P = 0.03$).

Table 2 shows the results of the main regression analysis. Models 1 and 2 present the results of the pre-registered confirmatory analysis, testing H1 and H2, respectively. Models 4 and 5 show the same analysis without teacher and test fixed effects. Regardless of the specification, teachers did not evaluate tests with Roma names less favourably than tests with non-Roma names ($\beta_1 = -0.011$, $P = 0.465$ in Model 1). Therefore, Hypothesis 1 is not supported.[8]

With regard to Hypothesis 2, in the pre-registered model, teachers were 2.6 percentage points less likely to recommend a secondary school track ending with the high school final exam for Roma students than for non-Roma students ($\beta_1 = -0.026$, $P = 0.042$, Model 2). Without teacher and test fixed effects, the ethnic difference is 4 percentage points ($\beta_1 = -0.040$, $P = 0.019$, Model 5). The difference between the parameter estimates obtained in Model 2 and Model 5 is statistically not significant. The estimated effect sizes are −0.057 SD and −0.086 SD of the outcome variable in Model 2 and Model 5, respectively. After applying the Benjamini-Hochberg (1995) procedure for multiple hypothesis testing, the coefficient is not significant in the pre-registered model at the 0.05 level. The lack of statistical significance in the pre-registered model is likely the consequence of the fact that we found an effect size that is half as large as the expected effect size we used in the power calculations to design our sample size. Therefore, we interpret this finding as evidence for Hypothesis 2, though the effect size is small.

Models 3 and 6 show that results for the secondary outcome variable—grade assigned to the test—are similar to that of test scores. Teachers did not assign lower grades to Roma than to non-Roma students.

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

Table 2 Results of the main regression analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Test score	Track recommendation	Grade	Test score	Track recommendation	Grade
Roma	−0.011	−0.026*	0.018	−0.164	−0.040*	−0.013
SE	(0.121)	(0.015)	(0.024)	(0.206)	(0.019)	(0.038)
P-value[a]	0.465	0.042	0.232	0.213	0.019	0.369
Male	0.327***	0.016	0.035	0.464**	0.028	0.057*
SE	(0.092)	(0.013)	(0.021)	(0.136)	(0.016)	(0.027)
P-value	<0.001	0.226	0.101	0.001	0.080	0.033
Test FE	Yes	Yes	Yes	No	No	No
Teacher FE	Yes	Yes	Yes	No	No	No
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,478	2,478	2,478	2,478	2,478	2,478
R-squared	0.781	0.425	0.683	0.002	0.003	0.001
Number of teachers	413	413	413	413	413	413

Notes: Models 1 and 2 present the results of the pre-registered confirmatory analysis. Robust standard errors clustered at the teacher level in parentheses. For the fixed effects models (Models 1–3), the within *R*-squared is reported. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.
[a]For the Roma variable, *P*-values of one-sided tests are reported.

Interaction effects with gender and test quality are presented in the Supplementary Material Appendix B. No significant interaction was found between gender and ethnicity (see Supplementary Table S5). However, Supplementary Figure S2 and Supplementary Table S6 suggest that discrimination in track recommendation is more pronounced in the middle range of test scores than at the bottom and the top of the performance distribution.

We repeated the main analysis by restricting the sample to those teachers who teach in schools with at least a 1 per cent share of Roma students and, thus, might have an opportunity to discriminate against Roma students in their everyday teaching practices. These schools constitute 83 per cent of the sample. The results are presented in Supplementary Table S7. For track recommendations, the parameter estimates and effect sizes are slightly smaller in this subsample (−0.049 SD and −0.076 SD of the outcome variable in Models 2 and 5, respectively). Though the estimates are statistically not significant in this restricted sample, the sign and the size of the parameters are very close to those estimated from the entire sample.

The role of the ethnic context

In this section, we explore how the ethnic context affects teacher discrimination. Specifically, we investigate differences between teachers who received tests with fewer Roma than non-Roma names and teachers who received tests with more Roma than non-Roma names.[9]

For test scores and grades as outcome variables, the interaction between the Roma dummy and the share of Roma names is not significant (Models 1 and 3 in Table 3). At the same time, discrimination in track recommendations is associated with the share of Roma names encountered by the respondents. Table 3 Model 2 shows that teachers who received tests with more Roma names were less likely to discriminate in track recommendations than teachers who received tests with fewer Roma names (*fR * Roma:* −0.10, $P = 0.045$). Specifically, the results show that teachers gave similar track recommendations for Roma students in both experimental conditions (the linear combination of the *fR* + *fR * Roma* variables: −0.02, $P = 0.530$). At the same time, they recommended a lower track for non-Roma students if they received tests with more Roma names compared with the condition when they received tests with fewer Roma names (the parameter estimate for the *fR* variable: 0.07, $P = 0.022$, the effect size is −0.151 SD of the outcome variable). In the former case, track recommendations for non-Roma students were as low as those for Roma students in both experimental conditions. These findings are graphically presented in Figure 2, which shows the estimated probabilities of track recommendations for Roma and non-Roma students by the number of Roma names in the experiment.

