sciendo
Swiss Journal of Sociology, 48 (1), 2022, 77–105

Discrimination at the Crossroads? Evidence from a Factorial Survey Experiment on Teacher's Tracking Decisions

Sandra Gilgen* and Milan Stocker*

Abstract: Using a factorial survey experiment, we investigate whether teachers discriminate against lower class and minority children. The vignettes include information on the pupil's gender, social and ethnic origin, academic potential, motivation, behaviour in class as well as parental educational aspirations. While the discrimination hypotheses are not supported, we find gender-specific effects of challenging behaviour, to the disadvantage of girls. Both the usefulness as well as potential pitfalls of the factorial survey approach are discussed.

Keywords: Discrimination, educational inequalities, factorial survey experiment, tracking decisions, gender bias

Discrimination lors du passage à l'école secondaire ? Constats d'une enquête factorielle expérimentale sur les décisions de recommandation des enseignant·e·s

Résumé : Utilisant une enquête factorielle expérimentale – information sur le genre, l'origine sociale et ethnique, le potentiel scolaire, la motivation et le comportement des élèves ainsi que les aspirations des parents – les recommandations des enseignant·e·s ont été analysées. Si les hypothèses de discrimination selon la classe sociale et l'origine ethnique ne sont pas confirmées, on trouve des effets spécifiques au genre d'un comportement difficile. Cet article illustre l'utilité et les potentiels problèmes de l'approche expérimentale.

Mots-clés : Discrimination, inégalités scolaires, enquête factorielle expérimentale, recommandation de transition, biais de genre

Diskriminierung am Übergang in die Sekundarschule? Erkenntnisse aus einem faktoriellen Survey Experiment zu den Übertrittsempfehlungen von Lehrpersonen

Zusammenfassung: Anhand eines faktoriellen Survey Experiments – Geschlecht, die soziale und ethnische Herkunft, das Leistungspotential, die Motivation und das Verhalten der Schüler*innen sowie die Bildungsaspiration der Eltern – wurden die Empfehlungen von Lehrpersonen untersucht. Während die Diskriminierungshypothesen nach sozialer und ethnischer Herkunft nicht bestätigt werden, finden wir geschlechtsspezifische Effekte von schwierigem Verhalten zu Ungunsten von Mädchen. Methodisch werden sowohl die Vorteile wie potentiellen Nachteile des Ansatzes diskutiert.

Schlüsselwörter: Diskriminierung, Bildungsungleichheit, faktorielles Survey Experiment, Übertrittsempfehlung, Gender Bias

* University of Bern, Department of Sociology of Education, CH-3012 Bern, sandra.gilgen@edu.unibe.ch, milan.stocker@edufr.ch

Seismo
DOI 10.2478/sjs-2022-0005

1 Introduction[1]

It has repeatedly and consistently been shown that lower class (Boudon 1974; Erikson and Jonsson 1996; Goldthorpe 2003; Becker 2011) and minority children (Becker and Beck 2012; Beck 2015; Diehl et al. 2016) end up with lower educational degrees compared to their peers with more privileged and non-minority backgrounds. These educational differentials translate into differences in overall life chances. In societies that subscribe to a meritocratic ideal, education serves as a legitimating factor for unequal access to prestigious jobs and as such to social status (Bourdieu and Passeron 1970; Jencks et al. 1972; Bowles and Gintis 1976; Meyer 1977; Hadjar 2008; Becker and Lauterbach 2013). People with a lower educational background face more difficulties on the job market, leading to various adversities such as unemployment or poverty (Quenzel and Hurrelmann 2010; Kristen et al. 2011; Duncan and Murnane 2011). These job-related adversities can also lead to social exclusion and precarious life situations of all kinds; problems that are overall much more likely to afflict those holding low educational degrees (Tsakloglou and Papadopoulos 2002; Ditton 2010, 67).

While the use of rational choice theory has been very widespread and successful in explaining the disadvantage of lower class children in particular (Boudon 1974; Erikson and Jonsson 1996, 199; Becker and Hecken 2007), the individual-deficit approach has failed to find evidence that the disadvantage of minority and lower-class students can be explained by their deficits (McKay and Devlin 2016). In contrast, discrimination theories start with the opposite premise, seeking the reasons for inequalities in the treatment of the less privileged by those in power and the better off.

While we know a lot about how inequality of educational opportunity is mediated by primary and secondary effects (Boudon 1974), we know very little about the role of teachers in this regard. Furthermore, while research has shown that especially early streaming consolidates educational inequalities (Müller and Karle 1993; Hanushek and Wössmann 2006; Pfeffer 2008; Blossfeld et al. 2016), the question remains whether teachers aggravate this further by discriminating by ethnic origin and social class. Relying on the theory of statistical discrimination as a theoretical backbone, this paper investigates whether teachers (unintentionally) promote educational inequalities by giving biased recommendations.

As a means of circumventing the numerous challenges of research on discrimination, a factorial survey experiment was embedded in an online survey of 5th and 6th grade teachers in the canton of Bern, Switzerland. In the vignette experiment, teachers were asked to rate the probability of recommending the children described in the vignettes for a higher versus a lower secondary school stream. Since, in the

1 Many thanks to the editors as well as anonymous reviewers for their many invaluable inputs as well as to Katrin Auspurg and Ben Jann for their support and guidance. Rolf Becker, Katrin Auspurg and Ben Jann.

canton of Bern, the transition decision is made in the middle of 6th grade and the teacher's recommendation is binding in most schools, the factorial survey experiment covers a critical point in a student's trajectory. Among the *D*-efficiently varied characteristics of the children, information on their social class (upper class, middle class and lower class) and minority status (Swiss name versus Slavic name) was included.

Additionally, based on previous research findings on stereotypical beliefs about boys and girls (Munsch et al. 2007; Mullola et al. 2012; Åhslund and Boström 2018; Pendergast et al. 2018), this article considers potential differential treatment by gender. To test for potential interaction effects of gender with other variables of interest, we estimate separate models. However, since the experiment in question was not originally designed for this purpose, this is not unproblematic from a methodological point of view. Accordingly, this will be taken as an occasion to critically discuss not only the opportunities, but also the limitations and pitfalls of factorial survey experiments.

2 Background

In the following, the proposed mechanisms of discrimination are presented. Next, some background information is given on educational inequalities across social class, minority status, and gender as well as on how institutional characteristics affect educational opportunities. This is supplemented by an outline of the Swiss institutional context and a short discussion of the relevant terminology. Last, a brief overview of the literature on discrimination at school, with a focus on the methods employed, is given.

2.1 Mechanisms of Discrimination

There are three prominent theories of discrimination: the theory of preference-based discrimination (Becker 1995), statistical discrimination (Phelps 1972; Arrow 1973) as well as institutionalized discrimination (Feagin and Feagin 1986; Carmichael/Ture and Hamilton 1992; Gomolla 2017). In this paper, the focus lies on the theory of statistical discrimination because it provides a rationale for understanding processes of discrimination without having to make any assumptions about preferences or hidden agendas. Although developed for the labour market context, it is easily applicable to discrimination in the school context in terms of a mechanism based approach (Coleman 2000; Elster 2015).

Phelps' (1972) and Arrow's (1973) theory of statistical discrimination (developed independently to describe labour market discrimination) is based on the assumption that prejudices arise from perceptions of reality (Arrow 1973). This is a major difference to, for example, Becker's theory of preference-based discrimination, since there is no need to assume the existence of antagonistic attitudes for discrimina-

tion to prevail. Because employers have only imperfect information about potential employees, they prefer employing individuals who belong to a group which – at least in the perception of the employer – shows higher average levels of productivity (Arrow 1973). This means that if employers observe that women are less successful in the labour market (for whichever reasons), they will be especially cautious when considering hiring a female employee, notwithstanding her actual potential (which is not fully observable in advance).

Analogous to employers, teachers have easily accessible information on, for example, the ethnicity and social class of their pupils. Based on their everyday experience in the classroom, teachers also have an idea of the distribution of achievement levels among children of different social classes. Because families of minority children are often socio-economically deprived, this mechanism can also be applied to them (Kalter 2008; Beck et al. 2010; Ditton and Aulinger 2011). Since the theory of discrimination was developed for the context of an informational deficit in regard to productivity levels, one could argue that the theory is not well suited to the school context, since teachers have a profound knowledge of their students. However, research has shown that prejudice is very persistent, and humans are slow in adapting their attitudes (Rydgren 2009, 74). In general, people strive to categorize the observations they make in a way that is consistent with their stereotypes (Dovidio et al. 2013a) and they are thus reluctant to make adaptions to their system of beliefs.

2.2 Stereotypes and Prejudice

For a better understanding of the mechanisms through which statistical discrimination works, it is important to have a basic understanding of the somewhat complexly related concepts of stereotypes and prejudice (Allport 1979; Dovidio et al. 2005; Dovidio et al. 2013a; Fiske and Taylor 2013). For the present purpose, prejudices are understood as individual-level attitudes, consisting of cognitive, affective, and conative components that are directed toward a target group or an individual member, because of their membership in that specific target group (Allport 1979; Dovidio et al. 2013a). Prejudice reflects the evaluation of a group that "creates or maintains hierarchical status relations between groups" (Dovidio et al. 2013a, 7).

Stereotypes are cognitive representations or a structured set of beliefs about the shared attributes of groups of people (Ashmore and Del Boca 1979; Hamilton et al. 1985; Correll et al. 2013). Warmth and competence have thereby been identified as two fundamental aspects of stereotype content, with warmth being attributed to ingroups/cooperative groups, and competence being associated with social status (Fiske et al. 2002). This is why those at the lower end of the income distribution are perceived as less competent and motivated and ethnic minorities are often typically regarded as lacking warmth, as soon as they are regarded as competitors (Fiske et al. 2002; Dovidio et al. 2013a). Stereotypical beliefs about different ethnic groups which underlie prejudice (Eagly and Diekman 2005) can thus lead to the

discrimination of members of specific groups without there having to be negative intentions involved (Fiske 2000; Dovidio et al. 2013c; Correll et al. 2013). There is substantial agreement in the literature that contrary to prior belief of negative feelings and intentions as primary causes of discrimination, subtle, ambiguous beliefs and automatic cognitive processing as well as ingroup bias are in fact often responsible (Dovidio et al. 2005; Dovidio et al. 2013b; Fiske and Taylor 2013).

