

Grading in Hungarian Primary Schools: Mechanisms of Ethnic Discrimination against Roma Students

Dorottya Kisfalusi ^{1,*}, Béla Janky^{1,2} and Károly Takács^{1,3}

¹Computational Social Science—Research Center for Educational and Network Studies (CSS—RECENS), Centre for Social Sciences, Eötvös Loránd Research Network, Budapest 1097, Hungary, ²Department of Sociology and Communication, Budapest University of Technology and Economics, Budapest 1111, Hungary and ³Institute for Analytical Sociology, Linköping University, Norrköping 601 74, Sweden

*Corresponding author. Email: kisfalusi.dorottya@tk.hu

Submitted August 2019; revised May 2021; accepted June 2021

Abstract

The present study aims to shed light on the mechanisms of ethnic discrimination in teacher assessments in Hungarian primary schools. For this purpose, we use data collected among Roma minority and non-Roma majority students. First, we identify a considerable ethnic difference in non-blind school grades, which is beyond the ethnic difference in blind standardized test scores. Then, we derive and empirically test predictions from different theories of discrimination that might explain the ethnic difference in grades. We find that stereotype-based theories of discrimination do not explain why minority students receive lower grades than majority students. We do not exclude the possibility that taste-based discrimination exists among teachers against Roma students. A considerable part of ethnic discrimination, however, is explained by teachers' indirectly discriminatory grading practices: Roma students' school behaviour is evaluated more negatively by teachers than that of non-Roma students and school behaviour seems to be taken into account in grading without legitimate justification. This practice does not only disadvantage Roma students, but boys and low status students as well.

Introduction

Results from comparative standardized large-scale assessment studies such as the PISA indicate that in many countries, children of some immigrant and ethnic minority groups lag behind majority students considerably with regard to acquired skills, competencies, and abilities (Ammermueller, 2007; Rangvid, 2007; Jonsson and Rudolphi, 2011). Studies spanning various countries, including the United States, UK, Brazil, Germany, Italy, and Sweden, found evidence for the existence of ethnic

differences in non-blind teacher assessments as well, even after controlling for standardized blind test scores. These latter differences were found both in observational (Lindahl, 2007; Ouazad, 2008; Rauschenberg, 2014; Burgess and Greaves, 2013; Kiss, 2013; Botelho, Madeira and Rangel, 2015; Triventi, 2019) and experimental studies (Hanna and Linden, 2010; Sprietsma, 2013; Hinnerich, Höglin and Johannesson, 2015; for a review, see Malouff and Thorsteinnsson, 2016). The ethnic difference in teacher assessments that is above the

test score difference indicates either *discrimination in grading* or unobserved differences between ethnic groups that affect students' school grades but not their test scores.

Our definition of discrimination encompasses both differential treatment, which is also called direct discrimination, and disparate impact, which is also called indirect discrimination. Differential treatment discrimination is the unequal treatment of students on the basis of a social category, when category membership is irrelevant to the original scope of assessment (Pager and Shepherd, 2008; Blank, Dabady and Citro, 2004; Bygren, 2020). Disparate impact discrimination occurs when teachers take into account inadequately justified factors in their evaluations that are not legitimate part of the grading process and, therefore, grading practices favour students from one social group over students from another (Blank, Dabady and Citro, 2004; Pager and Shepherd, 2008). The sociological understanding of discrimination in grading is highly important because of its detrimental consequences on minority students' self-confidence, educational achievement, and labour market opportunities at the micro level and on reproduction of inequalities at the societal level (Becker and Hecken, 2008; Terrier, 2015; Keller, 2016b; Holm, Hjorth-Trolle and Jæger, 2019). The fairness of grading is particularly important in largely stratified educational systems where grades matter for advancement to different tracks at the next school level and hence determine chances of educational attainment and socio-economic status (SES) (Kiss, 2013; Keller, 2016a).

The present study aims to shed light on the mechanisms of ethnic discrimination in grading. Previous studies mainly focused on the extent of ethnic differences in grades. Less attention has been devoted to the systematic analysis of social mechanisms that bring about these differences. The identification of mechanisms underlying discrimination does not only help us understand why minority students receive lower school grades but it is also crucial for the design of effective policy interventions to decrease social inequalities (Reskin, 2003; Blank, Dabady and Citro, 2004; Keuschnigg and Wolbring, 2016; Bozoyan and Wolbring, 2018; Horr, Hunkler and Kroneberg, 2018).

We analyse ethnic discrimination in grading using data collected among Roma minority and non-Roma majority students in Hungarian primary schools. First, we show that Roma students receive lower grades on average than non-Roma students conditional on blind test scores that measure competences. This association is found using a large stratified random sample of primary schools in towns and cities in Hungary. Then, we

analytically distinguish and empirically test predictions from different theories of discrimination using a smaller but more focused study. We investigate if explanations suggested by stereotype-based and taste-based theories of discrimination or explanations focusing on indirectly discriminatory practices of teacher assessments hold in our empirical context. We use a residual approach to differentiate between analytically distinct theoretical predictions. This enables us to gain insights into the underlying social mechanisms of discrimination in grading. As a consequence, the unexplained part of the ethnic gap in grades is largely reduced as we provide evidence for different mechanisms of discrimination against Roma minority students in Hungary.

Mechanisms of Discrimination in Teacher Assessments

Theories of Direct Discrimination 1: Stereotype-Based Theories of Discrimination

Stereotype-based theories suggest that prior beliefs or stereotypes about social categories underlie discrimination. These cognitive elements affect judgements and actions when they are consciously or unconsciously activated during a social interaction (Quillian, 2006).

Social psychological theories assume that impression formation takes place in a continuum from category-based to individuating processes (Fiske and Neuberg, 1990). That is, individuals integrate stereotypic categorization and individuating information in their person perception (Fiske and Neuberg, 1990; Fiske, 1998). Similarly, economic theories of statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977) emphasize the role of prior beliefs about group characteristics in the decision-making process if information about individual characteristics is not sufficiently available for the decision-maker.

Theories of statistical discrimination and models of stereotypic categorization thus suggest that, if teachers hold corresponding prior beliefs or stereotypes, minority students will receive lower grades than they would have received had they been majority students *when they first meet a teacher*, but, following these approaches, this difference should be expected to decrease over time. This is because teachers first make inferences based on their prior beliefs about the social groups students belong to. As teachers acquire more information on individual characteristics, they rely less on group characteristics in their assessments. Some observational and experimental studies support these explanations (Hanna and Linden, 2010; Botelho, Madeira and Rangel, 2015; Wenz and Hoening, 2020).

Status characteristics theory (Berger, Cohen and Zelditch, 1972) offers another explication of how prior beliefs about social categories underlie discrimination in grading. The theory suggests that ethnic and gender categories are diffuse characteristics that carry different status values. Certain states of these categories, e.g. being male or belonging to the majority, are evaluated more positively than others. Due to status generalization, teachers assume that students who belong to higher-valued categories will perform better than students who belong to lower-valued categories (Ridgeway, 1991; Correll and Ridgeway, 2006; Grow, Takács and Pál, 2016). Minority students thus have to provide better performance than majority students to prove the same knowledge and receive the same grades.

Theories of Direct Discrimination 2: Taste-Based Theories of Discrimination

In contrast to theories relying on cognitive factors, taste-based theories of discrimination suggest that individuals inherently prefer to treat certain social groups differentially. The most prominent theoretical framework was provided by Becker (1957) who has assumed that individuals with inherent preferences against a social group are willing to pay a cost to avoid interactions with its members. Becker's theory, however, does not specify where preferences originate from.

Social psychological and sociological theories dig deeper. Social identity theory (Turner, 1975; Tajfel and Turner, 1979; Tajfel, 1982) emphasizes the role of in-group favouritism in discrimination. Individuals aim to identify with positively valued social groups. To achieve this aim, they categorize others along several dimensions and make comparisons between the social categories. To establish a positive distinctiveness from less desired social groups, they evaluate in-group members more positively. Sociological explanations shed light on the factors that foster the formation of out-group prejudice (Blumer, 1958; Blalock, 1967; Quillian, 1995). Theories of perceived group threat emphasize that prejudice underlying discrimination is a response to the perceived threat a dominant social group faces by a subordinate group (Blumer, 1958; Quillian, 1995).

Taste-based theories of discrimination suggest that the cause behind ethnic differences in grades is that teachers who generally belong to the majority ethnic group favour in-group students and disfavour out-group students. In line with this argument, previous empirical studies have shown some evidence that teachers evaluate same-ethnic students more favourably (Ehrenberg, Goldhaber and Brewer, 1995; Ouazad, 2008).

