Is it the school of fish or the size of the pond that matters? An experimental examination of reference group effects in secondary school

Zangger, Christoph ; Gilgen, Sandra ; Moser, Nora

DOI: https://doi.org/10.1016/j.rssm.2023.100869

Research in Social Stratification and Mobility 89 (2024) 100869

Contents lists available at ScienceDirect

Research in Social Stratification and Mobility

journal homepage: www.elsevier.com/locate/rssm

Is it the school of fish or the size of the pond that matters? An experimental examination of reference group effects in secondary school

Christoph Zangger [a,*], Sandra Gilgen [b], Nora Moser [a]

[a] University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland
[b] University of Zurich, Andreasstrasse 15, 8050 Zürich, Switzerland

ARTICLE INFO

Keywords:
Reference group effects
Teacher recommendation
Choice experiment
Group size
Compositional effects

ABSTRACT

Do teachers consider not only an individual student's performance and abilities but also the number of motivated peers in a class as a frame of reference when deciding whether or not to recommend them for academic high school? Given the limited number of places available in such schools in the short run, we argue that a student's chances of being recommended depend on the number of competitors and especially whether they already secured a recommendation for themselves. Using choice experiments presenting groups of three to five students to pre-service secondary school teachers in Switzerland, we show how the individual probability for a recommendation for the most advantageous school track depends on the size of the reference group. Furthermore, individual chances are especially affected by the number of other students in the group that the teacher deems fit for academic high school: The higher the share of competitors in the reference group with a recommendation, the smaller the individual chances. These effects are robust across samples, methods, and with respect to alternative mechanisms.

1. Introduction

There is general agreement on the importance of educational transitions for the explanation of persistent inequalities in education. In this regard, many researchers have focused on intended and unintended consequences of purposive social actions made by the agents concerned: students, their parents, and teachers (Becker, 2003; Boone & Van Houtte, 2013; Breen & Goldthorpe, 1997, 1997; Stocke, 2007). A second line of research suggests that one's educational performance and attainment is also affected by the academic behavior and achievement of others, particularly by one's peers (Hanushek et al., 2003; Imberman et al., 2012; Sacerdote, 2011; Zimmerman, 2003). While the first perspective stresses the fact that there is a social, ethnic as well as a gender gradient in students' educational opportunities, the second focuses on reference-group effects caused by the social and academic composition of the student body (Alexander & Eckland, 1975; Marsh & Parker, 1984; Thrupp et al., 2002).

In this paper, we focus on the role of teachers as intermediate actors who influence both the effects of students' background characteristics as well as the workings of reference group effects. More specifically, we elaborate an alternative explanation for the finding that being surrounded by high-achieving peers can be detrimental to students' educational chances. This observation has often been attributed to the so-called Big-Fish-Little-Pond effect (Alicke et al., 2010; Davis, 1966; Marsh, 1987; Marsh & Hau, 2003; Marsh & Parker, 1984). However, instead of focusing on students' self-concepts or social learning from peers, we ask how the reference group of motivated students affects decisions regarding the recommendation of students to the highest secondary school track in Switzerland (academic high school). Our approach bears the advantage that individual actions of teachers are directly observable. Moreover, the literature on reference group effects mainly focuses on different compositional aspects such as the average achievement or social background of peers (Gröhlich & Guill, 2009; Huguet et al., 2009; Trautwein et al., 2006) and has largely neglected the second aspect of the BFLP effect: the size of the pond (Alicke et al., 2010; Zell & Alicke, 2009). In Switzerland, teachers in academic high schools cannot be substituted with teachers in secondary school from other tracks. Consequently, the number of available places in the highest track are fixed in the short run. In such a context, the bare number of competitors likely affects individual chances in a stratified education system. We therefore examine whether being part of a larger group of competitors has a detrimental effect on pupils' educational chances.

* Corresponding author.
E-mail address: christoph.zangger@unibe.ch (C. Zangger).

https://doi.org/10.1016/j.rssm.2023.100869
Received 31 May 2023; Received in revised form 21 November 2023; Accepted 1 December 2023
Available online 5 December 2023

More precisely, we investigate whether teachers use the number of students who aspire to go to the academic high school as a frame of reference when deciding whom to recommend. To this end, we collected data from pre-service teachers using choice experiments in the city of Bern, Switzerland. In this setting, track recommendations are of special interest since they enable direct access to the most advantageous school track in the Swiss school system without further examination. Using an experimental design, we are able to directly test for the role of teachers in the formation of reference group effects separately from other relative comparison mechanisms that work through students' achievement, self-concepts and classroom composition.

The remainder of this paper is organized as follows: The next section outlines the theoretical arguments and reviews the literature. The third section presents the data, the experimental design and the methods used to test the hypotheses. This is followed by a presentation of the results in the fourth section. The paper concludes with a discussion of the main findings in light of the literature as well as implications regarding educational opportunities in tracked education systems.

2. Background

Ever since Davis (1966) Frog-Pond, Marsh & Parker (1984) Big-Fish-Little-Pond effects and the Wisconsin model of status attainment (Sewell & Portes, 1969), sociologists and other social scientists have extensively studied reference group effects in the classroom context (for summaries see, for example, Epple & Romano, 2011; Sacerdote, 2011).

2.1. Reference group effects in education and the role of teachers

A prominent line of research on reference group effects focuses on the role of peers. A peer effect can be understood as any externality in which the current behavior, background or outcomes of peers affect individual outcomes under study (Sacerdote, 2011). Existing research evolves around the question how an individual student's academic performance, both in terms of achievement and attainment, is influenced by peers' performance or their background (e.g., Dumont & Ready, 2020; Epple & Romano, 2011; Sacerdote, 2011; Wilkinson, 2002). Depending on the mechanism studied, this research has generally found heterogeneous peer effects that often depend on a student's standing in the ability distribution (Ammermueller & Pischke, 2009; Lavy et al., 2012). An often referred to concept suggests that high performing students lower the self-concepts of their peers, which results in lower educational achievement (Davis, 1966; Gröhlich & Guill, 2009; Huguet et al., 2009; Marsh & Hau, 2003; Marsh & Parker, 1984; von Keyserlingk et al., 2019). Such Big-Fish-Little-Pond (BFLP) effects are furthermore found to be stronger in smaller groups (Alicke et al., 2010; Ready & Wright, 2011; Thijs et al., 2010; Zell & Alicke, 2009).

Another line of research on reference group effects focuses on the mediating role of teachers (Esser, 2016). As noted by Richer (1976), for a reference group effect to take place, the reference group must be salient and meaningful to the individual. While this is often assumed implicitly in studies on peer effects, these two conditions are explicitly met in the case of norm-referenced assessment where teachers evaluate one student in light of others (Dockx et al., 2019; Preckel & Brüll, 2010; Urhahne & Wijnia, 2021). In this regard, the literature on grading on a curve shows that teachers evaluate the performance of students in comparison to the achievement of their classmates (e.g., Andersen, 2018; Bresee, 1976; Calsamiglia & Loviglio, 2019; Dardanoni et al., 2009). One could assume that especially less experienced teachers could be "blinded" by the "reflected glory" of successful students in the peer group (Cialdini et al., 1976), which would result in positive spillover effects when judging lower achieving students that are surrounded by high achieving students.

However, other research on the topic has shown the opposite: that being surrounded by high achieving peers decreases individual grades due to a negative comparison, that is, a contrast effect. Moreover, as Bergold et al. (2022) summarize, there is also evidence for such contrast effects on teacher ratings of an individual student's perceived ability. In this case, the classroom context serves as a frame of reference for teachers' evaluations of their students (Hoge & Coladarci, 1989; Maaz et al., 2008).

On the other hand, in tracked and especially recommendation-based education systems, the same reference group processes will also affect educational attainment (Bergold et al., 2022). In this respect, Boone et al. (2018) work with data from Belgium suggests that teachers rely on classroom mean achievement as a frame of reference: While the ethnic and social composition have no effect on their recommendations regarding transitions to secondary school, the ability composition of students significantly influences their decisions. In line with a relative advantage explanation, the authors find a larger likelihood for low-ability students to be recommended for an academically orientated track when surrounded by a larger group of low-ability compared to high-ability peers. The same effects are also found in the German context, where Bergold et al. (2022) recently found contrast effects, that is, a negative reference group effect, on teacher ratings of students' ability, grades in math, and track recommendations. Finally, Baeriswyl et al. (2011) also report a negative contrast effect on the recommendation for the academic track in the Swiss context. This is in line with other evidence from different countries regarding reference group effects on measures of academic achievement as well as attainment (Gröhlich & Guill, 2009; for a contrasting finding of a positive reference group effect: Ready & Wright, 2011; Rothenbusch et al., 2016; Timmermans et al., 2015). However, all the mentioned studies on track recommendations rely on observational data.

We can think of several different mediating mechanisms that might explain such reference group effects on tracking decisions. For one, it has been suggested that teachers' experience and acquaintance with their students influences the size of the reference group effect. In this regard, Bergold et al. (2022) only find negative contrast effects among teachers who had been teaching their students for at least 3 years, with a positive effect for shorter acquaintance periods with their classroom. However, this positive "assimilation" effect was only found for ability and grades, but not for tracking recommendations. Meanwhile, teachers' overall teaching experience did not significantly alter the reference group effect on ability ratings, grades and tracking decisions. Likewise, Rothenbusch et al. (2016) also report that the negative contrast effect of being recommended for a gifted program also depends on teacher experience in a heterogeneous way, being – in some cases – stronger among teachers with more experience. In contrast, Ready & Wright (2011) report that teachers with little experience are more likely to rate students depending on the classroom context, leading to inaccurate evaluations of individual potential. Urhahne & Wijnia (2021), summarizing the research on accuracy in teacher judgments more generally, note that in most studies, experience effects are either week or absent. Batruch et al. (2023), for example, recently also found that the bias in teacher track recommendation is independent of experience. Consequently, teaching experience likely explains only part of reference group effects in recommendation decisions, and only in a non-trivial way.