Research from Germany and Switzerland has shown that stereotypes about people from the Balkans are particularly negative, with low values of both warmth and competence (Binggeli et al. 2014; Froehlich and Schulte 2019). At the same time, because of the interaction with gender stereotypes, existing evidence suggests that male minority group members might face particularly negative stereotypes (Munsch et al. 2007; Pager and Karafin 2009).

2.3 Inequality and Social Class

The educational expansion, starting roughly in the middle of the last century, has not only brought about a considerable increase in the average educational attainment of the population, but also a decrease of class disparities (Becker 2006; Breen et al. 2009; Becker and Müller 2011). Nonetheless, the remaining inequalities are relatively stable and we are far from a situation in which every child has the same educational opportunities: Working class children continue to have lower chances of succeeding in education (Pfeffer 2008; Becker 2011).

2.4 Inequality and Minority Status

In German-speaking countries there is a correlation between minority status and social class (Kalter 2008; Beck et al. 2010; Ditton and Aulinger 2011), implying that minority children achieve lower educational outcomes than majority children due to class-related disadvantage. However, minority status is also relevant in its own right. For example, minority children are more likely to be transferred to special needs schools (Powell and Wagner 2002; Kronig 2003) and are more often streamed into less challenging school tracks (Beck 2015; Dollmann 2016). Even when controlling for social class, minority children generally achieve lower educational degrees (Hopf 1981; Quenzel and Hurrelmann 2010; Kristen et al. 2011; Becker and Beck 2012; Hadjar and Hupka-Brunner 2013; Heath and Brinbaum 2014; Diehl et al. 2016).

2.5 The Effects of a Streamed School System

As noted above, at the institutional level, streamed school systems foster higher degrees of educational inequalities (Müller and Karle 1993; Hanushek and Wössmann 2006; Pfeffer 2008; Blossfeld et al. 2016). Streaming, or tracking, is a method of formal vertical curricular differentiation as an organizational tool of schools, in which

students are grouped by (presumed) ability levels and thus provided with unequal instructional resources and learning materials (Sørensen and Hallinan 1984).

In such streamed school systems, where children are allocated to different streams according to grades and teacher recommendations instead of test results, class disparities in educational outcomes are larger (Müller and Kogan 2010, 226). As predicted by Sørensen (1970, 375) organizational differentiation per se seems to strengthen the effects of social origin on educational attainment. In an early study on streaming in the US, students from lower socio-economic classes, African Americans, and boys were found to have lower chances of being in higher ability groups, all else equal (Dauber et al.1996). This pattern of disadvantage of minority and lower class children in regard to higher track placement has been repeatedly reported in more recent research and across countries (OECD 2012; Pietsch and Stubbe 2007; Klapproth et al. 2013).

2.6 Streaming – the Swiss Context

The streaming decision to one of two secondary school tracks[2] in Bern, Switzerland, is mainly based on the teacher's recommendation in the middle of 6th grade, when pupils are around 11 or 12 years old. This first transition[3] is especially important, since it directly influences the subsequent educational trajectory – thus indirectly impacting future job prospects and life-chances on the whole (Weber 1925; Dauber et al. 1996; Müller and Mayer 1976; Glauser 2015). This transition from primary to secondary school is also the one for which class disparities are greatest (Müller and Karle 1993, 15).

Teachers' decisions on the school stream they recommend their students for are mostly based on their assessment of the pupils' future performance development in the school subjects German, French, and math, but also on their judgement of the pupils' soft skills (Becker and Beck 2012, 143) such as motivation. In practice, teachers rely heavily on grade point averages (GPAs) to legitimate their recommendations. This decision is then communicated at a parent–teacher meeting at which the student in question is also present. The ideal outcome of this meeting is when all three parties come to an agreement on which stream is best suited for the student. In the canton of Bern, in case of dispute, the pupil has to pass an exam to qualify for a placement in the higher stream (ERZ 2013).

Teacher recommendations should thus largely be determined by the children's performance levels, of which grades are a useful and easy indicator. This leaves little room for the discrimination of those clearly in the uppermost and lowest grade

2 In some Swiss cantons there are three streams, but in Bern, as in most cantons, there are only two streams in public schools. Private schools sometimes have an additional higher stream, directly preparing children for the academic upper secondary school (German: Gymnasium).

3 After 9th grade, when students have completed mandatory education there is a second transition to upper secondary school. While the majority enrols in a vocational education and training programme, a minority continues in a general educational track (Gymnasium).

segments. While grades themselves could already be biased, we can therefore expect mechanisms of discrimination to be most relevant for children whose grades are in the middle range.

Another point that needs to be considered is that parents can have an effect on teacher recommendations (Pohlmann-Rother 2009; Schneider 2011). Due to status maintenance considerations, higher class parents have higher educational aspirations for their children than lower class parents (Breen and Goldthorpe 1997; Paulus and Blossfeld 2007). And since ideally, teachers, parents and the student will try to reach an agreement on which stream is most suited, it is to be expected that teachers will sometimes give in to the parents' wishes (Pohlmann-Rother 2009). Similarly, immigrant optimism (Kao and Tienda 1995), for which there are several possible explanations (Fernández-Reino 2016), leads us to expect teachers to adapt to the higher educational aspirations of immigrant parents. This stands in contradiction to the general expectation of lower recommendations for minority children, which is based on the observation that, in Switzerland, they are underrepresented in higher tracks (Becker et al. 2013). There are of course other non-academic factors, such as mere belonging to a group (e. g., working class), that can influence teacher recommendations, for example, through stereotypical beliefs.

2.7 Discrimination at School? Methods Matter!

Teacher judgements can be biased by a variety of stereotypes (Parks and Kennedy 2007) which in turn could lead to discrimination. Teachers have been observed to have lower expectations of minority children (Ho and Cherng 2018; Pendergast et al. 2018). In a study investigating whether teachers respond differently to Turkish children compared to German children, teachers were found to be biased when faced with stereotype-conforming immigrant children (Glock and Krolak-Schwerdt 2013). Research on how teachers perceive boys and girls has shown that in general, teachers have lower opinions of boys' academic skills (Åhslund and Boström 2018; Pendergast et al. 2018) and also have lower expectations in regard to boys' behaviour (Mullola et al. 2012). Further research on stereotypes at the intersection of ethnic background and gender in the German context shows that male minority children (especially those from the Middle East) are associated with macho attitudes and endorsement of violence. At the same time, minority girls are viewed as oppressed and voiceless (Munsch et al. 2007).

Studies on discrimination at school have mostly used observational data and typically find no effects of minority status (Becker and Beck 2012) or only in some cases (e. g., for grades in German but not in math; Kristen 2006; Becker et al. 2011). Some of these studies point out that there remain penalizing effects of social class (Kristen and Granato 2007; Schneider 2011; Becker and Beck 2012). Others find additional disadvantage when minority children also come from lower class backgrounds (Relikowski et al. 2010) but there are only few studies that consider

the possibility of discrimination on grounds of social class. Maaz and Nagy (2010) observe differential grading of children depending on their social class. A study on the importance of secondary effects of social class for children with a Turkish background in Germany even finds positive effects after controlling for social class and school achievement (Kristen and Dollmann 2010). The authors explain this finding with the higher educational aspirations of this population. However, since it is hard to compare educational certificates attained in different countries, they acknowledge that the result might merely be an effect of a measurement error in regard to social class (Kristen and Dollmann 2010).

Generally speaking, evidence from experiments seems to be methodologically sounder and also yields more pessimistic findings. Hanna and Linden (2009) randomly assigned age, gender, and caste to children's essays in India and find that teachers discriminate against girls in the highest achievement group as well as against lower caste children of both genders. Van Ewijk (2011) constructed a similar experiment for the Netherlands, assigning either Dutch or foreign sounding names (Maroccan or Turkish) to essays and finds no direct discrimination effects. Another similar experiment, conducted in Germany, finds negative effects of Turkish sounding names on teacher's grading as well as fictitious track recommendations. Sprietsma (2013) finds that the effects are driven by a small subset of teachers. No evidence is found for the assumption that the discriminatory behaviour is caused by negative attitudes (Sprietsma 2013). Perhaps unexpectedly – or at least in contradiction to contact theory (Allport 1979) – it is the teachers who have most contact with minority children in their day-to-day work, who tend to discriminate most. Schulze and Schiener (2011), who use vignettes and whose design has the highest similarity to the approach used here, find no effects of migration background in regard to streaming recommendations. However, they find negative effects for children from lower social classes (Schulze and Schiener 2011).

2.8 Hypotheses

According to the theory of statistical discrimination and due to structurally caused observable differences in achievement levels, teachers are expected to hold different stereotypical beliefs and possibly prejudices against minority groups and boys. It is likely teachers will expect less from lower class children. These children will be assigned lower success rates in higher education and thus, on average, receive lower stream recommendations. Because of the correlation with class (Meyer 2003; Swiss Federal Statistical Office 2008), teachers might expect less from minority students as well. These beliefs and attitudes can cause discrimination, understood as unintentional biased behaviour or unequal treatment on the basis of group belonging. We thus expect lower class (hypothesis 1) as well as minority children (hypothesis 2) to have lower probabilities for higher track recommendations compared to their peers with more privileged and non-minority backgrounds.

3 Data and Methods

Since social desirability bias is a challenge for research on discrimination (Auspurg and Hinz 2015), the prevailing method of investigating discrimination in the school context has been indirect deduction (Kristen 2006; Becker et al. 2011; Becker and Beck 2012). This has usually been accomplished by including all presumably relevant independent variables or controls – such as grades or social class – in regression models and interpreting remaining effects of, for example, minority status, as a possible sign of discrimination. A difficulty which arises is that distinguishing between disadvantage due to resources and discrimination is not at all straightforward. With regard to minority children, the prevalent correlation between minority status and social class often leads to the finding that no further effects of minority background can be found once class has been controlled for (Kalter et al. 2007).