Discussion

We found mixed evidence for discrimination against Roma students in test evaluations and track recommendations among grade 5–8 teachers in Hungary. In

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

Table 3 Regression results with the share of Roma names in the experiment as explanatory variables

	(1)	(2)	(3)
	Test score	Track recommendation	Grade
Roma	0.149	0.026	0.056
SE	(0.187)	(0.031)	(0.043)
P-value	0.425	0.403	0.200
Male	0.039	0.027	−0.057
SE	(0.212)	(0.022)	(0.050)
P-value	0.853	0.228	0.255
Fewer Roma names	0.203	0.074*	0.040
SE	(0.236)	(0.032)	(0.057)
P-value	0.390	0.022	0.490
Fewer Roma names * Roma	−0.103	−0.095*	−0.021
SE	(0.254)	(0.045)	(0.061)
P-value	0.685	0.036	0.729
Test FE	Yes	Yes	Yes
Teacher FE	No	No	No
Constant	Yes	Yes	Yes
Observations	1,239	1,239	1,239
R-squared	0.804	0.459	0.685
Number of teachers	413	413	413

Notes: Robust standard errors clustered at the teacher level in parentheses. *Fewer Roma names* is an indicator variable for test packages with two Roma and four non-Roma names (vs. four Roma and two non-Roma names). The analysis is restricted to the three mathematics and three grammar tests where Roma and non-Roma names occur in both experimental conditions. *$P < 0.05$.

our experiment, teachers did not award Roma students with lower test scores. However, they recommended a lower secondary track for Roma students than for similarly performing non-Roma students.

The estimated average treatment effect in track recommendations is substantively small but comparable to the findings of previous experimental studies on discrimination in teacher assessment from other countries. It is not possible to make a direct comparison between our effect sizes and those reported in previous studies on track recommendations due to differences in the dependent variables and analytical strategies employed. However, it is possible to compare our effect sizes on track recommendations with those reported in previous studies on grading. In our case, the estimated effect sizes are between −0.057 and −0.086 SD of the outcome variable, depending on the model specification. Previous studies have found effect sizes between −0.01 and −0.12 SD (van Ewijk, 2011; Hanna and Linden, 2012; Sprietsma, 2013; Wenz and Hoenig, 2020). Thus, our effect size lies in the middle of the distribution of effect sizes identified in other studies.

Our finding that teachers do not discriminate against Roma students in test evaluations but do so in track recommendations is in line with previous experiments conducted among preservice teachers studying at Hungarian universities (Takács, 2018; Bruneau *et al.*, 2020) and among in-service teachers in other countries (van Ewijk, 2011; Wenz and Hoenig, 2020). Although track recommendation is not binding in Hungary, families take into account teachers' advice in their secondary school choices (Suhajda, 2017). Discrimination in track recommendations might exacerbate educational inequalities, especially in highly stratified educational systems where track placement substantially determines educational and labour market outcomes (Buchmann and Park, 2009).

Exploratory analyses showed that discrimination in track recommendation is more pronounced in the middle range of the performance distribution than at the bottom and the top of the distribution. These findings align with those of Wenz and Hoenig (2020), who observed discrimination in the case of an average-quality essay but did not find discrimination in the case of a low-quality essay. We contributed to this line of research by showing that discrimination does not occur at the upper end of the performance distribution either. This may also elucidate discrepancies in the findings of different experimental studies, whereby the presence or absence of evidence for discrimination may