Disparate Impact Discrimination

The above two types of theories assume that teachers treat students with different ethnic background differentially; hence, they are able to explain direct discrimination. Discrimination, however, does not necessarily stem from differential treatment based on students' ethnicity. The formal or informal rules and procedures of teacher evaluations might disparately impact minority and majority students when inadequately justified factors are taken into account in grading. This is also called indirect discrimination (Blank, Dabady and Citro, 2004). The literature on grading practices identifies a potential source of disparate impact discrimination: teachers have been found to take into account students' dispositions, behaviour, and attitudes towards school in academic assessments and this practice negatively affects minority students' grades (Pedulla, Airasian and Madaus, 1980; Bennett et al., 1993). Minority students might have difficulties with conforming to school norms and expectations because their cultural capital including skills, dispositions, and the form of language they use differ from what is accepted by the dominant social group and what is therefore needed to achieve educational success (Bernstein, 1971; Bourdieu and Passeron, 1990; Jæger and Breen, 2016; Jæger and Møllegaard, 2017). As these characteristics are more easily recognizable in everyday school interactions than in a single blind test, minority students are expected to be under-assessed in schools if teachers rely on these characteristics in evaluations.

It is important to note, however, that the perception of students' behaviour and attitudes might be itself subject to cognitive biases and hence, direct discrimination. Teachers have been found to perceive students' behaviour differently depending on whether students belong to the ethnic in-group or an ethnic out-group (Downey and Pribesh, 2004; Dee, 2005). Furthermore, even the same perceived behaviour was sanctioned differently in the case of majority and minority students (Gregory, Skiba and Noguera, 2010).

Identifying Discrimination

In the current paper, we use observational data to study mechanisms of discrimination in grading. We control for standardized blind test scores, which we use as a proxy for students' competences. Test scores are observed by the researchers but unobserved by the teachers at the time of grading. The ethnic difference in grades that is beyond the ethnic difference in standardized blind tests is either the result of discrimination or the result of ethnic differences that influence grades but

not blind test scores (Altonji and Blank, 1999; Blank, Dabady and Citro, 2004). The latter kind of ethnic differences might be present because grades are not assigned based only on competence and acquired knowledge. Teachers take into account several other factors such as students' effort, diligence, motivation, class participation, development, actual performance in assignments and tests, or even out-of-school challenges, and family circumstances (Malouff, 2008; Hardré, 2014). These confounders are observed by the teachers but unobserved by the researchers. Some of these confounders such as effort, diligence, motivation, development, and class participation rightly appear in grades in case teachers use grades to motivate students and to provide feedback for their actual school performance. Hence, different grades for students with the same competence are not discriminatory, in case grades are used to serve the purpose of motivation and feedback. Therefore, the ethnic difference in grades, after controlling for test scores, cannot fully be attributed to discrimination in grading (Blank, Dabady and Citro, 2004; Elwert and Winship, 2014; Bozoyan, and Wolbring, 2015; Wenz, 2020).

At the same time, differences in these factors might themselves be the result of cumulative discrimination in school or other domains (Blank, Dabady and Citro, 2004). If minority students realize that they receive lower grades than majority students for the same achievement, they might lose motivation and lower their efforts, which negatively affect their grades. In this case, we underestimate the effect of discrimination. Discrimination is also underestimated if minority students underperform on standardized tests due to stereotype threat (Steele and Aronson, 1995; Wenz, 2020: p. 161) or if standardized tests are culturally biased and favour majority students over minority students (Phillips, 2006).

Moreover, it is important to note that grades have two separate functions in the Hungarian educational system: they are used to give feedback to students, but in the last two years of primary and secondary school, they are also used in the admission to the next educational level. Whereas taking into account factors such as diligence and effort is justified and therefore does not indicate discrimination if grades are used to provide feedback to students, they are unjustified and therefore indicate discrimination if students with the same knowledge and competences do not receive the same chance to continue their studies in an academic track that is in accordance with their competences.

After identifying the ethnic difference in grades conditional on test scores, we control for students' SES.

Stereotypes, prejudices, and indirectly discriminatory practices of teacher assessments do not only disadvantage ethnic minority groups, they have been shown to play a role in how teachers evaluate low-status students (Bourdieu and Passeron, 1990; Westphal et al., 2016). Since ethnicity and social status are highly correlated, ethnic discrimination in grading is confounded with discrimination based on SES (Wenz and Hoenig, 2020). It is thus important to disentangle how much of the ethnic difference in grading can be attributed to ethnic discrimination and how much to discrimination based on social status.

The applied residual approach has its limitations because it requires strong assumptions to identify discrimination and the exact causal paths between ethnicity and grades (Keuschnigg and Wolbring, 2016; Bozoyan and Wolbring, 2018; Small and Pager, 2020; Wenz, 2020). Our approach, however, is well suited to explore the *social* mechanisms of discrimination. We argue that from different theories of discrimination, different analytical predictions can be derived, and they allow us to shed light on the underlying social mechanisms behind discrimination in grading. The identification of mechanisms helps us in reducing the unexplained part of the ethnic gap in grades.

Identifying the Mechanisms of Discrimination

To reveal the social mechanisms of discrimination in grading, we derive and empirically test predictions from different theories of discrimination. First, theories of statistical discrimination and social psychological models of stereotypic categorization suggest that minority students receive lower grades than majority students conditional on blind competence test scores when they first meet a teacher, but the difference between majority and minority students decreases over time as more information on individual performance is available. In Hungary, specialized teachers of mathematics and literature meet their students first in the 5th grade.¹ In our study, the earliest test score data are from 8th grade in study 1 and 6th grade in study 2. In case we find significant ethnic differences between non-blind teacher assessments and blind test scores towards the end of primary school education, it suggests that statistical discrimination or stereotypic categorization models are not the explanation for why Roma students receive lower grades than non-Roma students, controlling for their test results.

Second, we investigate the possibility that teacher assessments are influenced by status generalization

processes. Status characteristics theory does not only suggest that minority students receive lower assessments for the same performance. It also suggests that girls have to provide better performance than boys to achieve the same grades, at least in STEM (science, technology, engineering, mathematics) subjects, which are considered as stereotypically male domains (Steele, 1997; Spencer, Steele and Quinn, 1999). If status characteristics theory explained the systematic grade differences, we would not only expect Roma students to receive lower grades than non-Roma students, but also that girls receive lower grades than boys conditional on blind test scores, at least in stereotypically male subjects such as mathematics. Therefore, we investigate potential gender differences in grades that are beyond test score differences to test whether status generalization processes occur among teachers.

Then, we examine the possibility that teacher assessments do not only take into account students' academic performance but also actual or perceived behaviour, and this practice has a disparate impact on minority and majority students' grades. In the Hungarian educational system, it is not justified to give lower grades because of students' inappropriate behaviour since this practice is explicitly prohibited by law.² To test the potential presence of this type of indirect discrimination, we include students' conduct mark in the analysis, which indicates how teachers perceive and evaluate students' school behaviour.

We assume that if stereotype-based explanations and disparate impact discrimination do not entirely explain ethnic differences in grading, then minority students are either treated differently because of teachers' tastes or prejudices, or the difference is caused by other differences between ethnic groups that only impact grades but not test results. The identification of differential treatment vs. disparate impact discrimination is helped by our decision to investigate whether teachers' perceptions of students' ethnicity are associated with students' grades over and above students' ethnic self-identification. Existing empirical studies on ethnic discrimination in grading relied on ethnic identification reported by students. In doing so, they implicitly assumed that self-identification represents a consensual classification of ethnicity (Botelho, Madeira and Rangel, 2015). Ethnicity, however, is not an unambiguous concept: people's ethnic self-identification and how their ethnicity is perceived by other people often differ (Telles and Lim, 1998; Saperstein, 2006; Penner and Saperstein, 2015; Saperstein, Kizer and Penner, 2016; Boda, 2019; Kisfalusi, Janky and Takács, 2019; Kisfalusi, Pál and Boda 2020).

The differentiation between self-declared and perceived ethnicity can help the identification of social mechanisms underlying discrimination. We agree with those scholars (e.g. Greiner and Rubin, 2011) who argue that immutable characteristics such as ethnicity can have causal effects because in the case of direct—stereotype-based and taste-based—discrimination, not students' actual group belongings matter but the way their ethnicity is perceived by those who discriminate, such as teachers. Hence, teachers' ethnic perceptions rather than students' self-identifications are assumed to play a significant role in discrimination due to differential treatment (Angrist and Pischke, 2009) because teachers have to perceive students to belong to the minority group in order to be able to discriminate them based on their stereotypes or prejudices. By contrast, ethnic self-identification rather than teachers' ethnic perceptions are assumed to be more relevant in case ethnic differences in grades are caused by indirectly discriminatory grading practices. If minority students receive lower grades because in the grading process, teachers take into account student characteristics in which minority and majority students differ from each other, e.g. dispositions or linguistic style, then actual group belongings are more relevant than teachers' *perceptions* of those belongings.