However, this approach, while acknowledging that, out of equal treatment concerns, it is illegitimate to discriminate against minority children, it inadvertently implies that the lower outcomes of lower-class children are unproblematic. The underlying logic is that, since we know that lower class children are disadvantaged because they have less familial resources to lean on, it makes sense that this is mirrored in lower outcomes. Even though the reasons for controlling for social class are usually justified with reference to primary and secondary effects (Boudon 1974), the logic of the methodological approach rules out the possibility of discrimination by class. Additionally, the operationalisation of social class can be contaminated by a foregoing discrimination of the child's parents on the job market (Kristen and Dollmann 2010).

3.1 Why a Factorial Survey Experiment?

Factorial survey or vignette experiments are a subtype of survey experiments. (Factorial) survey experiments are uniquely valuable in that they bring together the advantages of conventional population-based surveys and experiments (Angrist and Pischke 2009; Becker and Zangger 2015). Through the randomisation of treatments (dimensions), problems of unobserved heterogeneity are reduced and unbiased causal inferences can be drawn (Angrist and Pischke 2009; Auspurg and Hinz 2015; Mutz 2017). Also, since factorial surveys can easily be disseminated in paper pencil or online questionnaires and do not require participants to appear in a laboratory setting, they are more economical than many other experimental designs. At the same time, this allows researchers to reach many more people and more diverse populations, which is valuable in terms of external validity.

Furthermore, survey experiments are highly efficient in that they enable the collection of data on many aspects simultaneously. This is possible since a vignette consists of multiple dimensions and their levels. Additionally, by varying the different dimensions independently from each other, survey experiments make it possible to

identify the isolated effects of factors that are usually highly correlated in real life and are thus difficult to disentangle using observational data (Rossi and Anderson 1982).

Furthermore, factorial survey experiments counteract social desirability bias in a number of ways. For one thing, this is achieved by asking participants to evaluate hypothetical situations (Badgett and Folbre 2003). Second, since the attributes of interest are not explicitly brought to the attention of the respondents, but indirectly as part of an overall evaluation, the situation is closer to reality and respondents are not as aware of the experimental context. Third, respondents are not encouraged to focus on a single characteristic but are instead asked to evaluate a vignette as a whole, making the situation more lifelike. For these reasons, researchers can expect respondents to be more honest in their replies than is the case for traditional survey questions, but also for many laboratory settings in which, for example, respondents can feel observed and thus be induced to display socially desirable behaviours (Alexander and Becker 1978; Badgett and Folbre 2003; Wallander 2009). This, together with the many options researchers have in designing the experiment, is of course a valuable source of internal validity.

Besides these advantages, survey experiments also have disadvantages. In real life, teachers usually have a broad knowledge not only about the academic, but also the social competence of their pupils. In the case of a vignette experiment, all this information, which can be highly relevant for the decision at hand, can never fully be captured in all its complexity. Also, because the vignettes represent hypothetical situations, the decisions of the subjects have no real-life consequences and as such should be interpreted with caution; that is, the decisions made in a vignette experiment are not necessarily informative about how a person would behave in real-life situations (Sniderman 2011; Hainmueller et al. 2015). Furthermore, and as we shall see, as soon as there are flaws in the original setup of the experiment, researchers are very limited in their possibilities to correct the mistake in the analyses.

3.2 The Factors and their Operationalization

The vignettes are comprised of seven dimensions with two or three levels each. These variables were chosen for their properties of describing 6th grade pupils awaiting a recommendation for secondary school placement in a manner as efficient and natural-sounding as possible. While the key variables of interest are the social class and ethnic origin as well as the gender of the pupils, other relevant information for a streaming decision was included to provide necessary context. Sufficient background is important because when respondents feel they have not been provided with enough information to base their decisions on, they are forced to mentally supplement the information given in the vignettes (Auspurg and Hinz 2015). This is problematic since the researcher has no control over which elements the respondents are mentally adding. At best, this leads to lower precision (through noise) and at worst, to bias (when this additional information is correlated with the vignette values).

Table 1 Dimensions and Levels of Vignettes

Dimensions	Levels		
Ability	The teacher has a *low* opinion of [name's] academic ability		The teacher has a *high* opinion of [name's] academic ability
Gender and ethnic origin	Lara or Nora/Mirjana		Benjamin or David/Aleksandar
Motivation	In class, he/she is usually *not very motivated*		In class, he/she is usually *highly motivated*
Behaviour in class	and/but his/her behaviour is rather *challenging*		and/but his/her behaviour is rather *unremarkable*
Social class	His/her parents, who have jobs that require *no post-compulsory schooling*	His/her parents, who have jobs that require *completed vocational training*	His/her parents, who have jobs that require *tertiary education*
Parental educational aspirations	would like their son/daughter to be assigned to the *lower* track		would like their son/daughter to be assigned to the *higher* track

As the perhaps most central information, the pupil's ability and motivation in class were added, since they constitute selection criteria generally acknowledged as legitimate. Considering that next to these, other factors such as gender (Terrier 2015) and behaviour in class (Bennett et al. 1993; Maaz and Nagy 2010), have also been known to influence teacher recommendations, these variables were also included in the child descriptions. Additionally, parents' educational aspiration was added, as it can also be expected to influence a teacher's decision (Schneider 2011). Table 1 gives an overview of the vignette dimensions and levels.

Special attention was given to the operationalisation of the two main variables of interest – ethnic origin and social class – with the goal of minimising social desirability bias. As a measure of minority status, Slavic names were chosen as indicators of origins from Ex-Yugoslavia[4], one of the biggest immigrant communities in Switzerland. Slavic names were also chosen, because people from South Eastern Europe are typically associated with negative attributes in Switzerland (Wyssmüller 2005) and have been shown to be subject to discrimination in the Swiss context before (Fibbi et al. 2006; Hainmueller et al. 2015; Zschirnt 2020). Moreover, the Slavic names Aleksandar and Mirjana were chosen because they do not sound too blatantly foreign to German-speakers, the assumption being that teachers would

4 In Switzerland, immigrants from Ex-Yugoslavia make up 11 % (2017) of the total migrant population (residents without a Swiss passport), which is currently 25 %. In the school year 2015/16 (in Switzerland school starts in August and ends in July), 27 % of the children and adolescents attending school throughout Switzerland were not in possession of a Swiss passport (in Bern slightly less) (Swiss Federal Statistical Office 2017). Children with origins in Ex-Yugoslavia typically live in low-income households (in 2003, only immigrants from Turkey and Portugal had a lower average socio-economic status) (Amos et al. 2003).

more often respond in a socially desirable way if confronted with more obviously foreign sounding names. Since teachers were presented four vignettes and the design included decks in which girls or boys with a majority status were represented twice, additional "Swiss" names (common in German-speaking Switzerland) were added, so as not to have two children out of four with the same first names. Social class was operationalized indirectly through parental educational level, in order to achieve a more natural sounding description (as opposed to explicitly stating the child's class origin) without triggering social desirability bias.

Although there is no official GPA that has to be attained in order for a teacher to give a recommendation for the more demanding track, teachers nonetheless tend to rely heavily on this number. So as not to have a too dominant, all-absorbing factor, this problem was also circumvented by merely stating that the children had ambiguous GPAs. This was also a simple way of accounting for the expectation that discrimination is most likely to occur in less obvious cases when children have neither clearly very high nor very low grades.

3.3 Setup of the Factorial Survey Experiment

The vignette universe or the full-factorial, representing the Cartesian product of the dimensions and levels, consists of $2^6 3^1 = 192$ vignettes for this particular design. Because the expected sample size was only 50 teachers and each vignette should ideally be evaluated by at least four to five people (Auspurg and Hinz 2015), a small fractional factorial of 24 (six decks with four vignettes each) was constructed. Two-way interaction terms were included in the design for a variety of combinations, but not all, which will become a problem in some of the analyses below, as we shall see.[5]

With a value of 95.34, the *D*-efficiency[6] of the fractional factorial design was comparatively high, suggesting that the selected subsample represents the vignette universe well (Auspurg and Hinz 2015, 29). The dimensions are sufficiently independent of each other (see table A1 showing the correlation matrix in the appendix) except for parental aspirations and social class, where there is a correlation of 0.5, due to a deliberate restriction in the setup of the experiment. The combination of high social class (parents, who have jobs that require a tertiary degree) and low parental aspirations was excluded from the design, because it is unrealistic. Due to this restriction, the dimensions *social class* and *parental aspirations* are confounded by design and should thus always both be included in the models simultaneously.

The factorial survey experiment was presented to the respondents in an online questionnaire used as part of SECABS[7], a project on educational inequalities in Bern,

5 The setup included the following interaction terms: behaviour*gender, behaviour*ability, ethnic origin*ability, ethnic origin*motivation, social class*ability, social class*motivation and ethnic origin*social class.

6 *D*-efficiency is a measure of statistical precision ranging from 0 to 100.

7 «Selektivität und Effektivität des Chancenausgleichs an Berner Schulen» (SECABS).

Switzerland. The data were collected in the first half of 2015[8]. The teachers were shown four vignettes each and asked to state the probability that they would recommend a particular pupil for a more challenging versus a less demanding educational track on a scale from 0–100 %. To avoid framing effects (Kahneman and Tversky 2000), the teachers were asked to state the probabilities for both tracks (see figure A1 in the appendix). In addition to the randomly distributed decks, the four vignettes within decks were also randomised, so as to rule out potential effects of ordering.

3.4 Analytical Strategy

To account for the fact that the data has a hierarchical structure, with vignette characteristics referred to as level one and respondent characteristics as level two variables, multilevel models were estimated. If there were significant differences between the always consistent but less efficient fixed effects model and the more efficient but not always consistent random effects model (Allison 2009), the former was used. This was the case for the models including both male and female students. The separate models were estimated using random effects.