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

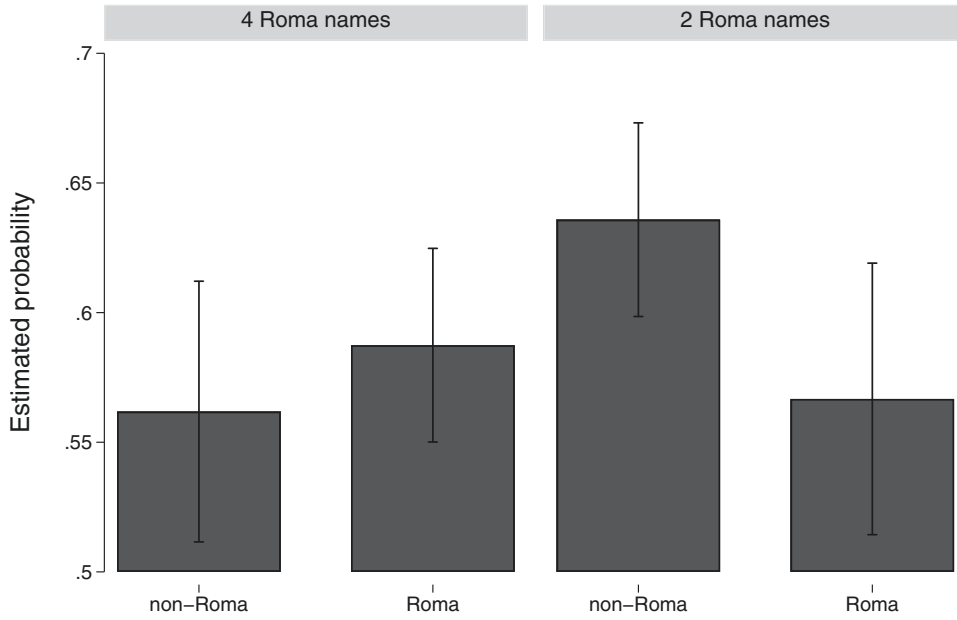

Figure 2 Estimated probabilities of track recommendations for Roma and non-Roma students by the number of Roma names in the experiment (with 95 per cent confidence intervals)

Note: The figure is restricted to the three mathematics and three grammar tests where Roma and non-Roma names occur in both experimental conditions.

be contingent on the quality of the tests employed in the experiment.

Further exploratory analyses showed that contextual factors play a substantial role in discrimination in track recommendations. On the one hand, we found that teachers who received tests with fewer Roma than non-Roma names tended to discriminate against Roma students in track recommendations. For this subsample, the effect size is considerably larger than for the entire sample (–0.151 SD of the outcome variable). On the other hand, teachers who received tests with more Roma than non-Roma names did not discriminate. Sprietsma (2013) found similar evidence for a moderating effect of ethnic composition: in Germany, teacher discrimination was smaller if teachers were presented with a higher number of ethnic-sounding names. In contrast, van Ewijk (2011) did not find such a moderating effect in the Netherlands.

We also investigated whether the ethnic composition of the names moderated the effect of the individual names by influencing track recommendations for Roma students, non-Roma students, or both. We found that teachers who encountered more Roma names in the experiment gave lower recommendations for non-Roma students than teachers who encountered fewer Roma names, whereas track recommendations for Roma students were relatively similar in the two experimental conditions. Therefore, when teachers received tests with more Roma names, Roma and non-Roma students received similarly low track recommendations. This finding is consistent with a recent vignette experiment (Keller, 2024), which showed that in classrooms with a higher share of Roma students, teachers evaluated the performance of both Roma and non-Roma students as lower than in classrooms with a lower share of Roma students.

One possible explanation for this finding is that a higher number of ethnic-sounding names have triggered perceptions of segregated schools: teachers interpreted the composition of the minority and majority names on the tests as signals of the ethnic composition of the schools these students presumably came from. Teachers might have relied on their prior knowledge that students in segregated schools have lower school performance on average and recommended a lower track for both Roma and non-Roma students who were associated with a segregated environment. This interpretation is consistent with the everyday experiences of teachers. In Hungary, segregated schools provide lower-quality education (Havas and Liskó, 2005), and in segregated schools, academic achievement is similar between Roma and non-Roma students (Kertesi and Kézdi, 2011, 2016).

It is important to note, however, that Roma students did not receive higher track recommendations in test packages with more non-Roma names relative to test

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

packages with fewer non-Roma names. This suggests that Roma ethnicity is a strong signal for teachers, and its negative effect is not moderated by the ethnic composition of the school (e.g., the share of Roma students in the test package).

Another possible explanation for the moderating effect of the ethnic composition of the names could be social desirability bias: perhaps teachers who received tests with more Roma names realized the true aim of our study, whereas teachers who received tests with fewer Roma names did not. However, in this scenario, we would expect that social desirability bias generates higher track recommendations for Roma students when the share of Roma names is high compared with recommendations for Roma students in the other case. Put differently, in the case of social desirability bias, track recommendations for non-Roma students should be stable, whereas track recommendations for Roma students should vary across the two experimental conditions. We find the opposite pattern; therefore, social desirability bias is not a likely explanation for our results.