The Situation of the Roma in the Hungarian Educational System

We investigate differences between grades of Roma and non-Roma Hungarian primary school students. The Roma constitute one of the largest ethnic minorities in Europe and the largest one in Hungary (O'Nions, 2016). In many countries, Roma people experience strong economic and social exclusion (Kertesi and Kézdi, 2011), residential and school segregation (Kemény and Janky, 2006; Kertesi and Kézdi, 2012), and widespread prejudice and discrimination (Váradi, 2014; Brüggemann and D'Arcy, 2017; FRA, 2019; Bruneau et al., 2020). Negative stereotypes about Roma people's cognitive abilities and intellectual skills are widely shared, also among teachers (Bordács, 2001; Ligeti, 2006). These processes may contribute to the persistent educational inequalities (FRA, 2018): Roma students' test scores lag behind those of non-Roma students (Kertesi and Kézdi, 2011; Hajdu, Kertesi and Kézdi, 2019; Keller and Takács, 2019) and they receive lower grades on average in school than majority students (Messing, Neményi and Szalai, 2010; Szalai, 2014; Hajdu, Kertesi and Kézdi, 2019).

Methods

Data

We use two independent datasets from Hungary that contain information on blind test scores and non-blind teacher assessments. Both studies are linked to data of the National Assessment of Basic Competences (NABC). The NABC is a blind test similar to the PISA test conducted to obtain a standardized measurement of reading comprehension and mathematics for all sixth-, eighth-, and tenth-grade students in the country.

Study 1

Data of study 1 are from the Inter-ethnic Friendship and Hostility survey³ conducted in the spring of 2010. The survey used a stratified random sample of 82 primary schools in 75 towns and cities in Hungary ($N = 3,430$ students). Data were collected among eighth-grade students in schools with at least 10 per cent and at most 90 per cent Roma students (for details see [Hajdu, Kertesi and Kézdi, 2019](#)).

Study 2

Data of study 2 are from the third wave of a six-wave panel study conducted among Hungarian primary school students ($N_{T3} = 1,054$ students, 53 classes in 34 schools in 28 settlements). Third-wave data were collected in the autumn of 2014 among sixth-grade students. The third wave of the research was selected because participating students also took part in the NABC in the same academic year. Students with parental permission (96.9 per cent) filled out a self-administered questionnaire under the supervision of trained research assistants. Schools with a high proportion of low-status and Roma students were overrepresented in the sample by design (for more details see [Kisfalusi, 2018](#)). Five schools were located in the capital city, 9 in small- and mid-size towns, and 20 in rural areas in central Hungary. Most schools in the sample are below the national average concerning output and performance measures.

For the purpose of the present analysis, we selected those classes from the sample for which student and teacher questionnaires, grades, and test scores were all available.⁴ Based on these selection criteria, our subsample consists of 33 classes from 23 schools (N students = 687, mean age = 12.0, SD age = 0.7) with a mean class size of 21 students (SD = 5.2). Students with missing data on relevant variables were not included in the regression models.

Measures

Grades (studies 1 and study 2)

In Hungarian schools, students receive summary grades ranging from 1 to 5 (1 = fail, 2 = pass, 3 = satisfactory, 4 = good, 5 = excellent) at the end of each semester. In study 1, grades in mathematics, Hungarian (measured as the mean of literature and grammar grades), and the grade point average (GPA) of all subjects were collected from class records for each student. In study 2, summary grades obtained at the end of the fall semester, after the third wave of data collection, were collected from class records. Moreover, GPA was calculated for every student based on summary grades in five subjects: mathematics, literature, Hungarian grammar, history, and foreign language.⁵ Students' grade in mathematics is used as dependent variable in the main regression models, whereas grade in Hungarian literature and students' GPA is used as dependent variables in robustness checks.

Test scores (studies 1 and 2)

Students' standardized test scores in reading and mathematics were obtained from the NABC. Test scores are neither known publicly, nor are they taken into account in grading⁶ or in the process of secondary school admission. Databases of studies 1 and 2 were linked to data of the nearest NABC.⁷ Test scores are nationally standardized (mean = 0 and SD = 1 in study 1 and mean = 1,500, SD = 200 in study 2).

Self-declared ethnicity (studies 1 and 2)

In study 1, students were asked to report their primary and secondary ethnic identification: 'What is your national or ethnic identity in the first place?'; 'What is your national or ethnic identity in the second place?' Students who identified themselves as Roma on any of the two questions are considered as Roma ([Hajdu, Kertesi and Kézdi, 2019](#)). In study 2, students were asked to classify themselves as 'Hungarian', 'Roma', 'both Hungarian and Roma', or members of 'another ethnicity'. Students who declared to be Roma or both Roma and Hungarian are coded as Roma, students who declared to be Hungarian or members of another ethnicity⁸ are coded as non-Roma.

Teacher perceptions of ethnicity (study 2)

A dummy variable indicates whether the student was classified as Roma by the homeroom teacher.⁹ Note that in most classes, data on grades reflect the assessment of another teacher. Here, we assume that ethnic

perceptions are shared among teachers. Nonetheless, as a robustness check, we repeat the analyses including those classrooms only, in which the homeroom teacher teaches the given subject.

Socio-economic status (study 2)

In study 2, students' socio-economic background was measured with a dummy variable: it indicates whether caretakers of students are entitled to regular child support allowance. The decision of child support allowance by the notary of the competent local government is based on the education level, employment situation, and living conditions of caretakers.

Conduct mark (study 2)

Students' school behaviour is graded on a five-point scale with a 'conduct mark' by the homeroom teacher. The conduct mark depends on whether students behave according to the rules of the school, e.g. whether students have written warnings and reprimands, but does not capture diligence during lessons, for instance. Conduct marks were collected from class records. Conduct marks from the fall semester were not available for six classes; these missing data were imputed with self-reported conduct marks for the same semester from the NABC dataset.

Gender (study 2)

Gender is measured based on students' self-reports. Girls are coded as 1.

Special educational needs (study 2)

A dummy variable controls for students' special educational needs.

Analytical Strategy

Study 1 is used to detect the extent of the ethnic difference in grades after controlling for blind test scores, using a large-scale dataset. The underlying mechanisms are analysed in study 2.

In the main analysis, students' grade obtained in mathematics is the dependent variable, while blind test score in mathematics is controlled for. We use mathematics in our main model because unlike the evaluation in literature and grammar classes, grades in mathematics are mostly based on exercises similar to the ones in blind competence tests (Botelho, Madeira and Rangel, 2015).

Since the dependent variable is ordinal, we estimate multilevel random intercept ordered logit models in our main analysis. Parameter estimates and odds ratios of nested ordered logit models are not directly comparable

with each other (Mood, 2010). Therefore, marginal effects, which are not seriously affected by rescaling bias, are calculated and compared in the different model specifications. For testing the statistical significance of the change between marginal effects of interest across different models, we use the newly developed method suggested by Mize, Doan and Long (2019), which uses seemingly unrelated estimation to combine estimates from multiple models using the `gsem` command in Stata.

Several robustness checks using grades in other subjects as dependent variable and estimating models with classroom fixed effects are available in Supplementary Tables S2–S15. The replication package containing the dataset of study 2 and the syntax files is attached to the article and is available at <https://osf.io/s3u8p/>.

Results

Descriptives

Table 1 presents the relationship of the dependent and independent variables with students' self-declared ethnicity. Consistent with previous findings, self-declared Roma students have significantly lower grades and test scores on average than non-Roma students in both studies. Important gender differences were also found with regard to teacher assessments in study 2. Descriptive statistics of these differences is found in Supplementary Table S1.

Ethnic Difference in Grades

Results of study 1 show that Roma students receive lower grades on average in mathematics than non-Roma students, even after controlling for blind test scores ($\beta = -1.10$, $P < 0.001$; see Table 2). Average marginal effects (AMEs) show that self-declared Roma students are 4.9 percentage points more likely to receive grade 1 ('Fail', $P < 0.001$) and 11.9 percentage points more likely to receive grade 2 ('Pass', $P < 0.001$) than non-Roma students with similar blind test scores. In the meantime, self-declared Roma students are 6.4 percentage points less likely to receive grade 4 ('Good', $P < 0.001$) and 9.3 percentage points less likely to receive grade 5 ('Excellent', $P < 0.001$) than non-Roma students with similar blind test scores. Although the difference for grade 3 ('Satisfactory') is statistically significant, it is not substantial in magnitude (AME = -0.010 , $P = 0.023$).