Another explanation for reference group effects in teachers' recommendation decisions and the heterogenous finding reported above can be attributed to the institutional setting of a stratified education system (Rosenqvist, 2018). In this regard, the idea of school tracks already reflects the assumption of reference group effects in general and negative contrast effects in particular (Dumont et al., 2019; Richer, 1976): Students are grouped according to ability in order to facilitate learning. Card & Giuliano (2016), for example, show that especially high achieving minority students benefit in terms of math and reading scores when placed in a gifted/high achievement class, because their true potential is underestimated and suppressed in regular classrooms. What is more, the selection process involved in promotion and recommendation decisions for the different tracks suggests to teachers, as pointed out by

Batruch et al. (2019), that status inequality should be maintained and only the "fittest" students should attend the highest track, even when achievement is equal. Consequently, even though students might benefit by learning from higher achieving peers, thus increasing individual academic achievement (e.g. Fruehwirth & Gagete-Miranda, 2019; Hanushek et al., 2003), their educational attainment could still be lowered. In addition, in a tracked education system, the number of available slots in the different educational alternatives is – in the short run – limited and fixed. This capacity limitation gives rise to competition effects, especially with regard to educational transitions to more advantageous schools and tracks (e.g., Boone et al., 2018; Calsamiglia & Loviglio, 2019; Zangger, 2018). Summing up these arguments, teachers have a pivotal role in translating reference group processes in a stratified education system into the educational achievement and attainment of their individual students.

2.2. The present study

Given a stratified, recommendation-based education system and the results from observational studies that focus on classroom composition (Bergold et al., 2022; Boone et al., 2018; Dockx et al., 2019), we expect that teachers evaluate a particular student's performance and suitability for different educational tracks by comparing them to other students in their class. Regarding recommendations for the highest school track, the frame of reference consists of other eligible students in the class who also want to attend the highest track. Since each classroom acts as an individual frame of reference and thus shapes teachers' perceptions of their students, we expect recommendation decisions to vary among teachers.

Furthermore, the probability for a recommendation likely depends on the number of competitors, that is, the size of the group of students who would also like to attend the highest track. As it takes time to recruit specialized teachers, the number of available slots in the highest track is fixed – in the short run at least. This argument is supported by previous research that demonstrates – for the same transition and context studied here – that educational aspirations are positively interdependent at the neighborhood level. Meanwhile, these higher aspirations of peers translate into a negative competition effect for the actual transition (Zangger, 2018). In a recommendation-based system, one explanation for this finding is that teachers incorporate supply-side information in their recommendation decisions. What is more, existing experimental evidence from Switzerland points out that the selective nature of making a recommendation decision for the highest track suggest to teachers that they should only recommend the most suitable candidates (i.e., a cream skimming process; Batruch et al., 2019). Consequently, we hypothesize that the individual propensity for a recommendation decreases with an increasing number of peers who also have high ambitions and exhibit necessary performance levels.

While peer effects are found to be more pronounced in smaller groups (Thijs et al., 2010), we assume here that a students' relative standing in the ability distribution of a given set of competitors is especially important in larger groups. To ensure a relative advantage in a large `pond' and a limited set of available places, it becomes especially important to be the `big fish'. Meanwhile, even if peers outperform a particular student in a smaller group, their chances for a recommendation for the highest track should be less affected given that the available slots need to be filled.

The contribution of this paper is thus twofold: First, it provides a unique experimental assessment of how reference group effects work through the decisions of teachers. Existing research on peer effects either neglects the mediating role of teachers or is based on observational data that do not allow for a direct testing of the underlying decision-making mechanisms (e.g., Bergold et al., 2022; Boone et al., 2018; Dockx et al., 2019). Using an experimental design, we manipulate the different sources of reference group effects in teachers' recommendation decisions: the social and achievement composition of the reference group as well as the number of competitors. Meanwhile, due to unobserved confounders, studies based on (cross-sectional) observational data cannot fully disentangle the different reference group mechanisms. Second, our approach goes beyond existing research on reference group bias in teacher recommendations. So far, existing research has focused exclusively on the ability or social composition of the reference group (e.g., Baeriswyl et al., 2011; Bergold et al., 2022; Boone et al., 2018). Instead, we explicitly test whether there is a competition effect resulting from the limited number of available slots in the highest track. Our approach highlights the importance of taking the institutional setting into account when studying relative comparison processes. We thereby identify a further cause of negative effects of high-achieving peers on educational attainment without contradicting the existing evidence on positive effects in other domains. Before describing the experimental design and the analytical approach, we will provide some additional information about the Swiss education system, the relevance of teacher recommendations, and on how the linkage of the two causes the suggested competition effects.

2.2.1. The Swiss education system

In Switzerland, compulsory education ends after nine years of schooling, not including two years of mandatory kindergarten. However, completing additional education at upper secondary level is the norm, and only a small share of at most 10 % is not employed, in education or training (NEET) directly after completing compulsory schooling (Hupka-Brunner et al., 2010). Of those not belonging to this small group, roughly two thirds go on to complete vocational education and training. The remaining third enrolls in general educational programs, including academic high schools, previously referred to as the highest track (FSO, 2020). Graduates of the academic secondary school track are the only ones with the right to attend any university and almost any field of study without further examination[1]. Academic high schools thus represent the most advantageous educational alternative in Switzerland in terms of future life chances. Consequently, the transition under study (to academic high schools vs. other tracks) can be regarded as the most crucial one for students' future lives as well as in terms of inequality of opportunity more generally. In contrast to other secondary school teachers in Switzerland, teachers in academic high schools hold a university degree in a specific subject (e.g., a Master's in mathematics) and have completed an additional two to three years of teacher training on top. Consequently, teachers in academic high schools cannot be substituted by teachers from other school tracks. Meanwhile, teachers making the recommendation for the academic high schools are secondary school teachers who studied at a specialized university of teacher education and usually received a Master's degree in secondary education. Since studying at a university of teacher education does not require a universal university entry diploma but only a vocational baccalaureate, secondary school teachers did not necessarily attend themselves the academic secondary school track in the Swiss school system.

Although there is some regional variation regarding requirements and entry procedures for admission to the academic secondary school track, in general, it is determined by either an exam or by teacher recommendations, both of which enable direct access (CSRE, 2018). Moreover, previous grades are also considered in both procedures. Teacher recommendations ought to be based on pupils' academic performance as well as on their general classroom and learning behavior, which includes, for example, learning motivation, concentration skills, or the capacity to understand new things (Erziehungsdirektion des Kantons Bern, 2017). In line with the relevant legal foundations, teachers' assessments should not primarily constitute a review of past achievements. Instead, recommendations should be based on teachers' prognostic evaluation of a student's fit in a given secondary school track (Erziehungsdirektion des Kantons Bern, 2017).

[1] For some subjects such as medicine and sports science, the number of study places is limited, there is the additional restriction of an aptitude test.

On average, less than 4 pupils per class make the transition to the academic secondary school track. Although the number of young people transitioning to the highest track has generally increased, the share has remained relatively stable over the past thirty years (FSO, 2020). Throughout the past decades, the rate of graduates from the most advantageous school track has only peaked when the number of students went down significantly for demographic reasons (CSRE, 2018). This suggests that, when making their recommendations, teachers consider the available places in a particular track and fill them up accordingly (similar to the context of Card & Giuliano's, 2016 study). Thus, especially in the case of a large reference group, that is, when many classmates want to attend the academic track, we would expect the individual probabilities for receiving a recommendation to be lower.

At the same time, teachers might have an idea of the workings of reference group effects in the classroom: Universities of teacher education urge future teachers to critically engage with issues such as selectivity and inequality in education and to learn about how contextual factors, such as social or ethnic origin, influence the educational opportunities of students (PH Bern, 2016). Especially teachers in training and those who graduated only recently might therefore be aware of how their behavior in general, and particularly the recommendations they make, can be influenced by compositional and other contextual effects.

3. Methods & data

3.1. Sample

To examine the effects of reference groups on teacher recommendations for the academic school track, we administered a short survey to prospective secondary school teachers in Bern, Switzerland, in 2017. In total, 70 pre-service teachers participated in the study, representing a response rate of 83 % (84 subjects were invited to take part in the study) and comprising about two thirds of a cohort of prospective secondary teachers at the institution where they were enrolled. Of the 70 subjects, one did not complete the experiments described below and another did not provide all the background information needed in the multivariate analyses. This leaves us with a total of 68 cases with complete information. Even though the sample stems from a single university of teacher education (of a total of 16 such institutions in Switzerland, the sample represents almost 10 % of the whole population of future secondary school teachers. Respondents are between 19 and 45 years old and were studying for a secondary school teacher's degree at the time of data collection. With said degree, these pre-service teachers will teach in public secondary schools (7th to 9th grade) and will make the respective track recommendation for their 8th grade students. Even though these are pre-service teacher, more than 85 % of the respondents already had experience teaching in secondary schools, while around one half of the remaining 15 % had previously worked as primary school teachers. On average, they had about 9 months of teaching experience, while 5 subjects had no experience at all and 4 had 48 or more months of teaching experience. For the analyses, it is thus crucial that we take this heterogeneity in teaching experience into account.