4 Findings and Discussion

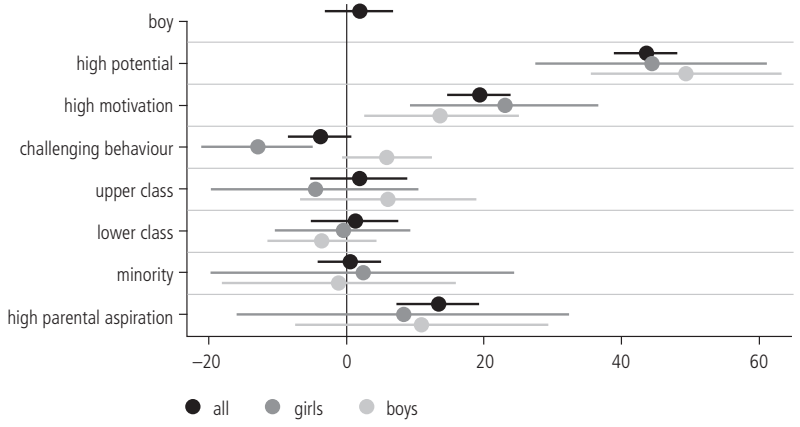
Given that 54 teachers participated in the online experiment, we have data for 216 hypothetical secondary school track recommendations. The teachers, of whom 29 identify as female and 23 as male, are between 26 and 66 years old with a mean age of 44. Only one of the teachers was born abroad (see table A2–A5 in the appendix for more information).

Because the teachers seem to have made quite different evaluations depending on the gender of the student, next to the analyses containing all vignette dimensions, separate models were estimated for boys and girls. However, unfortunately, the factorial survey experiment was not made with gender-interactions in mind. That is, the design did not include all interactions between gender and the other factors (see above), so that the separate models by gender need to be interpreted with caution. The results of all three models are presented in figure 1. The dependent variable is the probability (0–100 %) with which a teacher would recommend a student for the higher school stream.

As we can see in figure 1, only some of the information that was included in the factorial survey has a significant impact on teachers' responses. In accordance

8 The sampling strategy was a stratified random sample from all primary schools in the canton of Bern. Next to the teacher survey, the project included children and parent questionnaires as well as two achievement tests (German and math) and an intelligence test. Since participation required quite a high level of commitment and collaboration on the part of the teachers, those who participated in the study can be considered a rather selective group of highly motivated teachers.

Figure 1 Results Including All Vignette Dimensions



Reference categories: girl, low potential, low motivation, unremarkable behaviour, middle class, Swiss, high parental aspiration. **Models:** all: N = 216; $wR^2 = 0.73$; $bR^2 = 0.06$; $oR^2 = 0.60$; FE. Girls: N = 114; $wR^2 = 0.62$; bR2 = 0.06; $oR^2 = 45$; RE. Boys: N = 102; $wR^2 = 0.86$; $bR^2 = 0.61$; oR^2=0.79; RE.

with the cantonal guidelines (ERZ 2013), a student's potential and motivation have the greatest effect on the streaming decision. In the full model, we see that having a high perceived potential (44 percentage points) and high motivation (19 percentage points) raises the probability of receiving a recommendation for the higher stream considerably. While potential seems to be somewhat more important for boys than girls, it is the other way around for motivation, although these differences are not very substantial (and possibly biased, due to the mentioned design restriction).

Behaviour in class, which could partly be classified as another indicator of learning behaviour but can also be seen as an additional issue in its own right, appears to be especially important for girls. While displaying challenging behaviour in class lowers girls' chances of a recommendation for a higher track by 13 percentage points, the same behaviour leads to no significant effects for boys (again: these gender differences need to be interpreted with caution). A possible explanation for this imbalance is that boys generally show more externalizing behaviours than girls (Moffitt and Caspi 2001; Silver et al. 2005, 47) and are accordingly generally perceived as having more challenging behaviour (Bennett et al. 1993). This might lead teachers to perceive such behaviour as "normal", while they generally expect girls to behave "better" and thus penalize those who do not live up to these higher expectations.

High parental aspiration has a positive significant effect on the probability of a higher stream recommendation in the model including both genders (+ 13 percentage points). However, we do not find this effect in the gender-separated models. Furthermore, contrary to expectations, we find no significant effects of either social

class or ethnic origin. Consequently, the results do not support hypothesis 1, assuming that teachers are less likely to recommend lower class children for the more demanding track, all else equal. Hypothesis 2, claiming that minority children receive less favourable streaming decisions, all else equal, is also not supported.

Finally, all three models have very high R-squared values, with within R-squared values of .73 in the full model, and values of .62 for the model including just girls, and .86 for the model with boys. These high values of explained variance are reassuring in terms of the design of the factorial survey.

4.1 Limitations

There are two main limitations to the presented research and they are interrelated. Because, with roughly 50 teachers, the expected sample size was small, it was a challenge to find a "good" design that meets all criteria of a *D*-efficient setup while including the most relevant dimensions and excluding implausible combinations. The factorial survey experiment presented here was but one small part of a long questionnaire, and so as not to overburden the respondents with too many vignettes, it was only possible to achieve all these goals with a small fractional factorial design. In the end, a fractional factorial design of 24, not including all two-way interactions, out of a vignette universe of 192 was implemented. Because a design of this size runs the danger of not covering enough independent variation of all variables and possible interactions, especially when restrictions are also in place, we need to be cautious when interpreting the findings.

In further exploratory analyses (figures A2 and A3 in the appendix) containing either social class or parental aspirations, we found differences in the effects of the dimensions: class, ethnic origin, and educational aspirations, depending on the composition of the models. This led us to question whether there was a design-based confounding due to the limitations of the size of the fractional factorial. Taking a closer look at the 24 vignettes by gender presented in table 2, we see that indeed the effects of parental aspirations and ethnic origin are confounded in the subsample of boys. In this subsample, there is no independent variation of these two variables: There are no Swiss parents with low educational aspirations.

This explains the inconsistent findings in regard to these variables and is comparable to the problems identified by Auspurg (2018) in a similar case (Czymara and Schmidt-Catran 2016). Once again, we are reminded of the importance of using either the whole vignette universe or at least a large enough fractional factorial to include all two-way interactions. Also, to circumvent the resulting confounding effects, if not completely implausible, all possible combinations of vignette levels should be allowed. In hindsight, it would probably have been much less problematic and tedious to allow for the combination of upper class and low aspirations in this specific experiment.

Table 2 Fractional Factorial Design Containing 24 Vignettes by Gender

Gender	Potential	Motivation	Behaviour	Class	Ethnic Origin	Parental aspirations
Girl	low	motivated	unremarkable	middle class	Swiss	low
	low	not motivated	challenging	upper class	minority	high
	low	motivated	challenging	middle class	minority	high
	low	motivated	unremarkable	lower class	minority	high
	low	motivated	unremarkable	upper class	minority	high
	high	not motivated	challenging	lower class	minority	high
	high	not motivated	unremarkable	middle class	minority	high
	low	motivated	challenging	lower class	Swiss	low
	high	not motivated	challenging	middle class	Swiss	low
	high	not motivated	unremarkable	lower class	Swiss	low
	high	not motivated	unremarkable	upper class	Swiss	high
	high	motivated	challenging	upper class	Swiss	high
	low	motivated	unremarkable	middle class	Swiss	low
Boy	low	not motivated	unremarkable	lower class	Swiss	high
	low	not motivated	unremarkable	upper class	Swiss	high
	low	not motivated	challenging	middle class	Swiss	high
	high	motivated	unremarkable	middle class	Swiss	high
	high	motivated	challenging	lower class	Swiss	high
	low	motivated	challenging	upper class	Swiss	high
	low	not motivated	unremarkable	middle class	minority	low
	low	not motivated	challenging	lower class	minority	low
	high	not motivated	challenging	upper class	minority	high
	high	motivated	unremarkable	lower class	minority	low
	high	motivated	unremarkable	upper class	minority	high
	high	motivated	challenging	middle class	minority	low

A second relevant limitation is of course the relatively small sample size. Since all presented findings go back to only 54 teachers, the possibility of a selective sample is a realistic and weighty threat to the validity of the results. Owing to the high amount of effort required for the teachers to participate in the project (of which the teacher survey was only one element), it is to be assumed that the teachers who finally consented are not representative of the teacher population. Assuming these teachers are especially motivated in providing their pupils with the best teaching they can offer, if anything, this should lead to an underestimation of real-world discrimination.

Apart from that, there is an ongoing discussion in the literature on the ecological validity of survey experiments, which all in all seems to speak in favour

of the validity of the method (Eifler 2010; Barabas and Jerit 2010; Hainmueller et al. 2015). However, in this study, discriminatory behaviour was measured using vignettes; all the obvious advantages of survey experiments aside, of course this is critical from the point of view that we did not actually observe the behaviour in the context of interest. This gap between hypothetical and actual behaviour must always be kept in mind (Auspurg and Hinz 2015). Further research that addresses these issues is necessary for more robust conclusions on this sensitive topic. The gender-specificity of teacher evaluations that the findings here strongly suggest, definitely deserve further exploration. Future survey experiments on potential discrimination at school should ideally include three-way interactions between gender, social class and ethnic background, to explore the possible intersectionality of these dimensions.

Additionally, it would have perhaps been more intuitive for the teachers to rate the vignettes on a polar scale instead of giving two separate probabilities for a recommendation for the higher and lower tracks.

5 Conclusion

The aim of this article was to test whether discrimination by teachers could be an additional source of disadvantage for minority and lower-class children transitioning to secondary school. As a way of addressing the problems of social desirability bias and in order to measure the issue of interest as directly as possible, a factorial survey experiment was developed and included in a teacher survey in Bern, Switzerland. Relying on the theory of statistical discrimination and the therewith associated implications of stereotypical beliefs, we expected adverse effects for minority children as well as those from lower social classes. We find that the results do not confirm the hypotheses in regard to social class or ethnic origin.

However, an unexpected yet unsurprising finding is that girls with challenging behaviour are penalized, while boys are not. A possible explanation for this finding is that, in their decision-making process, teachers are guided by stereotypes regarding the gender of pupils, understood in terms of automatic cognitive processing (Dovidio et al. 2005; Dovidio et al. 2013b; Fiske and Taylor 2013). While in this case, the reactions to stereotype non-conforming girls was negative, it is possible that in other cases, people might react positively to someone challenging their expectations. In any case, an important implication of this gender-specific effect of classroom behaviour is that teachers seem to consider different factors to be relevant for the streaming decision depending on the gender of a child.