The associations are similar in study 2 (model 1 in Table 3), even after controlling for special educational needs and gender (model 2 in Table 3). Model 3 (in Table 3) shows that if students' SES is also included in the model, the effect of ethnicity on grade decreases

Table 1. Descriptive statistics about the variables according to ethnicity (self-declared Roma vs. self-declared non-Roma, perceived as Roma vs. perceived as non-Roma) in study 1 and study 2, wave 3

	Self-declared Roma		Self-declared non-Roma		Total		N	
	N = 711		N = 2,719		N = 3,430			
	Mean	SD	Mean	SD	Mean	SD		
GPA	2.94	0.88	3.79	0.84***	3.61	0.90	3,430	
Mathematics grade	2.37	0.93	3.31	1.14***	3.11	1.20	3,404	
Hungarian grade (mean of literature and grammar)	2.81	0.94	3.7	1.03***	3.52	1.10	3,427	
Mathematics test score (mean = 0, SD = 1)	-0.81	0.87	-0.13	0.96***	-0.27	1.00	3,199	
Reading test score (mean = 0, SD = 1)	-0.85	0.89	-0.18	0.95***	-0.31	1.00	3,198	
Study 2								
	Self-declared Roma		Self-declared non-Roma		Perceived as Roma		Perceived as non-Roma	
	N = 191		N = 426		N = 235		N = 452	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
GPA	2.55	0.90	3.63	0.99***	2.49	0.94	3.68	0.93***
Mathematics grade	2.49	1.01	3.53	1.15***	2.47	1.08	3.57	1.12***
Literature grade	2.52	1.07	3.65	1.15***	2.44	1.07	3.72	1.11***
Conduct mark	3.43	0.97	4.26	0.82***	3.34	0.96	4.31	0.78***
Mathematics test score (mean = 1,500, SD = 200)	1,358	166	1,484	181***	1,368	171	1,482	178***
Reading test score (mean = 1,500, SD = 200)	1,299	165	1,480	205***	1,301	165	1,485	205***
Low socio-economic status	81.7%		23.5%	***	83.4%		21.0%	***
Special educational need	7.9%		7.0%		8.1%		6.6%	
							42.4%	
							7.1%	

Notes: Grades, conduct mark, and GPA range from 1 (fail) to 5 (excellenr). Study 1: 181 classes, Study 2: 33 classes.
 $P < 0.1$, $P < 0.05$, $P < 0.01$, $***P < 0.001$.

Table 2. Parameter estimates of multilevel ordered logit regression models predicting grades in mathematics and average marginal effects for self-declared ethnicity (Roma vs. non-Roma) in study 1

	Model 1		
	Estimate	SE	P
Test score: mathematics	1.671	0.083	<0.001
Roma (self-declared)	-1.103	0.126	<0.001
Cutpoint 1	-5.020	0.186	<0.001
Cutpoint 2	-1.572	0.089	<0.001
Cutpoint 3	0.031	0.077	0.689
Cutpoint 4	1.790	0.101	<0.001
Class-level variance	0.770	0.136	
AIC	7,673.122		
BIC	7,715.568		
Average marginal effects for self-declared ethnicity (Roma vs. non-Roma)			
	AME	SE	P
Grade 1	0.049	0.008	<0.001
Grade 2	0.119	0.016	<0.001
Grade 3	-0.010	0.005	0.023
Grade 4	-0.064	0.009	<0.001
Grade 5	-0.093	0.010	<0.001

Notes: Test scores are nationally standardized (mean = 0, SD = 1). Standard errors are clustered around classes. $N = 3,177$ students, 180 classes.

significantly (differences between model 2 and model 3 in AMEs for ethnicity: grade 1: -0.018 , $P = 0.032$, grade 2: -0.025 , $P = 0.025$, grade 3: -0.002 , $P = 0.683$, grade 4: 0.020 , $P = 0.018$, grade 5: 0.024 , $P = 0.018$). However, the ethnic coefficient is still significant ($\beta = -0.98$, $P = 0.004$).

Figures 1 and 2 present marginal effects for ethnicity at different values of test scores calculated based on parameter estimates in model 3 in Table 3. Roma students are more likely than non-Roma students to receive grade 1 throughout almost the entire range of test scores, except in the case of very high test scores where probabilities are similarly low. They are also more likely to receive grade 2 in almost the entire range of test scores, except in the case of very low test scores where they are more likely to receive even worse grades. Meanwhile, Roma students are less likely than non-Roma students to receive grade 5 throughout the entire range of test scores, and they are also less likely to receive grade 4, except in the case of high test scores where non-Roma students are more likely to receive grade 5 (Figure 2).

Parameter estimates in model 3 in Table 3 also show that conditional on test scores, low status students are less likely than higher status students to receive higher grades ($\beta = -0.80$, $P = 0.003$). As an additional analysis, interaction effects between ethnicity and gender and

between ethnicity and SES were also investigated but found to be not significant (those models are not shown in Table 3).

Mechanisms of Ethnic Discrimination in Grading

Theories of statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977) and category-based models of impression formation (Fiske and Neuberg, 1990; Fiske, 1998) propose that teachers rely on their stereotypes until more accurate information on individual performances becomes available. The fact that the significant ethnic differences reported in Tables 2 and 3, model 3, are found among eighth-grade students (study 1) and sixth-grade students (study 2) suggests that these theories are not able to explain the ethnic difference in grades.

As a next step, we investigate whether teacher assessments are influenced by status generalization processes as suggested by status characteristics theory (Berger, Cohen and Zelditch, 1972; Ridgeway, 1991; Correll and Ridgeway, 2006). If status generalization processes underlie grade differences, not only minority students should receive lower assessments for the same performance but also girls, especially in stereotypically male domains such as mathematics. Model 3 in Table 3

Table 3. Parameter estimates of multilevel ordered logit regression models predicting grades in mathematics and average marginal effects for self-declared ethnicity (Roma vs. non-Roma) in study 2, wave 3

	Model 1			Model 2			Model 3			Model 4		
	Estimate	SE	P	Estimate	SE	P	Estimate	SE	P	Estimate	SE	P
Test score: mathematics	0.009	0.001	<0.001	0.009	0.001	<0.001	0.009	0.001	<0.001	0.008	0.001	<0.001
Roma (self-declared)	-1.171	0.307	<0.001	-1.289	0.318	<0.001	-0.982	0.340	0.004	-0.633	0.350	0.070
Girl				0.732	0.204	<0.001	0.759	0.193	<0.001	0.194	0.200	0.331
Special educational need				-0.412	0.432	0.341	-0.420	0.443	0.343	-0.422	0.390	0.279
Low socio-economic status							-0.799	0.272	0.003	-0.510	0.268	0.057
Conduct mark										1.097	0.119	<0.001
Curpoint 1	8.219	1.036	<0.001	8.474	1.019	<0.001	7.906	0.988	<0.001	10.972	1.109	<0.001
Curpoint 2	10.753	1.002	<0.001	11.081	0.985	<0.001	10.545	0.966	<0.001	13.957	1.148	<0.001
Curpoint 3	12.706	1.029	<0.001	13.086	1.002	<0.001	12.584	0.984	<0.001	16.190	1.193	<0.001
Curpoint 4	14.823	1.100	<0.001	15.257	1.092	<0.001	14.775	1.065	<0.001	18.543	1.269	<0.001
Class-level variance	1.663	0.513		1.752	0.523		1.731	0.483		1.528	0.507	
AIC	1,229.573			1,214.529			1,206.756			1,133.588		
BIC	1,259.227			1,252.657			1,249.120			1,180.188		
Average marginal effects for self-declared ethnicity (Roma vs. non-Roma)												
Grade 1	0.060	0.022	0.007	0.065	0.023	0.006	0.047	0.021	0.024	0.026	0.016	0.107
Grade 2	0.090	0.027	0.001	0.097	0.027	<0.001	0.072	0.027	0.007	0.042	0.023	0.070
Grade 3	0.018	0.011	0.106	0.019	0.012	0.104	0.017	0.009	0.048	0.012	0.007	0.087
Grade 4	-0.063	0.021	0.003	-0.068	0.021	0.001	-0.048	0.020	0.017	-0.025	0.016	0.115
Grade 5	-0.105	0.026	<0.001	-0.112	0.026	<0.001	-0.088	0.029	0.002	-0.055	0.029	0.055

Notes: Test scores are nationally standardized (mean = 1,500, SD = 200). Mathematics grade and conduct mark range from 1 (fail) to 5 (excellent). Differences between model 2 and model 3 in AMEs for ethnicity: grade 1: -0.018, $P = 0.032$, grade 2: -0.025, $P = 0.025$, grade 3: -0.020, $P = 0.683$, grade 4: 0.020, $P = 0.018$, grade 5: 0.024, $P = 0.018$. Differences between models 3 and 4 in AMEs for ethnicity: grade 1: -0.021, $P = 0.029$, grade 2: -0.029, $P = 0.041$, grade 3: -0.006, $P = 0.299$, grade 4: 0.023, $P = 0.026$, grade 5: 0.033, $P = 0.025$. $N = 511$ students, 33 classes.