3.2. Choice experiments

Using modified choice experiments, respondents evaluated three different groups of students from a set of hypothetical classes. A choice experiment is a type of survey experiment in which respondents are shown choice sets, each consisting of at least two alternatives and are asked to choose their preferred option. The options are described with the same set of attributes, which can take up one of several levels. Each attribute level corresponds to a treatment state and each alternative represents a combination of treatments (Ryan et al., 2008; Street, 2007). When combining the levels of attributes into the single alternatives, and the alternatives into choice sets, it is important to make sure that the single treatments are independent of each other. This can be achieved by using full factorials, which is the set of all possible combinations of attribute levels. Alternatively, one can combine the single treatments efficiently using a subsample of all possible combinations: a fractional factorial design.

The fractional factorial needs to be constructed in a way that allows for an unconfounded identification of all effects of interest. It is further recommendable to construct the alternatives and the choice sets in a statistically efficient way, such as by balancing the attribute levels and omitting dominant options[2] (Street, 2007). By choosing their preferred alternative in a choice experiment, respondents reveal their underlying implicit preferences and the trade-offs they make in their decisions. Consequently, choice experiments allow researchers to reach a causal understanding of people's preferences in a variety of settings. In this survey, we use a modified version of a choice experiment that will be described below.

An alternative strategy would have been to use a factorial survey experiment and present single vignettes of hypothetical students to teachers (a strategy used by, for example, Klapproth & Fischer, 2018, 2020; Pitten Cate et al., 2016). However, with this approach the reference group is only implicitly present (e.g., by stating that the vignette students all belong to the same class) while compositional elements would need to be explicitly stated (e.g., a particular student's standing in the ability distribution of the reference group), rendering the decision task rather artificial, and thus less reliable. Meanwhile, a choice experiment would also artificially force teachers to choose just one student. Our design, however, allows for teachers to express what is important to them when making a track recommendation in a more lifelike setting, as they are not explicitly restricted in the number of students they recommend and their comparisons can be affected by reference group effects, as in a real classroom.

3.3. Design of the modified choice experiment and operationalization

In this study, the choice sets consist of groups of 3, 4 or 5 hypothetical students of the same class who were described as wanting to attend the academic track in 9th grade. As elaborated above, in Switzerland less than 4 students per class are, on average, given the opportunity to attend this most advantageous school track. Thus, while in real-world contexts there will be cases with more than 5 students in a class seeking a recommendation, their teachers might also rarely find themselves in situations where only very few students or even no one wants to attend the academic track. Consequently, confronting the subjects with groups of 3–5 students comes close to the average in real life.

In the experiment, respondents were asked which pupils they would recommend for the academic high school. In contrast to traditional choice experiments where people are asked to choose only one among the available alternatives in each choice set, respondents were explicitly asked to make an individual recommendation for each pupil described in the vignettes. Moreover, they were told that recommending none or even all students was an equally valid option. This adaptation was made to accommodate for the fact that in real life teachers are expected to recommend all students they think have the potential to successfully complete the academic track.

The student vignettes were constructed as D-efficient fractional factorials which were then used to construct a series of choice-sets of three different group sizes (3 vs. 4 vs. 5 students; relative D-efficiency 95.71; Street, 2007). Table 1 depicts all characteristics used in the experimental setting. An example choice set of four students is shown in Fig. 1. Each teacher was presented with three such choice sets, one of each group size. To circumvent potential learning effects, the order of

[2] A dominant option is a combination of attribute levels that leads to an option of which it is reasonable to assume that a large majority would logically prefer over other options (Huber & Zwerina, 1996; Street, 2007).

Table 1
Experimentally altered treatments in the choice experiment.

Characteristics		Values
Students' characteristics		
	Gender	0 male, 1 female
	Grade last semester	4.5, 5.0, 5.5
	Performance	0 better, 1 about the same, 2 worse than before
	Participation during class	0 active, 1 not active
	Learning and study habits	0 learns new things easily, 1 needs some time
	Methods and learning style	0 independent & concentrated, 1 not very ind. & c.
	Parental education	0 attended grammar school, 1 didn't attend it
	Parental aspirations	0 grammar school, 1 vocational training
Classroom context		
	Group size	3, 4, 5
	Subject	0 math, 1 German

In what follows, three different 8. grade classes are presented. In the first class, 5 students would like to attend high school next year, in the second 4 students, and in the third 3 students. Please put yourself in the situation of being their subject teacher in mathematics. Based on the given information, try to decide for each student whether they should get a high school recommendation. The recommendation only concerns mathematics, the subject area that you teach.

Class 1:
Who of the following students would you recommend for high school in the subject area of mathematics?
Please put a cross below each student whom you would give a high school recommendation.

	Tom	Sarah	Lisa	Marc
Grade last semester (math)	5	5.5	4.5	5.5
Development of educational performance (compared with previous term)	Better than before	Same as before	Worse than before	Same as before
Participation during class	Actively contributes during class	Contributes only rarely during class	Contributes only rarely during class	Actively contributes during class
Pace of learning	Sometimes takes more time to learn new things	Learns new things easily	Sometimes takes more time to learn new things	Learns new things easily
Learning habits	Has trouble focusing and often needs assistance	Has no trouble focusing and can figure things out for herself/himself	Has trouble focusing and often needs assistance	Has no trouble focusing and can figure things out for herself/himself
Parental educational background	Attended grammar school	Did not attend grammar school	Did not attend grammar school	Attended grammar school
His/her parents would like their child to…	...complete grammar school	...complete a vocational degree	...complete grammar school	...complete a vocational degree
Grammar school recommendation	❑	❑	❑	❑

Fig. 1. Example choice set of hypothetical students wishing to attend the academic track.

the choice experiments, that is, the group / set size, as well as the order of the vignettes in each set were randomly altered. The median number of evaluations for each of the 16 choice sets in this study was 5, although 4 choice sets were only evaluated by two respondents.

Since a recommendation for the highest track should reflect an assessment of a student's future potential and not merely be based on prior academic performance (Erziehungsdirektion des Kantons Bern, 2017), grades as well as performance development in the past year should be key elements for teachers' decisions[3]. In the Swiss education system, grades range from 1, the lowest, to 6, the highest. The grade 4 equals a "pass", 5 is considered a "good" mark and 5.5 is given for a

[3] Klapproth & Fischer (2018) used a similar methodological approach to demonstrate that teachers' recommendation decisions are shaped by the development of students' GPAs.

"very good" achievement. In the experimental design, the grades range from 4.5 to 5.5. The latter is overrepresented since only students with a good to excellent performance can usually attend the most demanding track. The highest grade was not used in the design, since it is very difficult to achieve and would thus have been an overriding argument for a recommendation for the highest track. Characteristics that are so potent are a threat to survey experiments since respondents might fail to even register the values of the other treatments in their evaluations, which would potentially lead to biased results (Ryan et al., 2008).

Moreover, teachers are also asked to take students' learning habits and their classroom behavior into consideration when making the secondary school recommendations. To account for this, the choice experiment provides the respondents with information on students' participation during class, the pace with which they learn new things, and their learning habits. While the students are all characterized as highly motivated to attend the academic track, we altered parental aspirations as well as parents' educational background as part of the experimental setup. Finally, we also accounted for all two-way interactions in the design of the fractional factorial so that our experimental design also allows us to estimate interaction effects among the manipulated dimensions.

We opted against excluding any vignettes from the fractional factorial (implausible combinations) since doing so would violate the orthogonality of the design and lower its statistical power. More importantly, even the most "extreme" combinations, such as a good student with an advantageous background whose parents would not like them to attend the academic track, are not that implausible in the Swiss context where alternative educational routes such as VET are very common and, by international standards, have high reputations.

Figure A.1 in the appendix, depicting all pairwise correlations, shows that the experimentally altered dimensions are indeed independent of each other. To test the suggested competition effect for scarce places, we calculate the share of other students in the reference group of a particular vignette student i that secured a recommendation, that is, $\overline{y_i} = \sum_{j=1, j\neq i}^{J-1} y_j/(J-1)$, where J is the number of students in each choice set and y_j is a binary indicator of whether another student j received a recommendation or not. Similarly, for further robustness analyses, we also calculate an individual student's standing in the grade distribution of the reference group in the same way.

3.3.1. Respondent characteristics

Furthermore, additional data on respondents, in particular their teaching experience (Fig. A.2 in the appendix), knowledge about the recommendation process, as well as individual social and migration background were collected. Descriptive statistics for all the variables are summarized in the appendix (Table A.1). Although almost all respondents have teaching experience (more than 90 %), it should again be noted that they have not yet completed their teacher education for secondary school. Existing research using data from pre- and in-service teachers in Switzerland, however, finds no difference in their decision making when recommending students for the highest track (Batruch et al., 2019). Similarly, Klapproth & Fischer (2018) suggest that both pre-service and in-service secondary school teachers can be similarly affected by their implicit attitudes when making educational decisions. However, as with any experiment, we should nevertheless be careful when interpreting and generalizing the results. This issue will be addressed in detail in the conclusion.