The theory of statistical discrimination offers a basis for the understanding of how this mechanism of discrimination in the school context could work. From day-to-day experience, where they see that girls perform better on average, attain higher educational degrees (Buchmann et al. 2008; Voyer and Voyer 2014), and behave

better (Bennett et al. 1993; Moffitt and Caspi 2001; Silver et al. 2005, 47), teachers might come to expect more from girls than boys. However, while past research has shown that these stereotypical beliefs are especially disadvantageous in the case of lower class and minority boys (Munsch et al. 2007; Pager and Karafin 2009; Binggeli et al. 2014; Froehlich and Schulte 2019), in this study, we find no such effects.

Unfortunately, due to the small sample size, but also the restricted fractional factorial design, we cannot make any definite claims here. Further research avoiding the pitfalls discussed in this paper is direly needed for more confident conclusions. Using better, more robust designs, future research should target the interplay of gender with other factors more systematically. A more intersectional approach is needed to test whether teachers respond differently to children based on an interaction of gender, class, ethnic origin, and other factors.

From a methodological point of view, we hope to have demonstrated some of the benefits and potential problems of survey experiments. Although bearing many advantages, especially when dealing with sensitive issues such as discrimination, if not designed well, the possibilities for meaningful analyses are severely restricted. In particular, a very small fractional factorial design that does not account for all two-way interactions is hard to work with. Such a reduced design simply cannot reliably ensure an adequate degree of independent variation of dimensions. This problem is aggravated where there are interactions between dimensions in real life, that however, cannot be explored in the analyses. We also advise against making restrictions in regard to specific combinations of dimensions where not absolutely necessary in terms of plausibility. Restricting combinations always comes at the cost of confounding the effects of the respective dimensions, leaving us with no way of identifying their independent effects.

Although we do not find the expected disadvantages of a lower social class and minority background, the gender differences we find suggest that stereotypes can still be a cause of differential treatment in the school context. Because of the persistency of attitudes and beliefs, it would be naïve to expect teachers to be able to easily overcome their prejudices (Rydgren 2009; Dovidio et al. 2013a). Therefore, from a policy perspective, alternative allocation mechanisms, such as test based systems, might prove to generate more just outcomes. This, of course, is a very specific and rather technical type of solution for the reduction of educational inequalities. It remains without saying, that abstaining from early streaming in the first place, as well as levelling the degree of inequalities in society at large, would go a long way in reducing the inequalities we face today.

6 References

Åhslund, Ingela, and Lena Boström. 2018. Teachers' Perceptions of Gender Differences – What about Boys and Girls in the Classroom? *International Journal of Learning, Teaching and Educational Research* 17(4): 28–44. https://doi.org/10.26803/ijlter.17.4.2.

Alexander, Cheryl S., and Henry Jay Becker. 1978. The Use of Vignettes in Survey Research. *Public Opinion Quarterly* 42(1): 93–104. https://doi.org/10.1086/268432.

Allison, Paul David. 2009. *Fixed Effects Regression Models*. Quantitative Applications in the Social Sciences 160. Los Angeles: SAGE.

Allport, Gordon W. 1979. *The Nature of Prejudice*. Unabridged, 25th anniversary ed. Reading, MA: Addison-Wesley Pub. Co.

Amos, Jacques, Edi Böni and Mario Donati. (2003). *Parcours vers les formations postobligatoires : les deux premières années après l'école obligatoire. Résultats intermédiaires de l'étude longitudinale TREE*. Neuchâtel: Office fédéral de la statistique. http://edudoc.ch/record/2632 (15.09.2021).

Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. "The Experimental Ideal". In *Mostly Harmless Econometrics: An Empiricist's Companion*, 11–24. Princeton: Princeton University Press.

Arrow, Kenneth Joseph. 1973. The Theory of Discrimination. Pp. 3–33 in *Discrimination in Labor Markets*, edited by Orley Ashenfelter, Albert Rees, Princeton University, and Woodrow Wilson School of Public and International Affairs. Princeton, N. J: Princeton University Press.

Ashmore, Richard D., and Frances K. Del Boca. 1979. Sex Stereotypes and Implicit Personality Theory: Toward a Cognitive–Social Psychological Conceptualization. *Sex Roles* 5(2): 219–248. https://doi.org/10.1007/BF00287932.

Auspurg, Katrin. 2018. Konfundierte Ergebnisse durch ein zu stark beschränktes Design?: Anmerkungen zur Vignettenanalyse «Wer ist in Deutschland willkommen?"» von C. Czymara und A. Schmidt-Catran. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 70(1): 87–92. https://doi.org/10.1007/s11577-018-0515-0.

Auspurg, Katrin and Thomas Hinz. 2015. Multifactorial Experiments in Surveys. Pp. 291–315 in *Experimente in den Sozialwissenschaften*, edited by Marc Keuschnigg and Tobias Wolbring, 1st ed., Soziale Welt Sonderband 22. Baden-Baden: Nomos.

Badgett, Lee and Nancy Folbre. 2003. Job Gendering: Occupational Choice and the Marriage Market. *Industrial Relations: A Journal of Economy and Society* 42(2): 270–298.

Barabas, Jason and Jennifer Jerit. 2010. Are Survey Experiments Externally Valid? *American Political Science Review* 104(2): 226–242. https://doi.org/10.1017/S0003055410000092.

Beck, Michael. 2015. *Bildungserfolg von Migranten: Der Beitrag von Rational-Choice-Theorien bei der Erklärung von migrationsbedingten Bildungsungleichheiten in Bern und Zürich*. Prisma, Band 27. Bern: Haupt Verlag.

Beck, Michael, Franziska Jäpel, and Rolf Becker. 2010. Determinanten Des Bildungserfolgs von Migranten. Pp. 313–340 in *Bildungsverlierer: Neue Ungleichheiten*, edited by Gudrun Quenzel and Klaus Hurrelmann, 1st ed. Wiesbaden: VS, Verlag für Sozialwissenschaften.

Becker, Gary S. 1995. *The Economics of Discrimination*. 2. ed. Economics Research Studies of the Economics Research Center of the University of Chicago. Chicago: University of Chicago Press.

Becker, Rolf. 2006. Dauerhafte Bildungsungleichheiten Als Unerwartete Folge Der Bildungsexpansion? Pp. 27–61 in *Die Bildungsexpansion: Erwartete und unerwartete Folgen*, edited by Andreas Hadjar and Rolf Becker. Wiesbaden: VS Verlag für Sozialwissenschaften. http://dx.doi.org/10.1007/978-3-531-90325-5_2.

Becker, Rolf. 2011. Entstehung und Reproduktion dauerhafter Bildungsungleichheiten. Pp. 87–138 in *Lehrbuch der Bildungssoziologie*, edited by Rolf Becker. Wiesbaden: VS, Verlag für Sozialwissenschaften. http://link.springer.com/10.1007/978-3-531-92759-6_4. (09.09.2021)

Becker, Rolf, and Michael Beck. 2012. Herkunftseffekte oder Statistische Diskriminierung von Migrantenkindern in der Primarstufe? Pp. 137–163 in *Soziologische Bildungsforschung*, edited by Rolf Becker, and Heike Solga. Wiesbaden: Springer Fachmedien Wiesbaden. http://www.springerlink.com/index/10.1007/978-3-658-00120-9_6. (09.09.2021)

Becker, Rolf and Anna Etta Hecken. 2007. Studium oder Berufsausbildung? Eine Empirische Überprüfung der Modelle zur Erklärung von Bildungsentscheidungen von Esser sowie von Breen und Goldthorpe/University or Vocational Training? An Empirical Test of the Rational Choice Model of Educational Choices Suggested by Esser as Well as Breen and Goldthorpe. *Zeitschrift Für Soziologie* 36(2): 100–117.

Becker, Rolf, Franziska Jäpel, and Michael Beck. 2011. *Statistische und Institutionelle Diskriminierung von Migranten im Schweizer Schulsystem. Oder: Werden Migranten oder bestimmte Migrantengruppen in der Schule benachteiligt*. Bern: Institut für Erziehungswissenschaft der Universität Bern. http://edudoc.ch/record/96257/files/DiskrimierungMigrantenSchweiz.pdf. (21.07.2021)

Becker, Rolf, Franziska Jäpel, and Michael Beck. 2013. Diskriminierung durch Lehrpersonen oder herkunftsbedingte Nachteile von Migranten im Deutschschweizer Schulsystem? *Schweizerische Zeitschrift Für Soziologie* 39(3): 517–549.

Becker, Rolf and Wolfgang Lauterbach (eds). 2013. *Bildung als Privileg: Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit*. 4th rev. ed. Wiesbaden: Springer VS.

Becker, Rolf and Walter Müller. 2011. Bildungsungleichheiten nach Geschlecht und Herkunft im Wandel. Pp. 55–75 in *Geschlechtsspezifische Bildungsungleichheiten*, edited by Andreas Hadjar, 1st ed., Wiesbaden: VS, Verlag für Sozialwissenschaften.

Becker, Rolf, and Christoph Zangger. 2015. Experimentelle Bildungsforschung. Pp. 105–131 in *Experimente in den Sozialwissenschaften*. Soziale Welt Sonderband 22. Baden-Baden: Nomos.

Bennett, Randy E., Ruth L. Gottesman, Donald A. Rock, and Frances Cerullo. 1993. Influence of Behavior Perceptions and Gender on Teachers' Judgments of Students' Academic Skill. *Journal of Educational Psychology* 85(2): 347–356. https://doi.org/10.1037/0022-0663.85.2.347.

Binggeli, Steve, Franciska Krings, and Sabine Sczesny. 2014. Stereotype Content Associated with Immigrant Groups in Switzerland. *Swiss Journal of Psychology* 73(3): 123–133. https://doi.org/10.1024/1421-0185/a000133.

Blossfeld, Hans-Peter, Sandra Buchholz, Jan Skopek, and Moris Triventi (eds). 2016. *Models of Secondary Education and Social Inequality: An International Comparison*. EduLIFE Lifelong Learning Series 3. Cheltenham, UK; Northampton, MA: Edward Elgar Publishing.

Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality; Changing Prospects in Western Society*. Wiley Series in Urban Research. New York: Wiley.