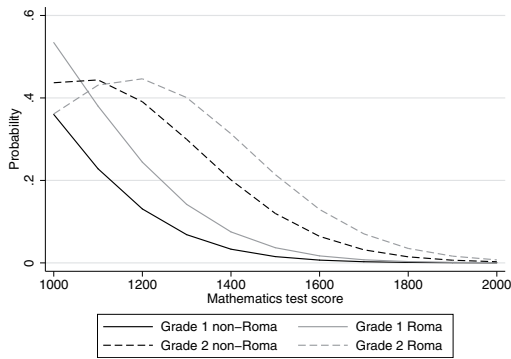


Figure 1. Predicted probabilities of self-declared Roma and non-Roma students for receiving grades 1 and 2 in mathematics at different test scores, controlling for gender, low socio-economic status, and special educational needs in study 2 (calculated based on Table 3, model 3, $N = 511$). Test scores are nationally standardized (mean = 1,500, SD = 200)

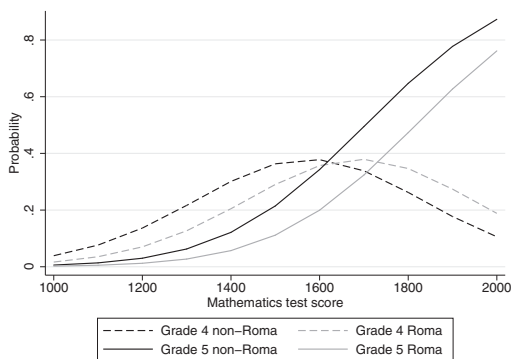


Figure 2. Predicted probabilities of self-declared Roma and non-Roma students for receiving grades 4 and 5 in mathematics at different test scores, controlling for gender, low socio-economic status, and special educational needs in study 2 (calculated based on Table 3, model 3, $N = 511$). Test scores are nationally standardized (mean = 1,500, SD = 200)

shows that conditional on test scores, girls are more likely than boys to receive higher grades in mathematics ($\beta = 0.76$, $P < 0.001$). This contradicts the idea that status generalization processes play a role in teacher assessments.

Then, we investigate a potential source of disparate impact discrimination: we assume that teachers inadequately take into account students' actual or perceived school behaviour when assigning grades. To test this assumption, we include students' conduct mark in the analysis, which indicates how teachers perceive and evaluate students' school behaviour. Model 4 in Table 3 shows that the inclusion of students' conduct mark

significantly decreases the effect of ethnicity on grades compared to model 3 (differences between models 3 and 4 in AMEs for ethnicity: grade 1: -0.021 , $P = 0.029$, grade 2: -0.029 , $P = 0.041$, grade 3: -0.006 , $P = 0.299$, grade 4: 0.023 , $P = 0.026$, grade 5: 0.033 , $P = 0.025$). Even in this model, however, Roma students are less likely than non-Roma students to receive higher grades, although this association is only significant at the 0.1 level ($\beta = -0.63$, $P = 0.070$). Similarly, low status students are less likely than higher status students to receive higher grades ($\beta = -0.51$, $P = 0.057$). In addition, students having higher conduct marks are more likely to receive higher grades than students having lower conduct marks ($\beta = 1.10$, $P < 0.001$). The gender difference in grading, however, is not significant after the inclusion of conduct marks ($\beta = 0.19$, $P = 0.331$).

Our findings are robust to several different model specifications, which can be found in Supplementary Tables S2–S7. In short, all models demonstrate that a large amount of ethnic difference in grades cannot be explained by an ethnic difference in competence measured with blind test scores. While students' SES and classroom behaviour evaluated by the homeroom teacher seem to be important factors behind the ethnic difference in grading, a small ethnic difference still remains in grades in most of our model specifications after accounting for these factors. In the case of GPA, this difference is significant at the 0.05 level, in the case of grade in mathematics, it is only significant at the 0.1 level, whereas in the case of grade in literature, it is not significant in our relatively small sample ($P = 0.144$).

A potential explanation for the remaining ethnic difference in grades can be taste-based discrimination against Roma students. Since we do not have data on teachers' preferences and prejudices, we cannot directly test whether taste-based discrimination occurs in grading. We can test, however, whether teachers' perceptions about their students' ethnicity explain ethnic differences in grades more than students' ethnic self-identification. Taste-based discrimination requires that teachers recognize minority students' group membership and evaluate those students more negatively whom they perceive to belong to the minority group. Therefore, in the last step, we include both students' ethnic self-identification and teachers' ethnic perceptions in the analysis.

There is a strong association between students' self-declared and perceived ethnicity: in most cases, teachers' categorization of students' ethnicity is consistent with students' ethnic self-identification. Still, there are 29 students who declared to be Roma in the third wave but were not perceived as Roma by the teacher, whereas there are 38 students who declared to be non-Roma in

the third wave but were perceived as Roma by the teacher. Due to missing data, only 23 and 27 of these students can be included in the analysis, respectively. This small number of inconsistent ethnic categorization does not allow us to detect statistically significant differences between parameter estimates of self-declared and perceived ethnicity, but some patterns can be observed based on the magnitude of these parameters.

Models including both students' self-identification and teachers' ethnic perceptions are found in [Supplementary Tables S8–S12](#). For an easier and clearer comparison,¹⁰ [Table 4](#) summarizes parameter estimates for self-declared and perceived ethnicity estimated with fixed effects linear models. In the case of mathematics, parameter estimates for self-declared and perceived ethnicity are similar in magnitude in model 1, but in models 2–4, self-declared ethnicity seems to have a larger effect on mathematics grade. In contrast, perceived ethnicity seems to show a larger effect on literature grade and GPA than self-declared ethnicity. Furthermore, in some models (literature: models 1–3, GPA: model 4), perceived ethnicity is significantly associated with grades whereas self-declared ethnicity is not, if both are included.¹¹

Conclusions and Discussion

In the present study, we aimed to shed light on the mechanisms of ethnic discrimination in grading. It is important to note that discrimination does not equal to mere inequality that describes average group differences. Discrimination covers two kinds of unjustified distinctions. First, direct discrimination is unequal treatment

on the basis of a social category. Theories attribute direct discrimination to existing stereotypes or taste-based differences. Second, disparate impact or indirect discrimination is systematic differentiation based on inadequately justified factors that favour members of one social category over members of another category ([Blank, Dabady and Citro, 2004](#); [Pager and Shepherd, 2008](#)).

We investigated the possible mechanisms of discrimination empirically using the situation of Roma students in the Hungarian education system as test case. In line with previous empirical studies from other countries ([Burgess and Greaves, 2013](#); [Kiss, 2013](#); [Botelho, Madeira and Rangel, 2015](#); [Hinnerich, Höglin and Johannesson, 2015](#); [Triventi, 2019](#)), we found evidence for a significant difference in the grades of Roma and non-Roma students conditional on standardized blind test scores and other relevant covariates such as parental SES. This finding is robust using two different studies with considerable differences in sample selection procedure and time of data collection, and to various alternative model specifications.

First, we discussed the possibility that prior beliefs or stereotypes play a role in teacher assessments. Since our data did not allow to investigate grading discrimination when teachers started to teach the students, we cannot rule out the possibility that statistical discrimination ([Phelps, 1972](#); [Arrow, 1973](#); [Aigner and Cain, 1977](#)) or category-based impression formation ([Fiske and Neuberg, 1990](#); [Fiske, 1998](#)) play a role in teacher assessments. The fact that we have found this considerable ethnic difference in grades towards the end of

Table 4. Parameter estimates for self-declared and perceived ethnicity (Roma vs. non-Roma) from fixed-effects linear regression models predicting grades in mathematics, literature, and GPA in study 2, wave 3

	Model 1			Model 2			Model 3			Model 4		
	Estimate	SE	P	Estimate	SE	P	Estimate	SE	P	Estimate	SE	P
Mathematics												
Roma (self-declared)	-0.333	0.144	0.027	-0.374	0.150	0.018	-0.294	0.146	0.052	-0.205	0.133	0.133
Roma (perceived)	-0.335	0.178	0.069	-0.334	0.179	0.071	-0.278	0.157	0.086	-0.089	0.136	0.516
Literature												
Roma (self-declared)	-0.309	0.205	0.142	-0.342	0.204	0.103	-0.250	0.229	0.283	-0.145	0.204	0.483
Roma (perceived)	-0.471	0.174	0.011	-0.463	0.168	0.009	-0.400	0.166	0.021	-0.172	0.151	0.262
GPA												
Roma (self-declared)	-0.267	0.111	0.022	-0.296	0.113	0.013	-0.229	0.123	0.071	-0.144	0.102	0.166
Roma (perceived)	-0.414	0.130	0.003	-0.407	0.126	0.003	-0.360	0.112	0.003	-0.175	0.087	0.054

Notes: Mathematics grade, literature grade, and GPA range from 1 (fail) to 5 (excellent). Model 1 controls for test score, model 2 controls for test score, gender, and special educational needs, model 3 controls for test score, gender, special educational needs, and low socio-economic status, and model 4 controls for test score, gender, special educational needs, low socio-economic status, and conduct mark. Class fixed effects are included. Standard errors are clustered around classes. The whole models can be found in [Supplementary Tables S8, S10, and S12](#). $N = 511$ for mathematics, 513 for literature, 512 for GPA, 33 classes.

primary school education suggests, however, that these are not the explanations for why Roma students receive lower grades than non-Roma students, controlling for test results.