Another alternative explanation that we cannot completely rule out is the minority hypothesis, which states that students who represent the numerical minority in a class are more salient and, therefore, attract more attention from teachers than students who represent the numerical majority (Ready and Wright, 2011; Kaiser, Südkamp and Möller, 2017). This increased attention could lead to more accurate teacher judgements for students in the minority position compared with students in the majority position (Kaiser, Südkamp and Möller, 2017). In our experimental design, we cannot observe how accurate teachers' evaluations are because we do not have an exact measure of student ability against which to compare the teacher-assigned test scores. We only observe the test scores that teachers assign to the students, which are endogenous. This is also true for track recommendation, as there is no exact cut-off above which students could be recommended to the higher school tracks. In sum, we cannot test whether the minority hypothesis played a role in our experimental setting.

Our findings are consistent with stereotype-based theories of discrimination for three main reasons. First, discrimination in track recommendations but not in test evaluations is more in line with stereotype-based than with taste-based discrimination (Wenz and Hoenig, 2020). For evaluating the tests, all information was available to the teachers. However, when recommending a secondary school track, teachers usually take into account several other factors besides students' academic achievement (Boone and Van Houtte, 2013). Qualitative feedback from the participating teachers also supports this: some teachers explicitly wrote that they were missing additional information about the individual students to give a proper track recommendation. Since students' efforts, parental support, and family background were not observable in the experimental situation, teachers needed to rely on their prior beliefs or stereotypes when recommending a track.

Second, the pattern of discrimination in track recommendations depending on the ethnic context provides further indirect evidence for stereotype-based discrimination. It suggests that instead of relying on ethnic prejudices, teachers were actively looking for additional information about the characteristics of students they needed to recommend a track. School composition provides additional information for teachers and can be used as an anchor for beliefs on individual student characteristics.

Third, the fact that discrimination occurred in the middle range of the performance distribution contradicts a simple model of taste-based discrimination and is more in line with social psychological theories of stereotype-based discrimination (Fiske and Neuberg, 1990; Fiske, 1998). In the case of a large deviation from the stereotypical performance level (very high performance for Roma students and very low performance for non-Roma students), teachers might have updated their category-based judgement with an assessment based on individual performance. In contrast, essays of average quality might not have been striking enough to move teachers' judgements from a category-based response, and therefore, teachers might have relied on their group-specific stereotypes (for a similar argument, see Wenz and Hoenig, 2020).

Experimental tests of discrimination have their limitations. First, teachers know that their test evaluations and track recommendations do not affect real students. Because of the low-stakes nature of the task, they may behave differently than they do in school. They might be less careful or pay less attention to students' background characteristics because they are told they are supposed to focus on grading. These design elements suggest the possibility of underestimating discrimination.

It is important to note that the external validity of identifying discrimination in an experimental situation is limited. The fact that teachers rely on stereotypes in the experiment does not imply that they discriminate in real-world track recommendations; as in the latter case, they do have further information on their individual students that is missing in the experiment. The results are still relevant, though. First, if teachers have strong prior beliefs on the typical characteristics of certain groups, they might discriminate in track recommendations despite the availability of individual information. Second, the experimental design closely resembles those situations where teachers need to

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

evaluate students they first meet, such as high school admission exams. Therefore, our results might signal discrimination in high school admission decisions where teachers act as gatekeepers and might not admit talented minority students to knowledge-intensive secondary tracks.

Another limitation of our study is that our experimental design did not distinguish between ethnic discrimination and discrimination based on socioeconomic status (Wenz and Hoenig, 2020; Crabtree *et al.*, 2022). Previous research suggests that ethnic minority students are discriminated against not only because of their ethnic minority status, but also because of their low social status (Berényi, Berkovits and Erőss, 2008; Wenz and Hoenig, 2020; Kisfalusi, Janky and Takács, 2021). Our estimate thus captures discrimination against Roma students without disentangling how much of it is due to ethnic discrimination and how much to discrimination based on socioeconomic status. However, due to the very strong correlation between ethnicity and socioeconomic status, Roma students are likely to face the combined effect of these two types of discrimination.

Our findings have important methodological implications for the design of discrimination experiments. Our results show that if participants encounter more than one name, they not only rely on the characteristics of the sole individuals but also consider the composition of all individuals in the experiment. As in our study, these results stem from exploratory analyses; future experimental studies should confirm these findings by conducting a pre-registered analysis. Nevertheless, our results suggest that experimental studies need to take into account that the social context participants infer from the experimental design might influence their discriminatory tendencies. Pre-registered experiments are also needed to confirm our findings on the moderating effect of test quality. Furthermore, future studies could design experiments to explicitly test the theoretical mechanism underlying discrimination.

By providing experimental evidence of discrimination in track recommendations among in-service teachers in Hungary, our study highlighted an important mechanism that can exacerbate educational inequalities. In this regard, our findings suggest that Roma students receive lower track recommendations, even if their performance is identical to that of non-Roma students. Raising teachers' awareness of potential biases and their consequences in student evaluation and track recommendations might reduce discrimination against minority students (Alesina *et al.*, 2024).