Bourdieu, Pierre, and Jean-Claude Passeron. 1970. *La Reproduction*. Paris : Les éditions de minuit.

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. London: Routledge & Kegan Paul.

Breen, Richard and John H. Goldthorpe. 1997. Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society* 9(3): 275–305. https://doi.org/10.1177/104346397009003002.

Breen, Richard, Ruud Luijkx, Walter Müller, and Reinhard Pollak. 2009. Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries. *American Journal of Sociology* 114(5): 1475–1521.

Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education". *Annual Review of Sociology* 34 (1): 319–37. https://doi.org/10.1146/annurev.soc.34.040507.134719.

Carmichael/Ture, Stokely/Kwame and Charles V. Hamilton. 1992. *Black Power: The Politics of Liberation in America*. Vintage ed. New York: Vintage Books.

Coleman, James Samuel. 2000. *Foundations of Social Theory*. 3rd ed. Cambridge (Mass.): Belknap Press of Harvard University Press.

Correll, Joshua, Charles M. Judd, Bernadette Park, and Bernd Wittenbrink. 2013. Measuring Prejudice, Stereotypes and Discrimination. Pp. 45–62 in *The Sage Handbook of Prejudice, Stereotyping and Discrimination*, edited by John F. Dovidio, Miles Hewstone, Peter Glick, and Victoria M. Esses. Los Angeles: Sage.

Czymara, Christian S. and Alexander W. Schmidt-Catran. 2016. Wer ist in Deutschland willkommen?: Eine Vignettenanalyse zur Akzeptanz von Einwanderern. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 68(2): 193–227. https://doi.org/10.1007/s11577-016-0361-x.

Dauber, Susan L., Karl L. Alexander, and Doris R. Entwisle. 1996. Tracking and Transitions through the Middle Grades: Channeling Educational Trajectories. *Sociology of Education* 69(4): 290–307. https://doi.org/10.2307/2112716.

Diehl, Claudia, Christian Hunkler, and Cornelia Kristen (eds). 2016. *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*. Wiesbaden: Springer VS.

Ditton, Hartmut. 2010. Selektion und Exklusion im Bildungssystem. Pp. 53–71 in *Bildungsverlierer: Neue Ungleichheiten*, edited by Gudrun Quenzel and Klaus Hurrelmann. Wiesbaden: VS Verlag für Sozialwissenschaften. http://dx.doi.org/10.1007/978-3-531-92576-9_3.

Ditton, Hartmut and Juliane Aulinger. 2011. Schuleffekte und institutionelle Diskriminierung – eine kritische Auseinandersetzung mit Mythen und Legenden in der Schulforschung. Pp. 95–119 in *Integration durch Bildung*, edited by Rolf Becker. Wiesbaden: VS Verlag für Sozialwissenschaften. http://link.springer.com/10.1007/978-3-531-93232-3_5. (28.08.2021)

Dollmann, Jörg. 2016. Der Übergang von der Primar- in die Sekundarstufe. Pp. 517–542 in *Ethnische Ungleichheiten im Bildungsverlauf*, edited by Claudia Diehl, Christian Hunkler, and Cornelia Kristen. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-04322-3_12.

Dovidio, John F., Peter Samuel Glick, and Laurie A. Rudman (eds). 2005. *On the Nature of Prejudice: Fifty Years after Allport*. Malden, MA: Blackwell Pub.

Dovidio, John F., Miles Hewstone, Peter Glick, and Victoria M. Esses. 2013a. Prejudice, Stereotyping and Discrimination: Theoretical and Empirical Overview. Pp. 3–28 in *The Sage Handbook of Prejudice, Stereotyping and Discrimination*, edited by John F. Dovidio, Miles Hewstone, Peter Glick, and Victoria M. Esses. Los Angeles: Sage.

Dovidio, John F., Miles Hewstone, Peter Glick, and Victoria M. Esses (eds). 2013b. *The Sage Handbook of Prejudice, Stereotyping and Discrimination*. Los Angeles: Sage.

Dovidio, John F., Miles Hewstone, Peter Glick, Victoria M. Esses, Naomi Ellemers, and Colette Van Laar (eds). 2013c. Individual Mobility. Pp. 561–576 in *The Sage Handbook of Prejudice, Stereotyping and Discrimination*, 1. paperback ed. Los Angeles: Sage.

Duncan, Greg J. and Richard J. Murnane (eds). 2011. *Whither Opportunity?* NY: Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610447515. (12.09.2021)

Eagly, Alice H. and Amanda B. Diekman. 2005. What Is the Problem? Prejudice as an Attitude-in-Context. Pp. 17–35 in *On the Nature of Prejudice*, edited by John F. Dovidio, Peter Glick, and Laurie A. Rudman. Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470773963.ch2.

Eifler, Stefanie. 2010. Validity of a Factorial Survey Approach to the Analysis of Criminal Behavior. *Methodology* 6(3): 139–146. https://doi.org/10.1027/1614-2241/a000015.

Elster, Jon. 2015. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Erikson, Robert, and Jan O. Jonsson (eds). 1996. *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Social Inequality Series. Boulder (Colo.): Westview Press. http://www.ub.unibas.ch/tox/HBZ/HT007441421/PDF. (25.09.2021)

ERZ, Die Erziehungsdirektion des Kantons Bern. 2013. *Direktionsverordnung über die Beurteilung und Schullaufbahnentscheide in der Volksschule (DVBS)*. Vol. 432.213.11. https://www.belex.sites.be.ch/frontend/versions/1303/embedded_version_content. (21.07.2021)

Feagin, Joe R., and Clairece Booher Feagin. 1986. *Discrimination American Style: Institutional Racism and Sexism*. 2nd augmented ed. Malabar, FL: R.E. Krieger Pub. Co.

Fernández-Reino, Mariña. 2016. Immigrant Optimism or Anticipated Discrimination? Explaining the First Educational Transition of Ethnic Minorities in England. *Research in Social Stratification and Mobility* 46 (December): 141–156. https://doi.org/10.1016/j.rssm.2016.08.007.

Fibbi, Rosita, Mathias Lerch, and Philippe Wanner. 2006. Unemployment and Discrimination against Youth of Immigrant Origin in Switzerland: When the Name Makes the Difference. *Journal of International Migration and Integration/Revue de l'integration et de La Migration Internationale* 7(3): 351–366.

Fiske, Susan T. 2000. Stereotyping, Prejudice, and Discrimination at the Seam between the Centuries: Evolution, Culture, Mind, and Brain. *European Journal of Social Psychology* 30: 299–322.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition. *Journal of Personality and Social Psychology* 82(6): 878–902. https://doi.org/10.1037/0022-3514.82.6.878.

Fiske, Susan T. and Shelley E. Taylor. 2013. *Social Cognition: From Brains to Culture*. 2nd ed. Thousand Oaks: SAGE.

Froehlich, Laura, and Isabel Schulte. 2019. Warmth and Competence Stereotypes about Immigrant Groups in Germany. Edited by Jennifer Steele. *PLOS ONE* 14 (9): e0223103. https://doi.org/10.1371/journal.pone.0223103.

Glauser, David. 2015. *Berufsausbildung oder Allgemeinbildung*. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09096-8.

Glock, Sabine and Sabine Krolak-Schwerdt. 2013. Does Nationality Matter? The Impact of Stereotypical Expectations on Student Teachers' Judgments. *Social Psychology of Education* 16(1): 111–127. https://doi.org/10.1007/s11218-012-9197-z.

Goldthorpe, John H. 2003. The Myth of Education-Based Meritocracy. *New Economy* 10(4): 234–239. https://doi.org/10.1046/j.1468-0041.2003.00324.x.

Gomolla, Mechtild. 2017. Direkte und indirekte, institutionelle und strukturelle Diskriminierung. Pp. 133–155 in *Handbuch Diskriminierung*, edited by Albert Scherr, Aladin El-Mafaalani, and Gökçen Yüksel. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10976-9_9.

Hadjar, Andreas. 2008. *Meritokratie Als Legitimationsprinzip: Die Entwicklung Der Akzeptanz Sozialer Ungleichheit Im Zuge Der Bildungsexpansion*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hadjar, Andreas and Sandra Hupka-Brunner (eds). 2013. *Geschlecht, Migrationshintergrund Und Bildungserfolg*. Weinheim: Beltz Juventa.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. Validating Vignette and Conjoint Survey Experiments against Real-World Behavior. *Proceedings of the National Academy of Sciences* 112(8): 2395–2400. https://doi.org/10.1073/pnas.1416587112.

Hamilton, David L, Patricia M Dugan, and Tina K Trolier. 1985. The Formation of Stereotypic Beliefs: Further Evidence for Distinctiveness-Based Illusory Correlations. *Journal Of Personality and Social Psychology* 48(1): 5–17.

Hanna, Rema and Leigh Linden. 2009. Measuring Discrimination in Education. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w15057.

Hanushek, Eric A. and Ludger Wössmann. 2006. Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries*. *The Economic Journal* 116(510): C63–76.

Heath, A. F. and Yaël Brinbaum (eds). 2014. *Unequal Attainments: Ethnic Educational Inequalities in Ten Western Countries*. Proceedings of the British Academy 196. Oxford: Published for the British Academy for Oxford University Press.

Ho, Phoebe and Hua-Yu Sebastian Cherng. 2018. How Far Can the Apple Fall? Differences in Teacher Perceptions of Minority and Immigrant Parents and Their Impact on Academic Outcomes. *Social Science Research* 74(August): 132–145. https://doi.org/10.1016/j.ssresearch.2018.05.001.

Hopf, Diether. 1981. Schulprobleme Der Ausländerkinder. *Zeitschrift Für Pädagogik* 27(6): 839–61.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. 1972. *Inequality – A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books, Inc.

Kahneman, Daniel, and Amos Tversky (eds). 2000. *Choices, Values, and Frames*. New York : Cambridge, UK: Russell sage Foundation; Cambridge University Press.

Kalter, Frank. 2008. Ethnische Ungleichheit auf dem Arbeitsmarkt. Pp. 303–332 in *Arbeitsmarktsoziologie*, edited by Martin Abraham and Thomas Hinz. VS Verlag für Sozialwissenschaften. http://dx.doi.org/10.1007/978-3-531-91192-2_10.