3.4. Analytical strategy

Since respondents were asked to rate the suitability of every hypothetical student in each of the choice sets, each vignette (student) can be treated as an individual observation. After omitting cases with missing values, this yields a total of 816 decisions (68 pre-service teachers who rated 3 choice sets with 3, 4, and 5 vignettes, respectively). We use multilevel logistic regression models to account for the fact that each pre-service teacher is responsible for 12 observed outcomes (recommendation: yes/no; Allison, 2009). Thus, we estimate the model

$$\boldsymbol{Y} = \alpha + \boldsymbol{X\beta} + \eta_j + \varepsilon_{ij},\ i = 1, \ldots, n_j,\ j = 1, \ldots, J, \tag{1}$$

where i denotes the hypothetical student portrayed in the vignette and j is the index of the teacher making a recommendation. The matrix $\boldsymbol{X}$ depicts observed characteristics of both the hypothetical students seeking a recommendation (i.e., the characteristics in Table 1) and the observed teacher background characteristics (e.g., their experience, gender, social and ethnic background). The vector of the corresponding effects is denoted by $\boldsymbol{\beta}$.

Meanwhile, η_j captures the variance at the respondent level, that is, the extent to which pre-service teachers differ in their decisions (e.g., due to unobserved attitudes and beliefs). Since the respondents in our sample can be regarded as a random sample from the wider population of pre-service secondary school teachers, we could treat η_j in Eq. (1) as random and thus opt for a random effects model (Allison, 2009; Clarke et al., 2015). Before assessing the suitability of a random effects model, we check whether we could settle for a simpler one-level model such as an ordinary logistic regression model with clustered standard errors.

The first row of Table 2 reports the result of a likelihood ratio test comparing the baseline random effect model (model 1 in Table 3) to a corresponding ordinary binary logistic regression model. The corresponding χ^2 value significantly exceeds the expected value of 1 for no difference between the two models. Thus, the random effects model fits the data better than an ordinary binary logistic regression model. However, while a random effects model seems to be the better choice in the present case, it is only adequate if the random effect is uncorrelated with the individual error term (Clarke et al., 2015). To test whether $\eta_j$is independent of ε_{ij}, we use the well-known Hausman test that examines whether the random and the fixed effects estimates both target the same values of $\boldsymbol{\beta}$ (i.e., we test the null hypothesis that $E[\widehat{\boldsymbol{\beta}}_{RE}] = E[\widehat{\boldsymbol{\beta}}_{FE}] = \boldsymbol{\beta}$ (Clarke et al., 2015)). From the second row in Table 2, we conclude that it is safe for us to rely on a random effects model: With a value of 1.35 and an empirical p-value of 0.930, we clearly fail to reject the null hypothesis that the random and fixed effects models target the same values of $\boldsymbol{\beta}$.

4. Results

As a starting point, Fig. 2 shows the number of students that were recommended by the size of the reference group. The very slight increase in the number of recommended students in larger reference groups is something that we would expect, because individual characteristics, such as students' grades, should also make a difference: Through the random assignment of students to sets, larger sets are more likely to comprise additional high achieving students (although the ability distribution, on average, remains about the same). Meanwhile, the important point is that the number of recommended students increases only slightly, by a factor much smaller than 1 for each additional student in the reference group. (also see Table B.6 in the online supplementary material for a more detailed discussion of this point). What is more, from this first impression, it seems that, as expected, respondents recommend on average around 3 students for the academic track. Since the characteristics of the students (vignette characteristics / treatments) as well as the vignettes as a whole (combinations of treatments) are independent from each other by design and since they are

Table 2
Testing different model specifications.

Test	Test statistic (df)	p-value
Logit vs. random effects (LR)	28.43 (1)	0.000
Random vs. fixed effects (Hausman)	1.35 (5)	0.930

Table 3
Random effect models (being recommended for academic high school).

	(1)	(2)	(3)
Grade last semester	0.540***	0.541***	0.482***
	(0.011)	(0.011)	(0.022)
Participation *(Reference: active)*			
not active	-0.081***	-0.085***	-0.081***
	(0.023)	(0.023)	(0.022)
Learning & study habits (*Reference: learns new things easily*)			
needs more time to learn	-0.070**	-0.069**	-0.063**
	(0.023)	(0.023)	(0.022)
Methods & learning style (*Reference: high independence & concentration ability*			
low independence & concentration ability	-0.179***	-0.184***	-0.174***
	(0.023)	(0.023)	(0.023)
Group size	-0.025+	-0.026+	-0.034*
	(0.014)	(0.014)	(0.014)
Teaching experience (months)		-0.001	-0.002
		(0.001)	(0.002)
Share others with a recommendation			-0.272***
			(0.068)
Teacher: knowledge		✓	✓
Teacher: background characteristics		✓	✓
η_j	1.147**	1.008**	2.371***
	(0.394)	(0.368)	(0.779)
N (groups)	816 (68)		

Note: Average Marginal Effects (AME); standard errors in parentheses.
+p < .10. *p < .05. **p < .01. ***p < .001.

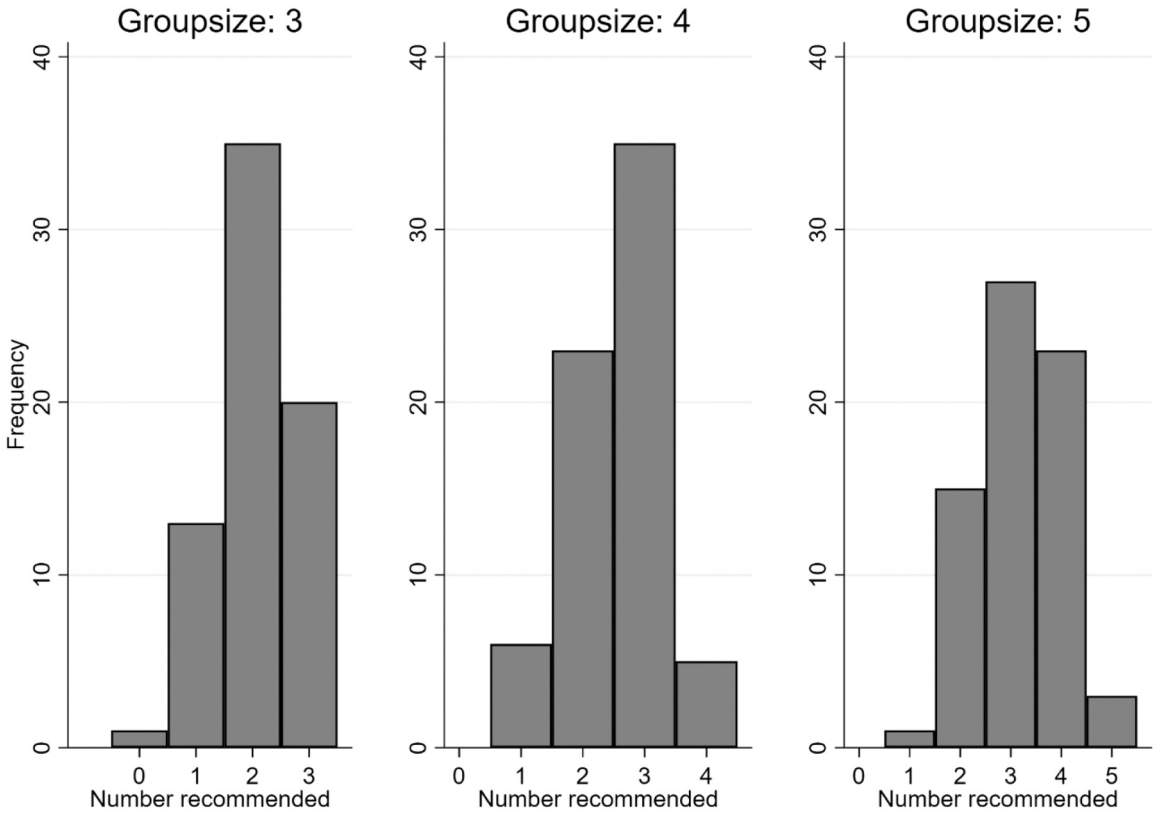

Fig. 2. Number of students recommended by size of choice set.

randomly assigned to choice sets of different sizes, this can be seen as a first indication that pre-service teachers also take into account the reference group. This would imply that respondents do not only consider individual ability and other traits of students when making a recommendation. More importantly, this also implies that being part of a larger group might indeed lower chances for a recommendation. Furthermore, it becomes apparent that respondents do not display behavior implying a forced choice mechanism, in which they try to avoid recommending all students of a given group – a point that will be taken up again in the robustness checks.