Notes

1. In Hungary, teachers are responsible for compiling the tests their students write. Given that most mathematics and grammar teachers only teach in one school, it is very unlikely that students in different schools will write the same tests in a given subject. Based on this institutional feature, it is reasonable to assume that if teachers are provided with different solutions for the same test, they will assume that the tests originate from the same student community.
2. Based on our power calculations using the Optimal Design program (Spybrook et al., 2011), we aimed to have at least 400 respondents in the experiment, each grading 6 tests. With 400 respondents, our design's minimum detectable effect size was 0.11. Since prior research found an effect size of 0.12 comparing Turkish and native German (Sprietsma, 2013; Wenz and Hoenig, 2020), our design was well powered to detect a similar-sized treatment effect among Roma and non-Roma Hungarians.
3. Compared to the pre-registered analysis plan, we use reversed coding, which is more intuitive concerning our substantive research question.
4. We deviate from our pre-registered models as we cluster standard errors at the teacher level (instead of the school level). Based on the paper by Abadie *et al.* (2023), clustering is more appropriate at the teacher level in our case. Our two pre-registered models are also presented with standard errors clustered at the school level in Supplementary Table S4. Clustering of the standard errors at the teacher level instead of the school level does not substantially alter our results.
5. Since the secondary analyses have exploratory purposes, we do not test pre-registered hypotheses. Therefore, we report two-sided *t*-tests in the tables.
6. The share of Roma students in the school stems from administrative data merged with our experimental data. It is estimated by principals in the school survey attached to the National Assessment of Basic Competences, which is a standardized yearly test measuring reading literacy and mathematics skills for the full population of sixth-, eighth-, and tenth-grade students in the country.
7. As the use of the teacher fixed effects holds the test evaluation/track recommendation for non-Roma students as constant across the two experimental conditions (fewer vs. more Roma names), it is not possible to estimate the difference in the test evaluation/track recommendation for non-Roma students by the two experimental conditions. In order to be able to estimate the effect of the share of Roma names separately for both Roma and non-Roma students, we have removed the teacher fixed effects from this estimation.
8. Although we did not have a pre-registered hypothesis on gender discrimination, it should also be noted that teachers tended to discriminate against female students in test evaluations. The finding that teachers discriminated against female students suggests that the absence of discrimination against Roma students in test evaluations is not because the tests we used did not provide enough room for subjective evaluation in teacher assessments.
9. It is important to note that half of the tests were paired with only Roma or only non-Roma names if we split the sample according to the composition of names in the experiment (e.g., test B was paired with only Roma names in the fewer Roma names condition and with only non-Roma names in the more Roma names condition, see Supplementary Table S3). Therefore, this analysis is restricted to the three

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

mathematics and three grammar tests (Tests A, C, and E in Supplementary Table S3) where Roma and non-Roma names occurred in both experimental conditions.

Supplementary data

Supplementary data are available at *ESR* online.

Acknowledgements

We thank Dániel Horn, Hedvig Horváth, Béla Janky, Róbert Károlyi, Gábor Kertesi, Hubert János Kiss, Vera Messing, Károly Takács, participants at the seminar series of the Institute for Analytical Sociology, Linköping University, members of the Economics of Education research group at the Institute of Economics, Centre for Economic and Regional Studies, and the editors and reviewers for their helpful suggestions to earlier drafts of the manuscript.

Author contributions

Dorottya Kisfalusi (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Writing—original draft [equal]), Zoltán Hermann (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Writing—original draft [equal]), and Tamás Keller (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Writing—original draft [equal])

Funding

This work was funded by the National Research, Development and Innovation Office—NKFIH (grant no. K124989, PI: Z.H.). Additional funding was provided by the Institute of Economics, Centre for Economic and Regional Studies. D.K. acknowledges the support of NKFIH grant no. FK137765, and T.K. acknowledges the support of NKFIH grant no. K135766 and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00569/21/9).

Data availability

Our statistical analysis follows the detailed pre-registration we submitted to the RCT Registry of the American Economic Association before beginning the fieldwork: https://doi.org/10.1257/rct.7838-2.0. Any deviations from the original pre-analysis plan are indicated in the paper.

We archived the data, analytic scripts, and an anonymized version of the pre-analysis plan on the project's page on the Open Science Framework: https://osf.io/743df/.

Data and questionnaires have also been reposited at the Research Documentation Centre of the HUN-REN Centre for Social Sciences: DOI: 10.17203/KDK558.

References

Abadie, A. *et al.* (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, **138**, 1–35.

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, **30**, 175–187.