Kalter, Frank, Nadia Granato, and Cornelia Kristen. 2007. Disentangling Recent Trends of the Second Generation's Structural Assimilation in Germany. Pp. 214–245 in *From Origin to Destination*, edited by Stefani Scherer, Reinhard Pollak, Otte Gunnar, and Markus Gangl. Frankfurt; New York: Campus.

Kao, Grace and Tienda. 1995. Optimism and Achievement: The Educational Performance of Immigrant Youth. *Social Science Quarterly* 76(1): 1–19.

Klapproth, Florian, Sabine Glock, Sabine Krolak-Schwerdt, Romain Martin, and Matthias Böhmer. 2013. Prädiktoren der Sekundarschulempfehlung in Luxemburg: Ergebnisse einer Large-Scale-Untersuchung. *Zeitschrift für Erziehungswissenschaft* 16(2): 355–379. https://doi.org/10.1007/s11618-013-0340-1.

Kristen, Cornelia. 2006. Ethnische Diskriminierung in Der Grundschule? *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58(1): 79–97.

Kristen, Cornelia and Jörg Dollmann. 2010. Sekundäre Effekte Der Ethnischen Herkunft: Kinder aus Türkischen Familien am ersten Bildungsübergang. Pp. 177–144 in *Vom Kindergarten bis zur Hochschule: Die Generierung von Ethnischen und Sozialen Disparitäten in der Bildungsbiographie*, edited by Birgit Becker and David Reimer. Wiesbaden: VS Verlag für Sozialwissenschaften. http://link.springer.com/chapter/10.1007/978-3-531-92105-1_5. (08.09.2021)

Kristen, Cornelia, Aileen Edele, Frank Kalter, Irena Kogan, Benjamin Schulz, Petra Stanat, and Gisela Will. 2011. The Education of Migrants and Their Children across the Life Course. *Zeitschrift für Erziehungswissenschaft* 14(S2): 121–137. https://doi.org/10.1007/s11618-011-0194-3.

Kristen, Cornelia and Nadia Granato. 2007. The Educational Attainment of the Second Generation in Germany: Social Origins and Ethnic Inequality. *Ethnicities* 7(3): 343–366. https://doi.org/10.1177/1468796807080233.

Kronig, Winfried. 2003. Das Konstrukt des leistungsschwachen Immigrantenkindes. *Zeitschrift für Erziehungswissenschaft* 6(1): 126–141. https://doi.org/10.1007/s11618-003-0008-3.

Maaz, Kai and Gabriel Nagy. 2010. *Der Übergang von der Grundschule in die Weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung Primärer und Sekundärer Herkunftseffekte*. Springer. http://link.springer.com/10.1007/978-3-531-92216-4_7.

McKay, Jade and Marcia Devlin. 2016. "Low Income Doesn't Mean Stupid and Destined for Failure": Challenging the Deficit Discourse around Students from Low SES Backgrounds in Higher Education. *International Journal of Inclusive Education* 20(4): 347–363. https://doi.org/10.1080/13603116.2015.1079273.

Meyer, John W. 1977. The Effects of Education as an Institution. *American Journal of Sociology* 45(2): 55–77.

Meyer, Thomas. 2003. Ungebildet in Die Wissensgesellschaft? Risiken Junger Migrantinnen und Migranten auf dem Weg zu einer Berufsqualifikation. *Terra Cognita* 3: 24–29.

Moffitt, Terrie E., and Avshalom Caspi. 2001. Childhood Predictors Differentiate Life-Course Persistent and Adolescence-Limited Antisocial Pathways among Males and Females. *Development and Psychopathology* 13(2): 355–375. https://doi.org/10.1017/S0954579401002097.

Müller, Walter, and Wolfgang Karle. 1993. Social Selection in Educational Systems in Europe. *European Sociological Review* 9(1): 1–23.

Müller, Walter, and Irena Kogan. 2010. Education. Pp. 217–289 in *Handbook of European Societies*, edited by Stefan Immerfall and Göran Therborn. Springer New York. http://dx.doi.org/10.1007/978-0-387-88199-7_9.

Müller, Walter and Karl Ulrich Mayer. 1976. *Chancengleichheit durch Bildung?: Untersuchung über den Zusammenhang von Ausbildungsabschlüssen und Berufsstatus*. Gutachten und Studien der Bildungskommission, Deutscher Bildungsrat; Bd. 42. Stuttgart: Klett.

Mullola, Sari, Niklas Ravaja, Jari Lipsanen, Saija Alatupa, Mirka Hintsanen, Markus Jokela, and Liisa Keltikangas-Järvinen. 2012. Gender Differences in Teachers' Perceptions of Students' Temperament, Educational Competence, and Teachability: Gender Differences and Temperament. *British Journal of Educational Psychology* 82(2): 185–206. https://doi.org/10.1111/j.2044-8279.2010.02017.x.

Munsch, Chantal, Marion Gemende, and Steffi Weber-Unger Rotino (eds). 2007. *Eva ist emanzipiert, Mehmet ist ein Macho: Zuschreibung, Ausgrenzung, Lebensbewältigung und Handlungsansätze Im Kontext von Migration und Geschlecht*. Geschlechterforschung. Weinheim: Juventa.

Mutz, Diana Carole. 2017. *Population-Based Survey Experiments*. Princeton: Princeton Scholarship Online. http://dx.doi.org/10.23943/princeton/9780691144511.001.0001.

OECD. 2012. *Equity and Quality in Education*. OECD Publishing. https://doi.org/10.1787/9789264130852-en.

Pager, Devah, and Diana Karafin. 2009. Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making. *The ANNALS of the American Academy of Political and Social Science* 621(1): 70–93. https://doi.org/10.1177/0002716208324628.

Parks, Felicia R., and Janice H. Kennedy. 2007. The Impact of Race, Physical Attractiveness, and Gender On Education Majors' and Teachers' Perceptions of Student Competence. *Journal of Black Studies* 37(6): 936–943. https://doi.org/10.1177/0021934705285955.

Paulus, Wiebke and Hans-Peter Blossfeld. 2007. Schichtspezifische Präferenzen oder sozioökonomisches Entscheidungskalkül? Zur Rolle elterlicher Bildungsaspirationen im Entscheidungsprozess beim Übergang von der Grundschule in die Sekundarstufe. *Zeitschrift für Pädagogik* 53(4): 491–508.

Pendergast, Laura L., Linda Nickens, Sofia Pham, Simone Miliaresis, and Gary L. Canivez. 2018. Race and Gender Differences in Teacher Perceptions of Student Homework Performance: A Preliminary Examination. *Contemporary School Psychology* 22(3): 294–302. https://doi.org/10.1007/s40688-017-0162-x.

Pfeffer, F. T. 2008. Persistent Inequality in Educational Attainment and Its Institutional Context. *European Sociological Review* 24(5): 543–565. https://doi.org/10.1093/esr/jcn026.

Phelps, Edmund S. 1972. The Statistical Theory of Racism and Sexism. *The American Economic Review* 62(4): 659–661.

Pietsch, Marcus and Tobias C. Stubbe. 2007. Inequality in the Transition from Primary to Secondary School: School Choices and Educational Disparities in Germany. *European Educational Research Journal* 6(4): 424–445. https://doi.org/10.2304/eerj.2007.6.4.424.

Pohlmann-Rother, Sanna. 2009. *Der Übergang am Ende der Grundschulzeit: zur Formation der Übergangsempfehlung aus der Sicht der Lehrkräfte*. Empirische Erziehungswissenschaft, Band 17. Münster: Waxmann.

Powell, Justin JW and Sandra Wagner. 2002. Zur Entwicklung Der Überrepräsentanz von Migrantenjugendlichen an Sonderschulen in Der BRD Seit 1991. *Gemeinsam Leben: Zeitschrift für Integrative Erziehung* 10(2): 66–71.

Quenzel, Gudrun and Klaus Hurrelmann (eds). 2010. *Bildungsverlierer: neue Ungleichheiten*. 1. Aufl. Wiesbaden: VS, Verlag für Sozialwissenschaften.

Relikowski, Ilona, Thorsten Schneider, and Hans-Peter Blossfeld. 2010. Primäre Und Sekundäre Herkunftseffekte beim Übergang in das gegliederte Schulsystem: Welche Rolle spielen Soziale Klasse und Bildungsstatus in Familien mit Migrationshintergrund? Pp. 143–167 in *Komparative Empirische Sozialforschung*, edited by Tilo Beckers. Wiesbaden: VS, Verlag für Sozialwissenschaften. http://link.springer.com/chapter/10.1007/978-3-531-92472-4_6. (09.09.2021)

Rossi, Peter H. and Andy B. Anderson. 1982. The Factorial Survey Approach: An Introduction. Pp. 1–25 in *Measuring Social Judgements*, edited by Peter H. Rossi, and Steven L. Nock. Beverly Hills: Sage Publications.

Rydgren, Jens. 2009. Beliefs. Pp. 72–93 in *The Oxford Handbook of Analytical Sociology*, edited by Peter Hedström and Peter S. Bearman. Oxford: Oxford University Press.

Schneider, Thorsten. 2011. Die Bedeutung Der Sozialen Herkunft Und Des Migrationshintergrundes Für Lehrerurteile Am Beispiel Der Grundschulempfehlung. *Zeitschrift Für Erziehungswissenschaft* 14(3): 371. https://doi.org/10.1007/s11618-011-0221-4.

Schulze, Alexander and Jürgen Schiener. 2011. Lehrerurteile Und Bildungsgerechtigkeit: Schullaufbahnempfehlungen angehender Lehrkräfte im Vignettenexperiment. *Zeitschrift Für Soziologie der Erziehung und Sozialisation* 31(2): 186–200.

Silver, Rebecca B., Jeffrey R. Measelle, Jeffrey M. Armstrong, and Marilyn J. Essex. 2005. Trajectories of Classroom Externalizing Behavior: Contributions of Child Characteristics, Family Characteristics, and the Teacher–Child Relationship during the School Transition. *Journal of School Psychology* 43(1): 39–60. https://doi.org/10.1016/j.jsp.2004.11.003.