Moving on to the multivariate analyses, Table 3 depicts the results from logistic regression models with a respondent random effect (accounting additionally for the choice set level does not improve model fit nor does it change results – see Table B.3 in the online appendix). Effects are reported as average marginal effects (AME; Cameron & Trivedi, 2005). The significant random effect η_j suggests that pre-service teachers

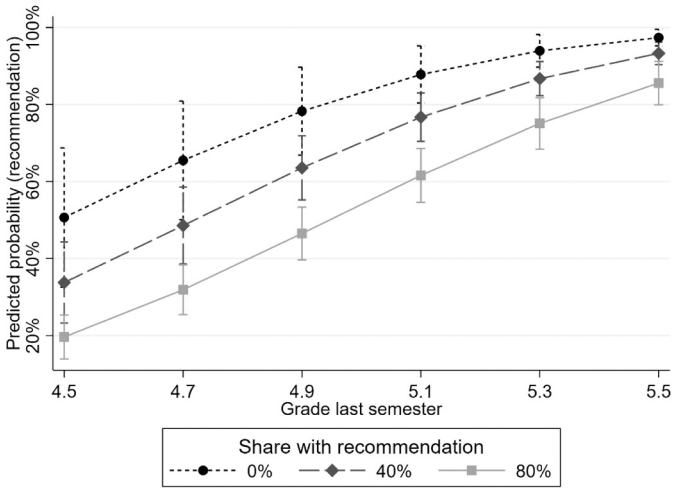

Fig. 3. Share with a recommendation by grades.

differ in their judgements of a pupil's potential. Although the magnitude of the random effect is somewhat reduced when controls for observed respondent characteristics, such as their teaching experience, knowledge about the recommendation process, or their social and ethnic background are included, their judgements seem to be further influenced by unobserved factors (comparing model 1 with model 2 in Table 3). Moreover, the models fit the data well (Table B.1 in the online appendix).

Turning our attention to the experimentally altered dimensions in the choice sets, we see that not all of them are significant predictors for a recommendation for the most demanding track[4]. While this is perhaps not surprising in the case of the subject (math or German), the missing effect of parental educational background, which serves as an indicator of social origin, and of parental educational aspirations needs explaining. First, the absence of an influence of parental education and aspirations could be explained by the fact that the educational aspirations of the students are held constant (they are all described as highly motivated to attend the academic secondary school track). Second, since we included students' grades, primary effects of social origin are already captured in all models (Becker, 2003; Breen & Goldthorpe, 1997). Third, as elaborated above, teachers are increasingly made aware of the enduring effects of social and ethnic origin as part of their education (PH Bern, 2016). Finally, the lack of significance only implies that respondents in our sample do not discriminate based on obvious measures of social background (DiMaggio, 1982). They might, for example, nevertheless punish challenging behavior for some social groups (Gilgen & Stocker, 2022).

Another variable that fails to have the expected effect is the development of students' educational performance (compared to the prior semester). The interaction of a student's grades with the development thereof has no significant effect either. This suggests that, in contrast to the intended procedure (Erziehungsdirektion des Kantons Bern, 2017), grades are taken as an absolute indicator of students performance, irrespective of students' potential for development.

All the other characteristics in the experimental design are highly significant predictors for recommendation decisions. A student's performance, measured in grades, is the most important predictor. On average, a unit change in a student's grade increases the likelihood of receiving a recommendation for the most demanding school track by 54 percentage points (p.p.). By contrast, students who do not participate actively in class, as well as those who take longer to learn new things have a 8 and 7 p.p. lower chance of being recommended. Similarly, respondents are 18 p.p. less likely to recommend students who are not very independent and have trouble concentrating. Meanwhile, respondents' teaching experience has no effect.

Of primary interest in the present context is that there is indeed a marginally significant negative effect of 2.5 p.p. for each additional pupil in the reference group of students with high ambitions. An additional model introducing the group size as a categorical variable shows that this can be attributed to the difference between 3 vs. 5 students per choice set. Comparing models 1 and 2 in Table 3, this effect, as well as all of the above mentioned, remain unchanged when including respondent-level control variables (e.g., respondents' teaching experience or their social background).

If we go one step further and introduce the share of other students with a recommendation, the effect sizes are generally reduced (model 3 in Table 3). This, however, is not the case for the group size effect, which increases in magnitude and reaches conventional levels of significance. Meanwhile, the share of other students with a high school recommendation has a strong negative effect on the individual chances of being recommended for the same track. As indicated in Fig. 3, this effect is strongest for students with grades around 5.0. Grades higher than 5.0 are often informally regarded as a necessary condition for a recommendation – although teachers are explicitly reminded not to base their recommendations solely on grades (Erziehungsdirektion des Kantons Bern, 2017). Unsurprisingly, for students who are characterized either by low or very high grades, the share of others with a recommendation does not matter as much. Nevertheless, there seems to be a tendency towards "filling up empty places" with even rather low-performing students if the peers in their reference group do not receive a recommendation (indicated by the significant differences for students with grades lower than 5.0. in Fig. 3).

We further explore whether the effect of peers with a recommendation depends on the size of the reference group. Since our sample size is too small for reliable evidence on interaction effects (Sommet et al., 2023), we abstain from any substantial interpretation. The corresponding effect in Table A.2 in the appendix suggests that the negative effect of an increasing share of peers with a recommendation is strongest in larger groups. However, a replication with a larger sample size is needed for any substantiated conclusions to be made.

Given the composition of our sample (all respondents are still in training to become accredited secondary school teachers), we additionally explore whether the competition effect found before differs with respondents' teaching experience. Again, the small sample size impedes any substantial interpretation of the corresponding interaction effect, reported in model 5 in table A.2 in the appendix. The combination of a negative main effect for respondents' teaching experience together with the positive interaction suggests that, on average, more experienced respondents are generally less likely to recommend a student, but they tend to be less influenced by the size of the reference group[5]. However, as mentioned above, this finding should be treated as a preliminary indication at most.

4.1. Robustness

We assess the robustness of our results based on five complementary strategies. First, to test whether our models are misspecified, we compare the results to more parsimonious models as well as to fixed

[4] The corresponding likelihood-ratio test comparing a model with all experimentally altered covariates to model 1 in Table 3 yields a p-value of 0.986. The two models can be considered equivalent. We therefore only present the result of models that do not include the insignificant covariates.

[5] When omitting the interaction between group size and share with a recommendation from model 5, however, the negative main effect of the size of the reference group is even more pronounced, pointing out that there is indeed considerable heterogeneity in its effect according to respondents' teaching experience.

effects regression models that account for unobserved heterogeneity. Moreover, we run three-level models that explicitly account for the student (vignette) level, choice set level, and respondent level. Second, we investigate whether the results are driven by a particular subset of the sample. Third, we test for alternative reference group mechanisms, which could potentially result in the same pattern as documented above. Fourth, we also check for potential learning effects of the respondents. Fifth, we perform additional analyses at the choice set level to investigate possible dependencies of the documented reference group effects on the setup of the experiment. Finally, we also check for effects of a wider frame of reference by using lagged quality characteristics of previous choice sets evaluated by the same respondent.

The results and patterns of the alternative models, especially when using respondent fixed effects are comparable to those of the random effects models reported above. Thus, unobserved heterogeneity seems not to be a concern (table B.2 in the supplementary material). Also, using student vignette fixed effects, we find a comparable, even slightly more pronounced competition effect of about a 5 p.p. lower probability of being recommended when the size of the reference group increases by one student. Also, explicitly accounting for the choice set level does not change results (table B.3 in the online appendix; the corresponding random effect is not significant). Nonetheless, it should be noted that we do not find the negative competition effect of the number of peers with a recommendation in the more parsimonious binary logistic regression models with clustered standard errors (Table B.1). However, since the pre-service teachers in our sample differ considerably in their decision-making, as we can see in the random effects models in Table 3, this is most likely a reflection of an underspecification of the true process.

Since our sample consists of individuals who have not yet completed their teacher education, only about 85 % of all respondents have experience teaching in secondary school. Moreover, model 5 in Table A.2 in the appendix suggests that the group size effect is attenuated when respondents have more teaching experience in general (e.g., as a primary teacher). Consequently, the results could be driven by the specific composition of our sample. Respondents without teaching experience in secondary school might be particularly prone to resorting to unobserved heuristics that could bias the results. However, rerunning the analyses on the subsample of respondents with teaching experience in secondary schools reveals an identical pattern of effects (Table B.4 in the online supplementary material).

Moreover, we investigated whether respondents fall back on a relative comparison process regarding students' performance. This could be an alternative explanation for the documented negative competition effect and a filling up of places in the highest track. To test this mechanism, we included two indicators: Whether a student's grade was below the peer group average and whether they had the lowest grade in their peer group. Both factors are not significant, strengthening our interpretation of the results (Table B.5 in the online appendix). Moreover, we also tested for potential ordering effects of the vignettes in the choice sets, which would indicate that the design was too demanding for respondents. We do not find any ordering effects.

Additional analyses on the choice set level further strengthens the interpretation of a contrast or competition effect (table B.6 in the online appendix). While the number of recommended students is generally higher in larger group, this positive effect is smaller than 1, resulting in diminishing individual chances for a recommendation. Moreover, pre-service teachers in our sample were less likely to recommend all students in larger groups, which also suggests that their choice of how many students to recommend was not driven by the design of the experiment.

Finally, and in a similar vein, we also checked whether the pre-service teachers in our sample might consider a wider frame of reference. Respondents might not only compare students within a given choice set but also to the ones from choice sets they evaluated before – corresponding to reference group effects in the real world that not only work within a given classroom but across classes taught by the same teacher. To test this assumption, we employed a lagged variable approach. However, we do not find any significant effects of the average grades or learning styles of the students rated in previous choice sets on an individual student's probability for a recommendation in the current choice set (Table B.7 in the supplementary material). Meanwhile, the competition effect remains unchanged in all models. Thus, the crucial frame of reference in our experiment is indeed at the level of a student's peers from the same class.

Together, these additional analyses make us confident that pre-service teachers indeed take the reference group of motivated students into account when deciding on whether or not to recommend a student for the academic high school track, and that the number of ambitious peers reduces individual chances, especially if they receive recommendations for the mentioned track. As in any competition, the more competitors one has, and the better they are, the harder it is for an individual to win the race.