Second, we tested whether status generalization processes underlie discrimination (Berger, Cohen and Zelditch, 1972; Ridgeway, 1991; Correll and Ridgeway, 2006). If this is the case, then we should not only find an ethnic difference in grades, but girls should also perform better than boys to achieve the same grades in stereotypically male domains such as mathematics (Steele, 1997; Spencer, Steele and Quinn, 1999). To the contrary, girls received better mathematics grades than boys with the same blind competence scores. This is in line with the results of previous empirical studies (Lindahl, 2007; Lavy, 2008; Kiss, 2013).

Third, based on the literature on grading practices (Malouff, 2008; Hardré, 2014), we have assumed that teacher assessments also indicate how students conform to school norms and expectations. We have argued that we can draw inference about conformity to school norms from individual conduct marks that indicate how the teacher perceives and evaluates student's school behaviour. The inclusion of conduct mark in the analysis decreased the ethnic difference in grades significantly.

Different causal mechanisms can explain this finding. It is possible that Roma students behave differently in schools, and they do not only receive lower conduct mark because of their behaviour, but they are also punished in their grades. Another possibility is that perception of students' behaviour is itself subject to cognitive biases and discrimination and majority teachers perceive minority students' behaviour less favourably than that of majority students (Downey and Pribesh, 2004; Dee, 2005). Therefore, minority students receive lower conduct mark and they are also punished in their grades because of their differently perceived behaviour. We do not have data on students' actual behaviour, we thus cannot rule out any of these possibilities. In both cases, however, Roma students are indirectly discriminated against in teacher assessments because teachers take into account a factor that is not related to students' competences, and this practise has a disparate impact on the assessment of minority and majority students. In the Hungarian educational system, it is prohibited to give lower grades because of students' inappropriate behaviour. Yet another possible explanation can be that Roma students perform below their ability level in school and receive both lower grades and lower conduct mark because of it. The literature on oppositional culture and the acting white hypothesis suggests that minority students perform below their ability level due to peer

pressure arising from anti-achievement norms (Fordham and Ogbu, 1986; Fryer and Torelli, 2010). Although we have not investigated this possibility in the current analysis, previous analyses on the association between Hungarian Roma students' academic achievement and social relationships using the same samples, we used in this study have shown that Roma students do not exclude or sanction their high-performing Roma peers (Habsz and Radó, 2018; Kisfalusi, 2018; Hajdu, Kertesi and Kézdi, 2019). This suggests that there is no peer pressure on Roma students to camouflage their efforts and competencies.

Since an ethnic difference in grades has still remained after accounting for students' conduct mark, we could not rule out the possibility that there is taste-based discrimination in teacher assessments against Roma students. The fact that teachers' ethnic perceptions seem to play a more important role in grading than students' ethnic self-identification, at least in the case of literature and GPA, also supports the possibility that teacher assessments are influenced by teachers' ethnic preferences and prejudices.

Our study is not without limitations. To study discrimination, we analysed observational data and followed a residual analytical approach to identify social mechanisms behind differential treatment and disparate impact discrimination. The residual approach has well-known weaknesses as it is vulnerable to the consequences of unobserved heterogeneity (Cancio, Evans and Maume, 1996; Small and Pager, 2020) as well as to the possibility of overcontrol (Pager and Shepherd, 2008; Keuschnigg and Wolbring, 2016; Neil and Winship, 2019; Wenz, 2020). Therefore, we were cautious in making causal interpretations, and we emphasized that our control variables might have been endogenous to the process of discrimination. Furthermore, it should be noted that study 2, which was used to investigate the mechanisms of discrimination, is not representative, because the sample intentionally overrepresented schools with a higher share of Roma students.

Some scholars highlight that it is difficult to clearly distinguish disparate impact discrimination from ethnic inequality (see e.g. Wenz, 2020: p. 67). Disparate impact discrimination occurs in grading when teachers take into account factors in their evaluations that are not legitimate parts of the grading process (Blank, Dabady and Citro, 2004). Grading systems may differ in what factors are allowed to be used in grading. Taking into account diligence and effort, for instance, is justified in systems where grades are used to provide feedback to students, but they are unjustified where students with the same knowledge and competences are to receive the same

chances to continue their studies in the same academic track. Taking into account student school behaviour in the grading process is clearly unjustified in the Hungarian educational context as this practice is explicitly prohibited by law. Taking this into consideration, our conclusions on discrimination hold only if the conduct mark does not exclusively correlate with that part of student ability that is not reflected in blind standardized test but is observed by teachers.

Another caveat concerning the measurement of perceived ethnicity is that whereas homeroom teachers were asked to classify students as Roma or non-Roma, students' performance was evaluated by the teacher who taught the given subject. For the majority of students in the sample, these two people were different. Although we assume that teachers are likely to share an opinion about who is Roma in a given class as social ties influence ethnic perceptions (Boda, 2018), it is of course possible that the mathematics or the literature teacher has a different perception about a student's ethnicity than the homeroom teacher. It is important to emphasize, however, that even with this way of measurement we found that students who are perceived as Roma by the homeroom teacher receive lower grades in the examined subjects. In the case of literature grade and GPA, moreover, perceived ethnicity had a larger effect on grades than self-declared ethnicity, although this difference was statistically not significant, probably due to low statistical power. In addition, teachers might not only influence ethnic perceptions but also the direct stereotype- or taste-based discrimination tendencies of each other (Takács, Bravo and Squazzoni, 2018).

Despite these limitations, our study is a significant contribution in multiple ways. We have demonstrated how to test for the presence of theoretical mechanisms that are able to explain discrimination in grading. As an innovative contribution, we have attempted to disentangle different mechanisms by looking at students' self-declared ethnicity and also at teachers' perception of students' ethnicity. These contributions go beyond the analysis we conducted for studying ethnic discrimination in grading in Hungary. The results suggest that Roma students are not the only social group that is discriminated against in teacher assessments: boys and low status students also receive lower grades than girls and higher status students with similar competences. The findings indicate that a considerable part of discrimination is caused by the disparate impact of indirectly discriminatory grading practices: despite the legislation, teacher assessments do not only take into account competencies but also how students conform to school norms and expectations. In the case of Roma students

and low status students, furthermore, we did not rule out the possibility that taste-based discrimination also exists. Discrimination in grading thus further reinforces already existing educational inequalities.

Notes

- 1 In the Hungarian educational system, the 8-year-long primary school education consists of two 4-year-long phases. In the first 4 years, the homeroom teacher teaches most subjects. From the fifth grade on, every subject is taught by specialized teachers. The teachers usually remain the same for the entire 4-year-long period.
- 2 See the Act CXC of 2011 on National Education 54.§(1).
- 3 The Inter-ethnic Friendship and Hostility (IEFH) survey was designed and supervised by Gábor Kertesi and Gábor Kézdi and was financed by the Educatio Kht, Hungary. The data were collected by the Adatgyűjtő Intézet, Hungary, in the spring of 2010. Hajdu, Kertesi, and Kézdi (2019) provide a detailed description of the study.
- 4 Teacher questionnaires are missing in the case of 8 classes, test scores are missing in the case of 14 classes, and grades are missing in the case of 4 classes.
- 5 Subjects in natural sciences such as physics, biology, and chemistry are incorporated in the curriculum at later grades.
- 6 The evaluation of the test is organized centrally by the Educational Authority. Therefore, teachers do not know the student whose test they correct. The test results are published several months later. Individual test scores can be accessed with an individual identification code. This is obtained for research purposes only if parental consent is received.
- 7 In study 2, grades obtained at the end of the semester when the 2015 NABC test was written are not available in the dataset. Hence, grades were measured earlier in time than test scores. We argue that this is not a problem in the case of our analysis for two reasons. First, we do not intend to identify a causal relationship between test scores and grades but argue that these two assessments are aimed to measure different elements of the same concept: students' academic achievement. We use test scores as a proxy for competence, which is assumed to be non-varying under a short time period. In the meantime, we acknowledge that grades do not only rely on competence but take into account other factors affecting academic achievement as well. Our

aim is to examine whether there are systematic ethnic differences in the process how these factors are translated into grades. Second, summary grades at the end of a semester summarize grades obtained earlier during the school year. In the Hungarian educational system, summary grades at the end of the fall semester rely on grades obtained between September and January, whereas summary grades at the end of the spring semester rely on grades obtained during the entire academic year between September and June, thus, also include grades from the fall semester. Therefore, the correlation between summary grades of the fall and the spring semester is usually very high. In the sample of study 2, for instance, the correlation coefficient between students' GPAs in the fall and spring semester was 0.948 ($P < 0.001$), whereas the correlation coefficient between students' mathematics grades in the fall and spring semester was 0.877 ($P < 0.001$) when students attended the fifth grade.