Alesina, A. *et al.* (2024). Revealing stereotypes: evidence from immigrants in schools. *American Economic Review*, **114**, 1916–1948.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A. (Eds.), *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press, pp. 3–33.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: The University of Chicago Press, available from: https://www.press.uchicago.edu/ucp/books/book/chicago/E/bo22415931.html [accessed 14 April 2020].

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing author. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **57**, 289–300.

Berényi, E., Berkovits, B. and Erőss, G. (2008). Iskolarendszer és szabad választás. A jóindulatú szegregációról. In Berényi, E. *et al.* (Eds.), *Iskolarend. Kiváltság És Különbségtétel a Közoktatásban*. Budapest: Gondolat, pp. 15–26.

Billings, S. B., Deming, D. J. and Rockoff, J. (2014). School segregation, educational attainment, and crime: evidence from the end of busing in Charlotte-Mecklenburg. *The Quarterly Journal of Economics*, **129**, 435–476.

Blalock, H. M. (1967). *Toward a Theory of Minority-Group Relations*. New York: Wiley.

Blank, R. M. *et al.* (Eds.) (2004). *Measuring Racial Discrimination*. Washington, DC: The National Academies Press.

Blumer, H. (1958). Race prejudice as a sense of group position. *The Pacific Sociological Review*, **1**, 3–7.

Bohren, J. A. *et al.* (2023). Inaccurate statistical discrimination: an identification problem. *The Review of Economics and Statistics*, 1–45.

Boone, S. and Van Houtte, M. (2013). Why are teacher recommendations at the transition from primary to secondary education socially biased? A mixed-methods research. *British Journal of Sociology of Education*, **34**, 20–38.

Borgna, C. and Contini, D. (2014). Migrant achievement penalties in Western Europe: Do educational systems matter? *European Sociological Review*, **30**, 670–683.

Borgna, C. *et al.* (2022). Old habits die hard? School guidance interventions and the persistence of inequalities. *Research in Social Stratification and Mobility*, **81**, 100728.

Botelho, F., Madeira, R. A. and Rangel, M. A. (2015). Racial discrimination in grading: evidence from Brazil. *American Economic Journal: Applied Economics*, **7**, 37–52.

Bruneau, E. *et al.* (2020). Beyond dislike: blatant dehumanization predicts teacher discrimination. *Group Processes & Intergroup Relations*, **23**, 560–577.

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

Brunello, G. and Rocco, L. (2013). The effect of immigration on the school performance of natives: cross country evidence using PISA test scores. *Economics of Education Review*, **32**, 234–246.

Buchmann, C. and Park, H. (2009). Stratification and the formation of expectations in highly differentiated educational systems. *Research in Social Stratification and Mobility*, **27**, 245–267.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, **31**, 535–576.

Bygren, M. (2020). Biased grades? Changes in grading after a blinding of examinations reform. *Assessment & Evaluation in Higher Education*, **45**, 292–303.

Caro, D. H. *et al*. (2009). The role of academic achievement growth in school track recommendations. *Studies in Educational Evaluation*, **35**, 183–192.

Crabtree, C. *et al*. (2022). Racially distinctive names signal both race/ethnicity and social class. *Sociological Science*, **9**, 454–472.

England, P. and Lewin, P. (1989). Economic and sociological views of discrimination in labor markets: persistence or demise? *Sociological Spectrum*, **9**, 239–257.

Engzell, P. (2019). Aspiration squeeze: the struggle of children to positively selected immigrants. *Sociology of Education*, **92**, 83–103.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In Gilbert, D. *et al*. (Eds.), *The Handbook of Social Psychology*. New York: McGraw-Hill, pp. 357–411.

Fiske, S. T. and Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. In Zanna, M. P. (Ed.), *Advances in Experimental Social Psychology*. Vol. **23**. Amsterdam, NL: Academic Press, pp. 1–74, available from: http://www.sciencedirect.com/science/article/pii/S0065260108603172 [accessed 19 March 2020].

Geven, S., Batruch, A. and van de Werfhorst, H. (2018). *Inequality in Teacher Judgements, Expectations and Track Recommendations: A Review Study*. Amsterdam: Universiteit van Amsterdam, available from: https://www.eumonitor.eu/9353000/1/j4nvgs5kjg27kof_j9vvik7m1c3gyxp/vkv5ilr6qrz7/f=/blg869161.pdf

Glock, S., Kovacs, C. and Pit-ten Cate, I. (2019). Teachers' attitudes towards ethnic minority students: effects of schools' cultural diversity. *The British Journal of Educational Psychology*, **89**, 616–634.

Hajdu, T., Hermann, Z., Horn, D., Hőnich, H. and Varga. J. (2022) *A Közoktatás Indikátorrendszere 2021 [The Indicator System of the Hungarian Public Education 2021]*. Budapest: KRTK KTI, available from: https://kti.krtk.hu/wp-content/uploads/2022/02/A_kozoktatas_indikatorrendszere_2021.pdf [accessed 12 July 2022].