5. Discussion and conclusion

This paper proposed and tested an alternative mechanism for reference group effects (e.g., Sacerdote, 2011; Zimmerman, 2003), such as the Big-Fish-Little pond effect (Huguet et al., 2009; Marsh & Parker, 1984). Reference group effects are often conceptualized as influencing individual educational achievement and attainment either directly through social learning and comparison processes among students or indirectly through students' self-concepts (e.g., Dumont & Ready, 2020; Feld & Zölitz, 2017; Gibbons & Telhaj, 2016; von Keyserlingk et al., 2019). Building on different streams of research, we tested an additional reference group effect that works through teachers' relative assessments of students.

Focusing on teacher recommendations for slots in the academic high school track which is the most demanding and most advantageous secondary school track in Switzerland, we argued that teachers' evaluations of a student's relative standing in the classroom guides their selection decisions. Given a – at least in the short run – limited number of slots in the highest track (teachers in academic high schools cannot be replaced with secondary school teachers from other tracks), we argued that teachers subsequently fill up places. Consequently, in addition to a student's absolute level of attainment and behavior, the individual chance for a secondary school recommendation should further depend on the number of peers who also have high aspirations and especially on whether they have already ensured a recommendation.

To test our assumptions, we analyzed data from a modified choice experiment and additional data collected in a survey of 68 pre-service secondary school teachers in Bern, Switzerland. Respondents were randomly allocated three different sets of three, four, and five hypothetical students wishing to attend the academic high school. They were then asked to decide, for each of these students individually, whether they would recommend them or not. Using logistic regression models with a respondent random effect, we find that the higher the number of ambitious peers, the lower the individual chances of obtaining a recommendation. However, this effect is rather small: An increase of one competitor amounts to about 3 percentage points lower chances of receiving a recommendation, which is about 1/15th of the effect of an increase in prior achievement by one grade. Furthermore, we show that this effect is mediated by the share of peers the teacher deems fit for the most demanding track. In line with research from the same (Batruch et al., 2019) as well as other contexts (Bergold et al., 2022; Boone et al., 2018; Pit-ten Cate et al., 2016), this effect confirms the assumed competition mechanism. Against the background of a more or less fixed number of available places in academic high schools, this result suggests that teachers take the (perceived) "supply side" needs into account and fill up the slots accordingly, recommending, on average, only about 3 students—irrespective of the number of students wishing to attend the academic high school track. This result is also in line with prior research from the US context, where Hallinan (1992) showed that the number of

students per track is to a large degree influenced by organizational factors that are independent of the achievement distribution of students. Unlike existing studies on reference group effects on recommendation decisions, using an experimental design enabled us to disentangle the suggested mechanisms from other, unobserved (peer group and individual) processes. What is more, we extend existing research on reference group effects on teacher recommendations by explicitly outlining and testing how variation in the class context is taken into account when making a decision.

Nevertheless, our study has several limitations. First, the results are based on a small sample of one cohort of prospective teachers in one out of 14 institutions in Switzerland. More importantly, although almost all respondents have teaching experience in either primary or secondary school, they are all still in education. A generalization to the wider population of in-service secondary school teachers is therefore not permissible. Meanwhile, other studies suggest only small or no differences between pre-service and in-service teachers in similar experiments (Klapproth & Fischer, 2018) and Batruch et al. (2019) show that the (SES-) bias in recommendation decisions in Switzerland is about the same for pre- and in-service secondary school teachers. Likewise, Bergold and colleagues (2022) recently showed for Germany that contrast effects, that is, negative reference group effects, are even stronger in cases where teachers are more acquainted with their class. Meanwhile, our own results suggest that more teaching experience could lead to less-pronounced reference group effects on the allocation of students to limited slots in academic high schools in Switzerland. However, given the small sample size and the associated potential problems in estimating interaction effects (Sommet et al., 2023), this finding clearly needs replication with a larger sample that comprises secondary school teachers with varying levels of experience.

Second, as with any experiment, the question of external validity arises. In real life, teachers usually do not directly compare the information they have on their ambitious students in the form of a table. In this regard, it is possible that the setup of the experiment in form of a modified DCE might have "forced" respondents to compare the portrayed students and suggested to them that only a limited number of students should be recommended. However, a variety of robustness checks suggests that this was not the case: In small reference groups of three students, for example, recommending all of them was the second most common choice, while the number of recommended students increased by less than 1 with each additional student in the reference group. More importantly, using a factorial survey as an alternative approach, the reference group would have to be operationalized even more explicitly in the setup, being even further away from real life situations. Additionally, since we find a comparable pattern of effects as in studies using observational data (Bergold et al., 2022; Boone et al., 2018) and since the very same competition effect has been found in the Swiss context at the neighborhood level – where a positive spatial interdependence of educational aspirations among neighbors translates into a negative effect for the actual transition to the highest track (Zangger, 2018) – we are confident that our results are indeed indicative of the underlying causal mechanisms.

A third concern was that the effect of grades could be overstated due to the hypothetical setting in the experiment. In real life, teachers have much more detailed knowledge on the ability and performance of students which allows them to judge their fitness for academic track more accurately. Meanwhile, lacking such in-depth knowledge about individual students in the experiment, teachers are more likely to resort to supposedly objective indicators such as grades. Nevertheless, in line with existing evidence and the formal prerequisites for making a recommendation decision in the Swiss context (Batruch et al., 2023; Urhahne & Wijnia, 2021), pre-service teachers in our sample also considered other aspects of student behavior and learning. In addition, even though about 8 % of students in a cohort usually achieve a 6, the best grade, we explicitly opted against including this grade in our experiment because we believed it would be a dominant characteristic: It is almost unthinkable for a student with this grade not to get a recommendation. Nevertheless, one could argue that including a wider range of grades in the experiment would be a stronger test for the assumed competition effect.

Fourth, even though the robustness checks indicated that respondents do not resort to a wider frame of reference, it is possible that teachers in the real world consider not only a small group of motivated students that are actively seeking a recommendation but rather use a wider frame of reference (e.g., students that they deem fit but who themselves do not want to attend the academic track). Likewise, the reference groups in this study are rather small. This implies that the competition measure – the share of peers with a recommendation – can take on only few values. Consequently, the share of peers with a recommendation might suffer from measurement error.

Fifth, the outlined processes are caused both by the structure of the education system (specialized teachers in academic high schools and a fixed number of places in the short run) and the relative comparisons teachers make when deciding on whether a student should be recommended for the highest track. Consequently, our findings only apply to rather static, track-based school settings that feature a recommendation-based promotion to advantageous school tracks (e.g., Austria; Belgium, Germany, Switzerland).

Finally, it is also worthwhile to note that the pre-service teachers in our sample differ considerably in their decisions. This implies that, aside from the abilities and characteristics of an individual student and their peers, a student's educational chances also depend on additional unobserved attributes of the teacher making the decision.

Overall, the finding that pre-service teachers compare students to their classroom peers when making a recommendation, leading them to have a more or less favorable impression of individual students independent of their performance and behavior, and that they fill up slots accordingly, points to an additional mechanism of educational inequalities in tracked education systems (Breen & Goldthorpe, 1997; Esser, 2016; Hallinan, 1992). Since teachers evaluate students' fitness for the highest track in relative instead of absolute terms, their chances of being recommended for the highest track decrease with the number of high-performing and ambitious competitors. Consequently, even high-performing students could be denied the opportunity of attending the highest secondary school track if they are part of a bigger pond with even bigger fish. This would be especially detrimental for minority students who have been shown to benefit the most from more demanding educational programs (Card & Giuliano, 2016).

Furthermore, since educational attainment and achievement are crucially linked to social origin (e.g., Becker, 2003; Stocke, 2007), this relative comparison process can be expected to be especially detrimental for disadvantaged students facing more privileged peers. In this regard, recent research by Bergold & Steinmayr (2023) shows pronounced negative effects on students' future development for students who are underestimated by their teachers. This finding is also in line with other studies that detect heterogeneous peer effects on educational achievement, with negative effects for disadvantaged students when surrounded by more privileged peers (e.g., Ammermueller & Pischke, 2009; Batruch et al., 2019; Fruehwirth & Gagete-Miranda, 2019). Irrespective of social origin, however, these reference group effects working through the choices teachers make can be regarded as problematic. They affect educational chances beyond the individual effort, motivation, and even performance levels of students and are thus out of line with the meritocratic ideal. In this regard, increasing the accountability of teachers has

proven to be an effective strategy to decrease such detrimental reference group effects (Leest et al., 2021; Pit-ten Cate et al., 2016).

Research ethics

We obtained informed consent from all individual participants included in the study and refrain from using any identifying information about participants so that participants' confidentiality is adequately protected.

Declaration of Competing Interest

None.

Appendix

Table A1
Descriptive statistics.

		Mean	SD	min	max
Experiment					
	Grade last semester	5.123	0.416	4.5	5.5
	Performance development	1.015	0.814	0.0	2.0
	Participation during class	0.484	0.500	0.0	1.0
	Learning and study habits	0.521	0.500	0.0	1.0
	Methods and ways of working	0.498	0.500	0.0	1.0
	Parental education	0.515	0.500	0.0	1.0
	Parental aspirations	0.496	0.500	0.0	1.0
	Student's gender	0.496	0.500	0.0	1.0
	Subject	0.412	0.492	0.0	1.0
	Group size	4.167	0.800	3.0	5.0
Teacher characteristics					
	Total teaching experience (months)	9.265	12.350	0.0	51.0
	Knowledge recommendation process	2.882	1.132	0.0	4.0
	Female (reference: *male*)	0.515	0.500	0.0	1.0
	Age	24.574	4.339	19	45
	Migration background (reference: *none*)	0.235	0.424	0.0	1.0
	Subjective social status	1.574	0.626	1.0	3.0
N (Teachers)*		816 (68)			

*After listwise deletion of 24 vignettes (2 respondents) with item nonresponse.