- 8 Two students declared to belong to another ethnicity: Russian and Finnish.
- 9 In the Hungarian educational system, the homeroom teacher is responsible for a class size group of students—designated to the same homeroom—in terms of administrative tasks, keeping contact with parents and other teachers, organizing out-of-school activities for students, and conducting discussion classes. Usually, the homeroom teacher also teaches one or more subjects to the class.
- 10 Since log odds/odds ratios cannot be directly compared (Mood, 2010), we could only compare marginal effects of the ordered logit models. But because our dependent variable is an ordered variable with five categories, five separate average marginal effects belong to both self-declared and perceived ethnicity, which makes the parameter comparison much less straightforward in the case of the ordered logit models.
- 11 The results are robust to the exclusion of classes where the person who reports students' perceived ethnicity (the homeroom teacher) is not the same who teaches the given subject and awards grades (mathematics: [Supplementary Table S9](#), $N = 100$ students, 6 classes; literature: [Supplementary Table S11](#), $N = 179$, 10 classes).

Supplementary Data

[Supplementary data](#) are available at *ESR* online.

Acknowledgements

We thank Gábor Kertesi, Gábor Kézdi, and Tamás Hajdu for providing us access to the database of the Inter-ethnic Friendship and Hostility survey. We are grateful to the four anonymous reviewers and the editors for their helpful comments and suggestions. The article is dedicated to the memory of Gábor Kézdi (1971–2021), a pioneer of quantitative research on discrimination against the Roma in the public education and the labour market.

Funding

This work was supported by the project 'From Talent to Young Researcher project aimed at activities supporting the research career model in higher education', identifier EFOP-3.6.3-VEKOP-16-2017-00007 co-supported by the European Union, Hungary, and the European Social Fund to D.K., the ÚNKP-16-4 New National Excellence Program of the Ministry of Human Capacities to D.K., the Lendület program of the Hungarian Academy of Sciences (Project title: 'Competition and Negative Networks' to K.T.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648693 to K.T.).

References

- Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, 30, 175–187.
- Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. In Ashenfelter, O. C. and Card D. (Eds): *Handbook of Labor Economics*, Vol. 3. Amsterdam: Elsevier, pp. 3143–3259.
- Ammermueller, A. (2007). Poor background or low returns? Why immigrant students in Germany perform so poorly in the programme for international student assessment. *Education Economics*, 15, 215–230.
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, Oxford: Princeton University Press.
- Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter O. and Rees A. (Eds.), *Discrimination in Labor Markets*. Princeton, New Jersey: Princeton University Press, pp. 3–33.
- Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: The University of Chicago Press.
- Becker, R. and Hecken, A. E. (2008). Why are working-class children diverted from universities?—An empirical assessment of the diversion thesis. *European Sociological Review*, 25, 233–250.
- Bennett, R. E. et al. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85, 347–356.
- Berger, J., Cohen, B. P. and Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 37, 241.
- Bernstein, B. (1971). *Class, Codes and Control. Volume 1: Theoretical Studies towards a Sociology of Language*. London and New York: Routledge and Kegan Paul.

- Blalock, H. M. (1967). *Toward a Theory of Minority-Group Relations*. New York: Wiley.
- Blank, R. M., Dabady, M., and Citro, C. F. (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.
- Boda, Z. (2018). Social influence on observed race. *Sociological Science*, 5, 29–57.
- Boda, Z. (2019). Friendship bias in ethnic categorization. *European Sociological Review*, 35, 567–581.
- Bordács, M. (2001). A pedagógusok előítéletességének vizsgálata roma gyerekeket is tanító pedagógusok körében [Study of prejudices among teachers of Roma children]. *Új Pedagógiai Szemle*, 2, 70–89.
- 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.
- Bourdieu, P. and Passeron, J.-C. (1990). *Reproduction in Education, Society and Culture*. London and Beverly Hills: SAGE.
- Bozoyan, C. and Wolbring, T. (2015). The usefulness of directed acyclic graphs: what can DAGs contribute to a residual approach to weight-related income discrimination? *Schmollers Jahrbuch*, 135, 83–96.
- Bozoyan, C. and Wolbring, T. (2018). The weight wage penalty: a mechanism approach to discrimination. *European Sociological Review*, 34, 254–267.
- Brüggemann, C. and D'Arcy, K. (2017). Contexts that discriminate: international perspectives on the education of Roma students. *Race Ethnicity and Education*, 20, 575–578.
- Bruneau, E. et al. (2020). Beyond dislike: blatant dehumanization predicts teacher discrimination. *Group Processes and Intergroup Relations*, 23, 560–577.
- 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 and Evaluation in Higher Education*, 45, 292–303.
- Cancio, A. S., Evans, T. D. and Maume, D. J. (1996). Reconsidering the declining significance of race: racial differences in early career wages. *American Sociological Review*, 61, 541.
- Correll, S. J. and Ridgeway, C. L. (2006). Expectation states theory. In Delamater J. (Ed.), *Handbook of Social Psychology*. Boston, MA: Springer US, pp. 29–51.
- Dee, T. S. (2005). A teacher like me: does race, ethnicity, or gender matter? *The American Economic Review*, 95, 158–165.
- Downey, D. B. and Pribesh, S. (2004). When race matters: teachers' evaluations of students' classroom behavior. *Sociology of Education*, 77, 267–282.
- Ehrenberg, R. G., Goldhaber, D. D. and Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from NELS. *Industrial and Labor Relations Review*, 48, 547–561.
- Elwert, F. and Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, 40, 31–53.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In Gilbert D., Fiske S. T., and Lindzey G. (Eds.), *The Handbook of Social Psychology*. Boston: 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: Academic Press, pp. 1–74.
- Fordham, S. and Ogbu, J., U. (1986). Black students' school success: coping with the “burden of ‘acting white’”. *The Urban Review*, 18, 176–206.
- FRA. (2018). *Transition from Education to Employment of Young Roma in Nine EU Member States*. Luxembourg: European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-eu-midis-ii-roma-transition-education-employment_en.pdf. Last accessed: 10 July 2021.
- FRA. (2019). *Fundamental Rights Report 2019*. Luxembourg: European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-fundamental-rights-report-2019_en.pdf. Last accessed: 10 July 2021.
- Fryer, R. G., Jr. and Torelli, P. (2010). An empirical analysis of ‘acting white’. *Journal of Public Economics*, 94, 380–396.
- Gregory, A., Skiba, R. J. and Noguera, P. A. (2010). The achievement gap and the discipline gap: two sides of the same coin? *Educational Researcher*, 39, 59–68.
- Greiner, D. J. and Rubin, D. B. (2011). Causal effects of perceived immutable characteristics. *Review of Economics and Statistics*, 93, 775–785.
- Grow, A., Takács, K. and Pál, J. (2016). Status characteristics and ability attributions in hungarian school classes: an exponential random graph approach. *Social Psychology Quarterly*, 79, 156–167.
- Habsz, L. D. and Radó, M. (2018). Ez egy gyönyörű barátság kezdete? Iskolai ellenkultúra hiánya a magyarországi roma tanulók körében. *Szociológiai Szemle*, 28, 52–74.
- 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 and Policy*, 19, 1–17.
- Hanna, R. and Linden, L. (2010). *Measuring Discrimination in Education*. Working Papers. eSocialSciences. <https://ideas.repec.org/p/ess/wpaper/2541.html>. Last accessed: 10 July 2021.
- Hardré, P. L. (2014). Checked your bias lately? Reasons and strategies for rural teachers to self-assess for grading bias. *The Rural Educator*, 35. <https://core.ac.uk/reader/215221120>. Last accessed: 10 July 2021.
- Hinnerich, B. T., Höglin, E. and Johannesson, M. (2015). Discrimination against students with foreign backgrounds: evidence from grading in Swedish public high schools. *Education Economics*, 23, 660–676.