Sniderman, Paul M. 2011. The Logic and Design of the Survey Experiment. Pp. 102–114 in *Cambridge Handbook of Experimental Political Science*, edited by Arthur Lupia, Donald P. Greene, James H. Kuklinski, and James N. Druckman. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511921452.008.

Sørensen, Aage B. 1970. Organizational Differentiation Of Students And Educational Opportunity. *Sociology of Education* 43(4): 355–376. https://doi.org/10.2307/2111838.

Sørensen, Aage B. and Maureen Hallinan. 1984. Effects of Race on Assignment to Ability Groups. Pp. 85–106 in *The Social Context of Instruction: Group Organization and Group Processes*, edited by Penelope L. Peterson, Louise Cherry Wilkinson, and Maureen T. Hallinan. Educational Psychology. New York: Academic Press.

Sprietsma, Maresa. 2013. Discrimination in Grading: Experimental Evidence from Primary School Teachers. *Empirical Economics* 45(1): 523–538. https://doi.org/10.1007/s00181-012-0609-x.

Swiss Federal Statistical Office. 2008. *Ausländerinnen und Ausländer in der Schweiz [2008]: Bericht 2008*. Neuchâtel: Swiss Federal Statistical Office.

Terrier, Camille. 2015. Giving a Little Help to Girls? Grade Discrimination and Its Effect on Students' Achievement. Working Paper. Paris School of Economics. https://www.parisschoolofeconomics.eu/IMG/pdf/jobmarket-paper-terrier-pse.pdf. (27.07.2021)

Tsakloglou, Panos and Fotis Papadopoulos. 2002. Aggregate Level and Determining Factors of Social Exclusion in Twelve European Countries. *Journal of European Social Policy* 12(3): 211–225.

van Ewijk, Reyn. 2011. Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review* 30(5): 1045–1058.

Voyer, Daniel and Susan D. Voyer. 2014. Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin* 140(4): 1174–1204. https://doi.org/10.1037/a0036620.

Wallander, Lisa. 2009. 25 Years of Factorial Surveys in Sociology: A Review. *Social Science Research* 38(3): 505–520. https://doi.org/10.1016/j.ssresearch.2009.03.004.

Weber, Max. 1925. *Wirtschaft Und Gesellschaft*. 2nd ed. Tübingen: J. C. B. Mohr.

Wyssmüller, Chantal. 2005. *Menschen «Aus dem Balkan» in Schweizer Printmedien: Diskursive Konstruktion und (Re)Produktion von Raum- und Identitätsbildern und deren Bedeutung für die Soziale Integration*. Forschungsbericht / Gruppe Für Sozialgeographie, Politische Geographie Und Gender Studies, Geographisches Institut der Universität Bern. Geographisches Institut, Universität Bern. https://books.google.ch/books?id=dF8MPwAACAAJ. (27.07.2021)

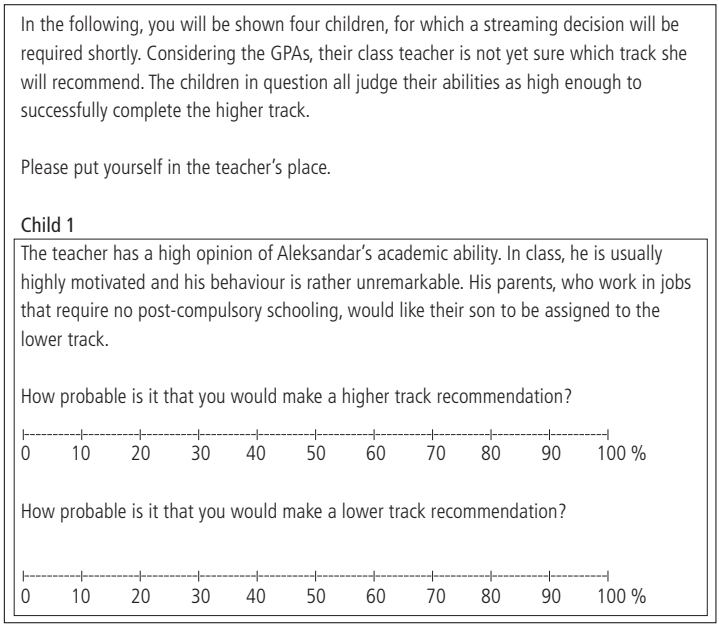
Zschirnt, Eva. 2020. Evidence of Hiring Discrimination Against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market. *Journal of International Migration and Integration* 21(2): 563–585. https://doi.org/10.1007/s12134-019-00664-1.

8 Appendix

Table A1 Correlation Matrix of the Vignette Dimensions

	Gender	Ethnic origin	Behaviour	Ability	Motivation	Social class	Parental aspirations
Gender	1.0000						
Ethnic origin	0.0484	1.0000					
Behaviour	0.0323	0.0000	1.0000				
Ability	–0.0484	–0.0161	0.0484	1.0000			
Motivation	0.0484	–0.0484	0.0000	0.0000	1.0000		
Social class	0.0683	–0.0877	–0.0195	0.0584	–0.0292	1.0000	
Parental Aspirations	–0.0255	–0.0598	–0.0085	–0.0085	–0.0598	0.4851	1.0000

Figure A1 Example Vignette


In the following, you will be shown four children, for which a streaming decision will be required shortly. Considering the GPAs, their class teacher is not yet sure which track she will recommend. The children in question all judge their abilities as high enough to successfully complete the higher track.

Please put yourself in the teacher's place.

Child 1

The teacher has a high opinion of Aleksandar's academic ability. In class, he is usually highly motivated and his behaviour is rather unremarkable. His parents, who work in jobs that require no post-compulsory schooling, would like their son to be assigned to the lower track.

How probable is it that you would make a higher track recommendation?

0 10 20 30 40 50 60 70 80 90 100 %

How probable is it that you would make a lower track recommendation?

0 10 20 30 40 50 60 70 80 90 100 %

Table A2 Educational Background of Teachers' Parents

	Frequency
No qualification	1
Primary school	3
Secondary school	6
Vocational education	15
Commercial high school	1
Grammar school without higher education entrance qualification	2
Grammar school with higher education entrance qualification	2
Higher vocational education	5
University of applied sciences / technical school	11
University/ETH	6
Total	52
Missing	2
Total	54

Table A3 Teachers' Migration Background

	Frequency
No	45
Yes, born abroad	1
Yes, parents born abroad	6
Total	52
Missing	2
Total valid	**54**

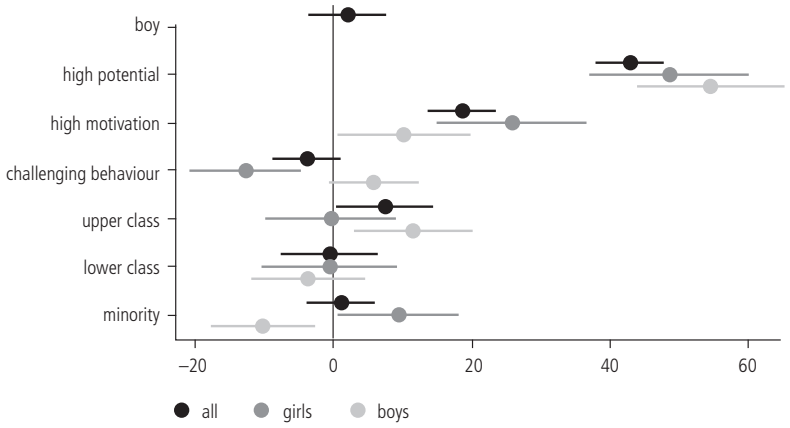
Table A4 Teachers' Gender

	Frequency
Male	23
Female	29
Total	52
Missing	2
Total valid	**54**

Table A5 Teachers' Age

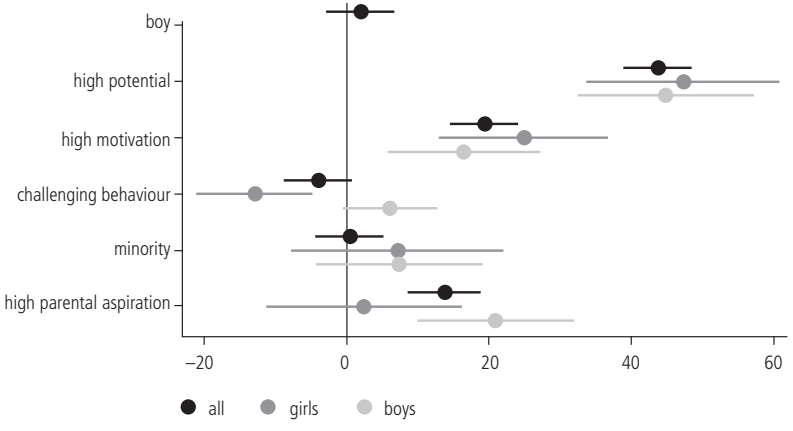
	Frequency
25–30	5
31–40	15
41–50	20
51–60	5
61–70	8
Total	53
Missing	1
Total valid	**54**

Figure A2 Results Using Class, but not Parental Aspirations



Reference categories: girl, low potential, low motivation, unremarkable behavoiur, middle class, Swiss, high parental aspiration. **Models:** all: N = 216; $wR^2 = 0.70$; $bR^2 = 0.06$; $oR^2 = 0.58$; FE. Girls: N = 114; $wR^2 = 0.62$; $bR^2 = 0.06$; $oR^2 = 0.44$; RE. Boys: N = 102; $wR^2 = 0.85$; $bR^2 = 0.61$; oR^2=0.78; RE.

Figure A3 Results Using Parental Aspirations, but not Class



Reference categories: girl, low potential, low motivation, unremarkable behavoiur, middle class, Swiss, high parental aspiration. **Models:** all: N = 216; $wR^2 = 0.73$; $bR^2 = 0.06$; $oR^2 = 0.60$; FE. Girls: N = 114; $wR^2 = 0.62$; $bR^2 = 0.06$; $oR^2 = 45$; RE. Boys: N = 102; $wR^2 = 0.86$; $bR^2 = 0.58$; oR^2=0.78; RE.