Table A2
RE models with interaction effects.

	(4)	(5)
Grade last semester	0.466***	0.459***
	(0.024)	(0.024)
Participation (*Reference: active*)		
not active	-0.078***	-0.080***
	(0.022)	(0.022)
Learning & study habits (*Reference: learns new things easily*)		
needs more time to learn	-0.059**	-0.060**
	(0.022)	(0.022)
Methods & learning style (*Reference: high independence & concentration ability*		
low independence & concentration ability	-0.170***	-0.169***
	(0.023)	(0.022)
Group size	0.060	0.003
	(0.040)	(0.044)
Teaching experience (months)	-0.002	-0.015 * *
	(0.002)	(0.005)
Share others with a recommendation	0.240	0.061
	(0.214)	(0.223)
Group size*share with recommendation	-0.144*	-0.104+
	(0.057)	(0.058)
Group size*teaching experience (months)		0.003**
		(0.001))
Teacher: knowledge	✓	✓
Teacher: background characteristics	✓	✓
η_j	2.972***	3.275***
	(0.947)	(1.025)
N (groups)	816 (68)	

Note: Average Marginal Effects (AME); standard errors in parentheses.
+p < .10. *p < .05. **p < .01. ***p < .001.

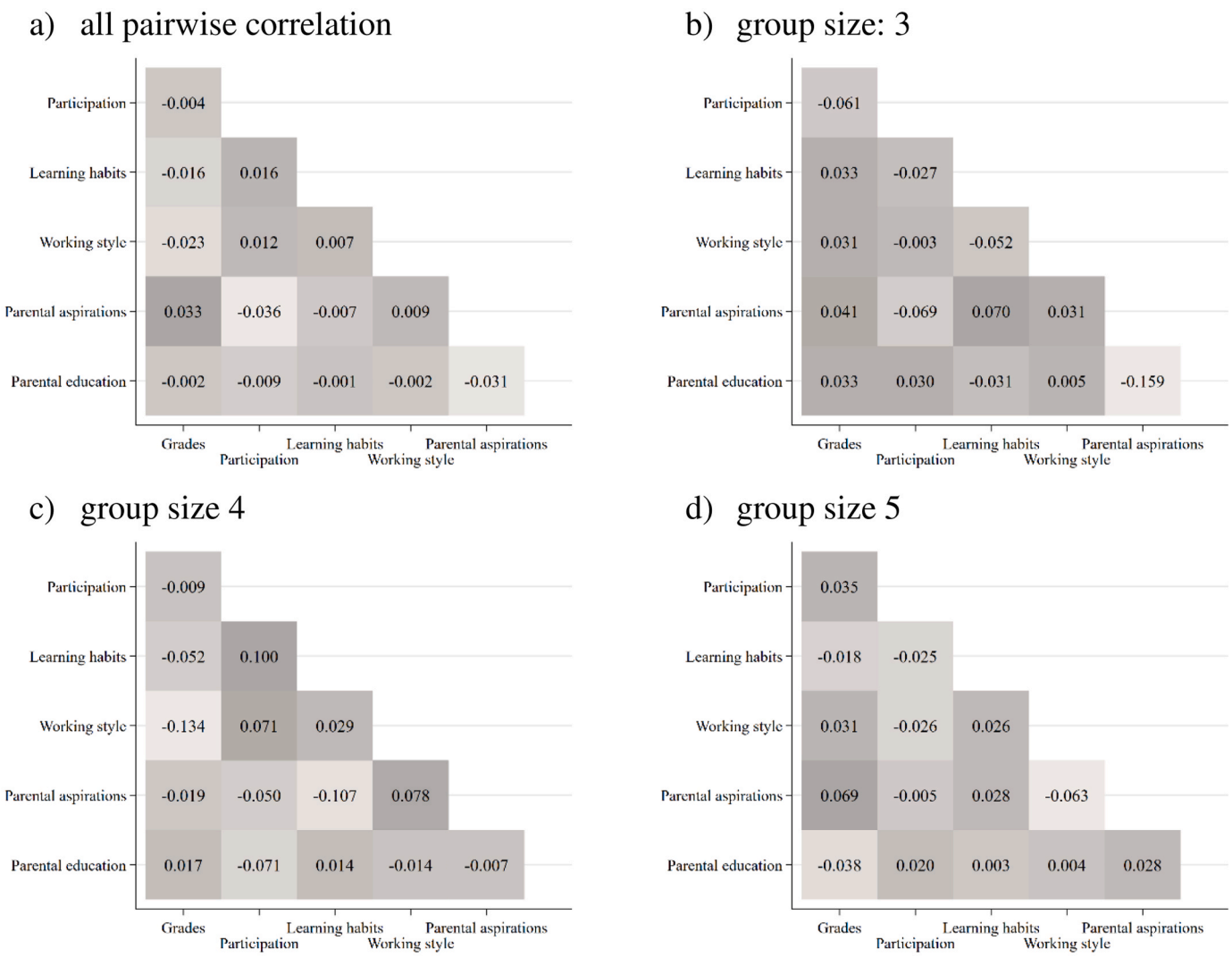

Figure A.1. All pairwise correlations of dimensions in the experiment.

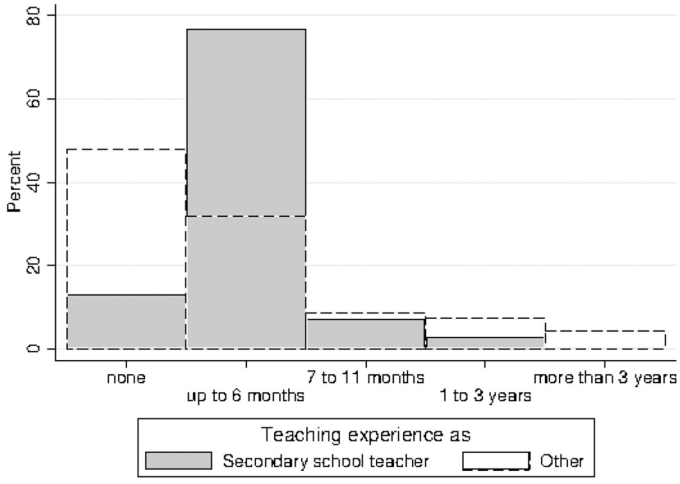

Figure A.2. Respondents' teaching experience.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.rssm.2023.100869.

References

Alexander, K., & Eckland, B. K. (1975). Contextual effects in the high school attainment process. *American Sociological Review, 40*(3), 402. https://doi.org/10.2307/2094466

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science, 21*(2), 174–177.

Allison, P. D. (2009). *Fixed effects regression models*. SAGE.

Ammermueller, A., & Pischke, J. (2009). Peer effects in european primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics, 27*(3), 315–348. https://doi.org/10.1086/603650

Andersen, I. G. (2018). Pygmalion in instruction? Tracking, teacher reward structures, and educational inequality. *Social Psychology of Education, 21*(5), 1021–1044. https://doi.org/10.1007/s11218-018-9452-z

Baeriswyl, F., Wandeler, C., & Trautwein, U. (2011). Auf einer anderen Schule oder bei einer anderen Lehrkraft hätte es für's Gymnasium gereicht": Eine Untersuchung zur Bedeutung von Schulen und Lehrkräften für die Übertrittsempfehlung. *Zeitschrift Für Pädagogische Psychologie, 25*(1), 39–47. https://doi.org/10.1024/1010-0652/a000028

Batruch, A., Autin, F., Bataillard, F., & Butera, F. (2019). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin, 45*(3), 477–490. https://doi.org/10.1177/0146167218791804

Batruch, A., Geven, S., Kessenich, E., & van de Werfhorst, H. G. (2023). Are tracking recommendations biased? A review of teachers' role in the creation of inequalities in tracking decisions. *Teaching and Teacher Education, 123*, Article 103985. https://doi.org/10.1016/j.tate.2022.103985

Becker, R. (2003). Educational expansion and persistent inequalities of education: Utilizing subjective expected utility theory to explain increasing participation rates in upper secondary school in the Federal Republic of Germany. *European Sociological Review, 19*(1), 1–24. https://doi.org/10.1093/esr/19.1.1

Bergold, S., & Steinmayr, R. (2023). Teacher judgments predict developments in adolescents' school performance, motivation, and life satisfaction. *Journal of Educational Psychology, 115*(4), 642–664. https://doi.org/10.1037/edu0000786

Bergold, S., Weidinger, A. F., & Steinmayr, R. (2022). The "big fish" from the teacher's perspective: A closer look at reference group effects on teacher judgments. *Journal of Educational Psychology, 114*(3), 656–680. https://doi.org/10.1037/edu0000559

Boone, S., Thys, S., Van Avermaet, P., & Van Houtte, M. (2018). Class composition as a frame of reference for teachers? The influence of class context on teacher recommendations. *British Educational Research Journal, 44*(2), 274–293. https://doi.org/10.1002/berj.3328

Boone, S., & Van Houtte, M. (2013). Why are teacher recommendations at the transition from primary to secondary education socially biased? A mixed-methods research. *British Journal of Sociology of Education, 34*(1), 20–38. https://doi.org/10.1080/01425692.2012.704720

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society, 9*(3), 275–305. https://doi.org/10.1177/104346397009003002

Bresee, C. W. (1976). On "grading on the curve. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 50*(3), 108–110. https://doi.org/10.1080/00098655.1976.9956921

Calsamiglia, C., & Loviglio, A. (2019). Grading on a curve: When having good peers is not good. *Economics of Education Review, 73*, Article 101916. https://doi.org/10.1016/j.econedurev.2019.101916

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Card, D., & Giuliano, L. (2016). Can tracking raise the test scores of high-ability minority students? *American Economic Review, 106*(10), 2783–2816. https://doi.org/10.1257/aer.20150484

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology, 34*(3), 366–375.