Hajdu, T., Kertesi, G. and Kézdi, G. (2014). Roma fiatalok a középiskolában. Beszámoló a TÁRKI Életpálya-felmérésének 2006 és 2012 közötti hullámaiból [Roma youth in secondary school]. In Kolosi, T. (Ed.), *Társadalmi Riport 2014*. Budapest: TÁRKI, pp. 265–302.

Hajdu, T., Kertesi, G. and Kézdi, G. (2019). Inter-ethnic friendship and hostility between Roma and non-Roma students in Hungary: the role of exposure and academic achievement. *The B.E. Journal of Economic Analysis & Policy*, **19**, 1–17.

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

Havas, G. and Liskó, I. (2005) *Szegregáció a Roma Tanulók Általános Iskolai Oktatásában [Segregation in the Education of Roma Primary School Students]*. Budapest: Felsőoktatási Kutatóintézet, available from: http://www.biztoskezdet.hu/uploads/attachments/havas_lisko_szegregacio_altalanos.pdf [accessed 12 July 2022].

Heath, A. and Brinbaum, Y. (2007). Guest editorial: explaining ethnic inequalities in educational attainment. *Ethnicities*, **7**, 291–304.

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, **12**, 101–116.

Hermann, Z. and Kisfalusi, D. (2023). School segregation, student achievement, and educational attainment in Hungary. *International Journal of Comparative Sociology*, 00207152231198434.

Jonsson, J. O. and Rudolphi, F. (2011). Weak performance--strong determination: school achievement and educational choice among children of immigrants in Sweden. *European Sociological Review*, **27**, 487–508.

Jussim, L. *et al*. (2009). The unbearable accuracy of stereotypes. In Nelson, T. D. (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination*. New York, NY: Psychology Press, pp. 199–227.

Kaiser, J., Südkamp, A. and Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, **109**, 871–888.

Keller, T. (2024). Teachers' perceptions of students' school performance: the impact of classroom composition. Evidence from a survey experiment. *Education Economics*, 1–17.

Keller, T., Takács, K. and Elwert, F. (2022). Yes, you can! Effects of transparent admission standards on high school track choice: a randomized field experiment. *Social Forces*, **101**, 341–368.

Kemény, I. and Janky, B. (2006). Roma population of Hungary 1971–2003. In Kemény, I. (Ed.), *Roma of Hungary. East European Monographs*. New York: Columbia University Press, pp. 70–225.

Kertesi, G. and Kézdi, G. (2009). Roma és nem roma fiatalok középiskolai továbbtanulása. In Fazekas, K. (Ed.), *Oktatás És Foglalkoztatás*. KTI Könyvek 12. Budapest: KTI, pp. 122–136.

Kertesi, G. and Kézdi, G. (2011). The Roma/non-Roma test score gap in Hungary. *American Economic Review*, **101**, 519–525.

Kertesi, G. and Kézdi, G. (2012). Ethnic segregation between Hungarian schools: long-run trends and geographic distribution. *Hungarian Statistical Review*, **90**, 18–45.

Kertesi, G. and Kézdi, G. (2016). On the test score gap between Roma and non-Roma students in Hungary and its potential causes. *Economics of Transition*, **24**, 135–162.

Kisfalusi, D. (2023). Roma students' academic self-assessment and educational aspirations in Hungarian primary schools. *British Journal of Sociology of Education*, **44**, 879–895.

Kisfalusi, D., Janky, B. and Takács, K. (2021). Grading in Hungarian primary schools: mechanisms of ethnic discrimination