- Holm, A., Hjorth-Trolle, A. and Jæger, M. M. (2019). Signals, educational decision-making, and inequality. *European Sociological Review*, *35*, 447–460.
- Horr, A., Hunkler, C. and Kroneberg, C. (2018). Ethnic discrimination in the German housing market. *Zeitschrift für Soziologie*, *47*, 134–146.
- Jæger, M. M. and Breen, R. (2016). A dynamic model of cultural reproduction. *American Journal of Sociology*, *121*, 1079–1115.
- Jæger, M. M. and Møllegaard, S. (2017). Cultural capital, teacher bias, and educational success: new evidence from monozygotic twins. *Social Science Research*, *65*, 130–144.
- Jonsson, J. O. and Rudolph, F. (2011). Weak performance–strong determination: school achievement and educational choice among children of immigrants in Sweden. *European Sociological Review*, *27*, 487–508.
- Keller, T. (2016a). If grades are not good enough—the role of self-assessment in the transition to tertiary education. *International Journal of Educational Research*, *77*, 62–73.
- Keller, T. (2016b). Sticky assessments—the impact of teachers’ grading standard on pupils’ school performance. *Educational Studies*, *42*, 493–518.
- Keller, T. and Takács, K. (2019). Peers that count: the influence of deskmates on test scores. *Research in Social Stratification and Mobility*, *62*, 100408.
- Kemény, I. and Janky, B. (2006). Roma population of Hungary 1971–2003. In Kemény I. (Ed.), *Roma of Hungary*. Columbia University Press, pp. 70–225.
- 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.
- Keuschnigg, M. and Wolbring, T. (2016). The use of field experiments to study mechanisms of discrimination. *Analyse and Kritik*, *38*, 179–202.
- Kisfalusi, D. (2018). Bullies and victims in primary schools. The associations between bullying, victimization, and students’ ethnicity and academic achievement. *Intersections*, *4*, 133–158.
- Kisfalusi, D., Janky, B. and Takács, K. (2019). Double standards or social identity? The role of gender and ethnicity in ability perceptions in the classroom. *The Journal of Early Adolescence*, *39*, 745–780.
- Kisfalusi, D., Pál, J. and Boda, Z. (2020). Bullying and victimization among majority and minority students: the role of peers’ ethnic perceptions. *Social Networks*, *60*, 48–60.
- Kiss, D. (2013). Are immigrants and girls graded worse? Results of a matching approach. *Education Economics*, *21*, 447–463.
- Lavy, V. (2008). Do gender stereotypes reduce girls’ or boys’ human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, *92*, 2083–2105.
- Ligeti, G. (2006). Sztereotípiák és előítéletek [Stereotypes and prejudices]. In Kolosi T., Tóth I. G., and Vukovich G. (Eds.), *Társadalmi Ríport 2006*. TÁRKI, pp. 373–389.
- Lindahl, E. (2007). *Comparing Teachers’ Assessments and National Test Results—Evidence from Sweden*. Working Paper Series 2007:24. IFAU—Institute for Evaluation of Labour Market and Education Policy. https://ideas.repec.org/p/hhs/ifauwp/2007_024.html. Last accessed: 10 July 2021.
- Malouff, J. M. (2008). Bias in grading. *College Teaching*, *56*, 191–192.
- Malouff, J. M. and Thorsteinsson, E., B. (2016). Bias in grading: a meta-analysis of experimental research findings. *Australian Journal of Education*, *60*, 245–256.
- Messing, V., Neményi, M. and Szalai, J. (2010). *Ethnic Differences in Education in Hungary: Survey Report*. Budapest: Central European University, Center for Policy Studies. http://www.edumigrom.eu/sites/default/files/field_attachment/page/node-5387/community-reporthungary.pdf. Last accessed: 10 July 2021.
- Mize, T. D., Doan, L. and Long, J. S. (2019). A general framework for comparing predictions and marginal effects across models. *Sociological Methodology*, *49*, 152–189.
- Mood, C. (2010). Logistic regression: why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*, 67–82.
- Neil, R. and Winship, C. (2019). Methodological challenges and opportunities in testing for racial discrimination in policing. *Annual Review of Criminology*, *2*, 73–98.
- O’Nions, H. (2016). *Minority Rights Protection in International Law: The Roma of Europe*. London: Routledge.
- Ouazad, A. (2008). *Assessed by a Teacher like Me: Race, Gender and Subjective Evaluations*. INSEAD Working Paper No. 2008/57/EPS, Available at SSRN: <https://ssrn.com/abstract=1267109>. Last accessed: 10 July 2021.
- 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.
- Pedulla, J. J., Airasian, P. W. and Madaus, G. F. (1980). Do teacher ratings and standardized test results of students yield the same information? *American Educational Research Journal*, *17*, 303–307.
- Penner, A. M. and Saperstein, A. (2015). Disentangling the effects of racial self-identification and classification by others: the case of arrest. *Demography*, *52*, 1017–1024.
- Phelps, E. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*, 659–661.
- Phillips, M. (2006). Standardized tests aren’t like T-shirts: one size doesn’t fit all. *Multicultural Education*, *14*, 52–55.
- 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. JSTOR.
- Rangvid, B. S. (2007). Sources of immigrants’ underachievement: results from PISA—Copenhagen. *Education Economics*, *15*, 293–326.

- Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives*, 22, 1-38.
- Reskin, B. F. (2003). Including mechanisms in our models of ascriptive inequality: 2002 presidential address. *American Sociological Review*, 68, 1.
- Ridgeway, C. (1991). The social construction of status value: gender and other nominal characteristics. *Social Forces*, 70, 367.
- Saperstein, A. (2006). Double-checking the race box: examining inconsistency between survey measures of observed and self-reported race. *Social Forces*, 85, 57-74.
- Saperstein, A., Kizer, J. M. and Penner, A. M. (2016). Making the most of multiple measures: disentangling the effects of different dimensions of race in survey research. *American Behavioral Scientist*, 60, 519-537.
- Small, M. L. and Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34, 49-67.
- Spencer, S. J., Steele, C. M. and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics*, 45, 523-538.
- Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Szalai, J. (2014). The emerging 'ethnic ceiling': implications of grading on adolescents' educational advancement in comparative perspective. In Szalai J. and Schiff C. (Eds.), *Migrant, Roma and Post-Colonial Youth in Education across Europe*. Basingstoke, Hampshire: Palgrave Macmillan UK, pp. 67-83.
- 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, California: Brooks/Cole Publishing, pp. 33-47.
- Takács, K., Bravo, G. and Squazzoni, F. (2018). Referrals and information flow in networks increase discrimination: a laboratory experiment. *Social Networks*, 54, 254-265.
- Telles, E. E. and Lim, N. (1998). Does it matter who answers the race question? Racial classification and income inequality in Brazil. *Demography*, 35, 465.
- Terrier, C. (2015). *Giving a Little Help to Girls? Evidence on Grade Discrimination and Its Effect on Students' Achievement*. CEP Discussion Paper No. 1341. Centre for Economic Performance. <https://ideas.repec.org/p/cep/cepdps/dp1341.html>. Last accessed: 10 July 2021.
- Triventi, M. (2019). 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.
- Váradi, L. (2014). *Youths Trapped in Prejudice*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Wenz, S. E. (2020). *Discrimination in Education: Methodology, Theory, and Empirics of Teachers' Stereotypes, Prejudice, and Discriminatory Behavior*, Vol. 26. GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssaoar.67307>. Last accessed: 10 July 2021.
- 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.
- Westphal, A. et al. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: the role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology*, 46, 218-227.

Dorottya Kisfalusi is a research fellow at the Centre for Social Sciences, Computational Social Science—Research Center for Educational and Network Studies, Budapest, Hungary. Her research interests include sociology of education, adolescents' peer relationships, interethnic relations, discrimination, and ethnic classification processes. She has published recently in *Social Networks*, *Network Science*, *Group Processes and Intergroup Relations*, and the *Journal of Early Adolescence*.

Béla Janky is an associate professor at the Budapest University of Technology and Economics, and a senior research fellow at the Centre for Social

Sciences, Computational Social Science—Research Center for Educational and Network Studies, Budapest. His current research interests are in rational decision-making, framing, stereotypes, and attribution processes. He has published in *Economics Letters*, *Sociological Research Online*, *Journal of Early Adolescence*, and *International Journal of Public Opinion Research* recently.

Károly Takács is a senior lecturer at the Institute for Analytical Sociology, Linköping University and a senior research fellow at the Centre for Social Sciences, Computational Social Science—Research Center for Educational and Network Studies,

Budapest. He received his Ph.D. from the University of Groningen/ICS. His main research interests are the theoretical, experimental, and empirical analysis of the dynamics of social networks and status relations, in relation to problems of cooperation and

conflict. He published recently in *Social Networks*, *Research in Social Stratification and Mobility*, *Journal of Early Adolescence*, *Social Psychology Quarterly*, *Journal of Research on Adolescence*, *PLOS One*, *Scientometrics*, and *Scientific Reports*.