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed- and random-effects models: Some considerations for policy-relevant education research. *Education Economics, 23*(3), 259–277. https://doi.org/10.1080/09645292.2013.855705

CSRE, S. C. C. for R. in E. (2018). *Swiss Education Report*. CSRE.

Dardanoni, V., Modica, S., & Pennsi, A. (2009). Grading across schools. *The B E Journal of Economic Analysis & Policy, 9*(1). https://doi.org/10.2202/1935-1682.2065

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72*(1), 17–31.

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review, 47*(2), 189–201. https://doi.org/10.2307/2094962

Dockx, J., De Fraine, B., & Vandecandelaere, M. (2019). Tracks as frames of reference for academic self-concept. *Journal of School Psychology, 72*, 67–90. https://doi.org/10.1016/j.jsp.2018.12.006

Dumont, H., Klinge, D., & Maaz, K. (2019). The many (Subtle) ways parents game the system: Mixed-method evidence on the transition into secondary-school tracks in Germany. *Sociology of Education, 92*(2), 199–228. https://doi.org/10.1177/0038040719838223

Dumont, H., & Ready, D. D. (2020). Do schools reduce or exacerbate inequality? How the associations between student achievement and achievement growth influence our understanding of the role of schooling. *American Educational Research Journal, 57*(2), 728–774. https://doi.org/10.3102/0002831219868182

Epple, D., & Romano, R. E. (2011). Peer effects in education: A survey of the theory and evidence. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *Handbook of social economics, 1B* pp. 1053–1163). Elsevier.

Erziehungsdirektion des Kantons Bern. (2017). Gemeinsame Grundlagenformulare GYM1, FMS, WMS, BM 1, IMS. *Aufnahmeverfahren für Schulbeginn ab August 2018*. Erziehungsdirektion des Kantons Bern.

Esser, H. (2016). Sorting and (much) more: Prior ability, school effects and the impact of ability tracking on educational inequalities in achievement. S. 95–114 in Education systems and inequalities. International comparisons, hrsg. von Andreas Hadjar und Christiane Gross. In A. Hadjar, & C. Gross (Eds.), *Education systems and inequalities: International comparisons* (pp. 95–114). Bristol: Policy Press.

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics, 35*(2), 387–428. https://doi.org/10.1086/689472

Fruehwirth, J. C., & Gagete-Miranda, J. (2019). Your peers' parents: Spillovers from parental education. *Economics of Education Review, 73*(Online First), 1–14. https://doi.org/10.1016/j.econedurev.2019.101910

FSO, F.S. O. (2020). *Ausbildungswahl auf der Sekundarstufe II – Daten des Indikators*. FSO. ⟨https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungssystem.assetdetail.12527170.html⟩.

Gibbons, S., & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in England. *Oxford Bulletin of Economics and Statistics, 78*(4), 548–575. https://doi.org/10.1111/obes.12095

Gilgen, S., & Stocker, M. (2022). Discrimination at the crossroads? Evidence from a factorial survey experiment on teacher's tracking decisions. *Swiss Journal of Sociology, 48*(1), 77–105. https://doi.org/10.2478/sjs-2022-0005

Gröhlich, C., & Guill, K. (2009). How stable are reference group effects of secondary school track recommendations? *Journal for Educational Research Online, 1*(1), 154–171.

Hallinan, M. T. (1992). The organization of students for instruction in the middle school. *Sociology of Education, 65*(2), 114–127. https://doi.org/10.2307/2112678

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527–544. https://doi.org/10.1002/jae.741

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313. https://doi.org/10.2307/1170184

Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 18*(4), 307–317.

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156–170. https://doi.org/10.1037/a0015558

Hupka-Brunner, S., Sacchi, S., & Stalder, B. E. (2010). Social origin and access to upper secondary education in Switzerland: A comparison of company-based apprenticeship and exclusively school-based programmes. *Swiss Journal of Sociology, 36*(1), 11–31.

Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review, 102*(5), 2048–2082. https://doi.org/10.1257/aer.102.5.2048

Klapproth, F., & Fischer, B.D. (2018). Preservice teachers' evaluations of students' achievement development in the context of school-track recommendations. *European Journal of Psychology of Education*, Online First. https://doi.org/10.1007/s10212–018-0405-x.

Klapproth, F., & Fischer, B. D. (2020). Achievement development is less important for school-placement recommendations when students are stereotyped. *Social Psychology of Education, 23*(6), 1483–1504. https://doi.org/10.1007/s11218-020-09593-9

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal, 122*(559), 208–237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Leest, A., Hornstra, L., Tartwijk, J., & Pol, J. (2021). Test- or judgement-based school track recommendations: Equal opportunities for students with different socio-economic backgrounds? *British Journal of Educational Psychology, 91*(1), 193–216. https://doi.org/10.1111/bjep.12356

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280–295.

Marsh, H. W., & Hau, K.-T. (2003). Big-Fish—Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective

schools. *American Psychologist, 58*(5), 364–376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*(1), 213–231.

Pit-ten Cate, I. M., Krolak-Schwerdt, S., & Glock, S. (2016). Accuracy of teachers' tracking decisions: Short- and long-term effects of accountability. *European Journal of Psychology of Education, 31*(2), 225–243. https://doi.org/10.1007/s10212-015-0259-4

Preckel, F., & Brüll, M. (2010). The benefit of being a big fish in a big pond: Contrast and assimilation effects on academic self-concept. *Learning and Individual Differences, 20*(5), 522–531. https://doi.org/10.1016/j.lindif.2009.12.007

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335–360. https://doi.org/10.3102/0002831210374874

Richer, S. (1976). Reference-group theory and ability grouping: A convergence of sociological theory and educational research. *Sociology of Education, 49*(1), 65. https://doi.org/10.2307/2112394

Rosenqvist, E. (2018). Two functions of peer influence on upper-secondary education application behavior. *Sociology of Education, 91*(1), 72–89. https://doi.org/10.1177/0038040717746113

Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology, 108*(6), 883–897. https://doi.org/10.1037/edu0000085

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In M. Ryan, K. Gerard, & M. Amaya-Amaya (Eds.), *Using discrete choice experiments to value health and health care* (Vol. 11, pp. 13–46). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5753-3_1.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. In *Handbook of the Economics of Education* (Vol. 3, pp. 249–277) Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00004-1

Sewell, W. H., & Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review, 34*(1), 82–92.

Sommet, N., Weissman, D. L., Cheutin, N., & Elliot, A. J. (2023). How many participants do i need to test an interaction? Conducting an appropriate power analysis and achieving sufficient power to detect an interaction. *Advances in Methods and Practices in Psychological Science, 6*(3). https://doi.org/10.1177/25152459231178728

Stocke, V. (2007). Explaining educational decision and effects of families' social class position: An empirical test of the breen goldthorpe model of educational attainment. *European Sociological Review, 23*(4), 505–519. https://doi.org/10.1093/esr/jcm014

Street, D. J. (2007). *The construction of optimal stated choice experiments* (1st ed.). Wiley & Sons.

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A further examination of the big-fish–little-pond effect: Perceived position in class, class size, and gender comparisons. *Sociology of Education, 83*(4), 333–345. https://doi.org/10.1177/0038040710383521

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*(5), 483–504. https://doi.org/10.1016/S0883-0355(03)00016-8

Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology, 85*(4), 459–478. https://doi.org/10.1111/bjep.12087

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review, 32*, Article 100374. https://doi.org/10.1016/j.edurev.2020.100374

von Keyserlingk, L., Becker, M., & Jansen, M. (2019). Academic self-concept during the transition to upper secondary school. *Contemporary Educational Psychology, 56*, 152–160. https://doi.org/10.1016/j.cedpsych.2019.01.001

Wilkinson, I. A. G. (2002). Introduction: Peer influences on learning: where are they? *International Journal of Educational Research, 37*(5), 395–401. https://doi.org/10.1016/S0883-0355(03)00012-0

Zangger, C. (2018). Bringing space into the equation: Modelling the social and spatial interdependence of neighborhood effects on educational outcomes. *Research in Social Stratification and Mobility, 55*, 63–74.

Zell, E., & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology, 97*(3), 467–482. https://doi.org/10.1037/a0015453

Zimmerman, D. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics, 85*(1), 9–23.

P.H. Bern. (2016). *Studienbereich des Studiengangs Sekundarstufe I. Erziehungs- und Sozialwissenschaften*. PH Bern, Institute for Secondary Level I. ⟨https://www.phbern.ch/sites/default/files/2019–09/27102016_Bereichsflyer_ESW_IS1.pdf⟩.