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025

against Roma students. *European Sociological Review*, **37**, 899–917.
Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics*, **21**, 447–463.
KSH (2022). *Oktatási Adatok, 2021/2022 (Előzetes Adatok)*, available from: https://www.ksh.hu/docs/hun/xftp/idoszaki/oktat/oktatas2122e/index.html [accessed 4 May 2022].
Lindahl, E. (2007). *Comparing Teachers' Assessments and National Test Results—Evidence from Sweden*. 2007:24, Working Paper Series. Uppsala, Sweden: IFAU—Institute for Evaluation of Labour Market and Education Policy, available from: https://ideas.repec.org/p/hhs/ifauwp/2007_024.html [accessed 9 February 2017].
Lorenz, G. (2021). Subtle discrimination: Do stereotypes among teachers trigger bias in their expectations and widen ethnic achievement gaps? *Social Psychology of Education*, **24**, 537–571.
Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, **79**, 280–295.
McKown, C. and Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology*, **46**, 235–261.
Nicolas, G., de la Fuente, M. and Fiske, S. T. (2017). Mind the overlap in multiple categorization: a review of crossed categorization, intersectionality, and multiracial perception. *Group Processes & Intergroup Relations*, **20**, 621–631.
O'Nions, H. (2016). *Minority Rights Protection in International Law: The Roma of Europe*. London, UK: Routledge.
Ouazad, A. (2008). *Assessed by a Teacher Like Me: Race, Gender and Subjective Evaluations*. London, UK: Centre for the Economics of Education, available from: https://eric.ed.gov/?id=ED530049 [accessed 9 February 2017].
Pager, D. and Shepherd, H. (2008). The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, **34**, 181–209.
Phelps, E. (1972). The statistical theory of racism and sexism. *The American Economic Review*, **62**, 659–661.
Quillian, L. (1995). Prejudice as a response to perceived group threat: population composition and anti-immigrant and racial prejudice in Europe. *American Sociological Review*, **60**, 586–611.
Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology*, **32**, 299–328.
Ready, D. D. and Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: the role of child background and classroom context. *American Educational Research Journal*, **48**, 335–360.
Salikutluk, Z. (2016). Why do immigrant students aim high? Explaining the aspiration–achievement paradox of immigrants in Germany. *European Sociological Review*, **32**, 581–592.
Small, M. L. and Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, **34**, 49–67.
Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics*, **45**, 523–538.
Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., Raudenbush, S., and TO, A. (2011). Optimal design plus empirical evidence: Documentation for the "Optimal Design" software, available from: https://www.stat.cmu.edu/~brian/463-663/week14/od/od-manual-20111016-v300.pdf [accessed 9 February 2020].
Südkamp, A. and Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum. *Zeitschrift für Pädagogische Psychologie*, **23**, 161–174.
Suhajda, C. J. (2017). *A Pályaorientációs Tevékenység Változása És Megvalósulása a Köznevelésben a Rendszerváltozástól Napjainkig, Különös Tekintettel Az Információs Folyamatokra. Doktori Értekezés*. Pécs: Pécsi Tudományegyetem.
Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, **33**, 1–39.
Tajfel, H. and Turner, J. (1979). An integrative theory of intergroup conflict. In Austin, W. G. and Worchel, S. (Eds.), *The Social Psychology of Intergroup Relations*. Monterey: Brooks/Cole, pp. 33–47.
Takács, J. (2018). Személynév—etnikai sztereotípia—előítélet. *Névtani Értesítő*, **40**, 77–89.
Timmermans, A. C. *et al*. (2018). Track recommendation bias: gender, migration background and SES bias over a 20-year period in the Dutch context. *British Educational Research Journal*, **44**, 847–874.
Triventi, M. (2020). Are children of immigrants graded less generously by their teachers than natives, and why? Evidence from student population data in Italy. *International Migration Review*, **54**, 765–795.
Turner, J. C. (1975). Social comparison and social identity: some prospects for intergroup behaviour. *European Journal of Social Psychology*, **5**, 1–34.
van Ewijk, R. (2011). Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review*, **30**, 1045–1058.
Wenz, S. E. and Hoenig, K. (2020). Ethnic and social class discrimination in education: experimental evidence from Germany. *Research in Social Stratification and Mobility*, **65**, 100461.
Zanga, G. and De Gioannis, E. (2023). Discrimination in grading: a scoping review of studies on teachers' discrimination in school. *Studies in Educational Evaluation*, **78**, 101284.

Dorottya Kisfalusi is a senior research fellow at the HUN-REN Centre for Social Sciences in Budapest, Hungary, and the Institute for Analytical Sociology, Linköping University, Sweden. Her current research interests include educational inequalities, discrimination, segregation, and interethnic relations. She has published recently in *Social Networks*, *British Journal of Sociology of Education*, *Group Processes and Intergroup Relations*, and *International Journal of Comparative Sociology*.

Zoltán Hermann is a senior research fellow at the HUN-REN Centre for Economic and Regional Studies, Budapest, and associate professor at the Corvinus University of Budapest. Current research interests comprise education inequalities, segregation, and teacher effects. His work has been published in the journals *Education Economics*, *Research Papers in Education*, *International Journal of Comparative Sociology*, and *Learning and Instruction*.

Tamás Keller, PhD, is a senior researcher at the HUN-REN Centre for Social Sciences in Budapest, Hungary, and also affiliated with the Institute of Economics at the HUN-REN Centre for Economic and Regional Studies. His research focuses on education and social inequality.

Downloaded from https://academic.oup.com/esr/advance-article/doi/10.1093/esr/jcae044/7902057 by Central Library of Budapest University of Economic Sciences user on 06 March 2025