Journal of Applied Developmental Psychology 66 (2020) 101089

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

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

Differential associations of school practices with achievement and sense of belonging of immigrant and non-immigrant students

Jia He[a,b,*], Jessica Fischer[a]

[a] DIPF | Leibniz Institute for Research and Information in Education, Germany
[b] Tilburg University, the Netherlands

ARTICLE INFO

Keywords:
Immigrants
Ability grouping
Grading practices
Extracurricular activities
Multilevel

ABSTRACT

We are interested in identifying "malleable" school and classroom practices to enhance immigrant students' learning. Using PISA 2015 data from Germany, Italy, and Spain we test the differential associations of school–level practices with achievement and sense of belonging at school for students with and without an immigrant background. We found that (1) in-school ability grouping was invariably, negatively related to achievement of both student groups, and the effects were stronger for immigrant than nonimmigrant students; (3) grading based on "hard" factors was not related to achievement, but it showed differential associations with sense of belonging in Germany; (4) grading based on "soft" factors and provision of extracurricular activities also showed mixed associations with the outcomes across countries and did not fulfil the potential to enhance immigrant students' outcomes. We discuss these findings and implications.

Introduction

In recent years, the number of immigrants, broadly referring to individuals who were born outside their country of residence (i.e., first-generation immigrants) and individuals with at least one parent born outside their country of residence (i.e., second-generation immigrants), increased on an unprecedented scale across the globe. Schools, a vital developmental context for immigrant youth to integrate and prepare for the workforce, are challenged to respond to the diverse influx of students and facilitate their learning and flourishment. Schools have the potential to compensate for often existing disadvantages of immigrant students[1] and to reduce the achievement gap between immigrant and nonimmigrant students. Research has highlighted school factors such as teachers' expectations and competence, classroom and school climate, and between-school tracking as being relevant for immigrant students' learning outcomes (Schachner, Juang, Moffitt, & van de Vijver, 2018). Other malleable, policy-amenable school factors aiming at enhancing student learning outcomes are still to be identified and promoted for immigrant education. Even though many instructional practices to enhance student outcomes are documented to be effective, it is rarely investigated if they operate to enhance or frustrate the learning experience of immigrant and nonimmigrant students in the same way. It is important to understand if and how school and classroom practices exacerbate or reduce inequalities to effectively foster learning outcomes of immigrant students.

Drawing on the dynamic model of educational effectiveness (Creemers & Kyriakides, 2006), we aim to shed light on how in-school ability grouping, grading practices, and the provision of extracurricular activities are related to the outcomes of immigrant compared to nonimmigrant students in multiple European countries. Substantive contributions of our study lie in (1) the expanded scope of student outcomes (academic achievement and sense of belonging at school), (2) the unfolding of associations of these practices with the learning outcomes for immigrant and nonimmigrant students, and (3) the application of an innovative multilevel analysis (i.e., students nested in schools) with a grouping variable at student level (i.e., immigrant background), which allows us to study the differential associations of school-level practices with outcomes of immigrant and nonimmigrant students in their intact ecology of school settings.

In the following, we first describe immigrant students' learning context in popular European destination countries, which leads to a selection of three countries for our study, namely, Germany, Italy, and Spain. We then introduce the dynamic model of educational effectiveness in the context of immigrant education, which guides the theoretical and methodological considerations of a differential impact of school and classroom practices for immigrant compared to nonimmigrant students. Next, we review studies on effects of the targeted school-level practices and highlight expected differences for immigrant

* Corresponding author at: Rostocker Strasse 6, Frankfurt 60323, Germany.
E-mail address: Jia.he@dipf.de (J. He).

[1] In this paper we use the term "immigrant students" to refer to students with an immigrant background.

https://doi.org/10.1016/j.appdev.2019.101089
Received 22 March 2019; Received in revised form 1 October 2019; Accepted 3 November 2019
Available online 10 November 2019
0193-3973/

and nonimmigrant students.

Immigrant education in Germany, Italy, and Spain

Following the surge of humanitarian migrants, the number of immigrants has increased in Europe in the recent years. Popular destinations for immigrants are the UK, France, Germany, Italy, and Spain (Migration Policy Institute, 2015). According to the OECD's indicators of immigrant integration (OECD, 2018), these destination countries can be grouped based on similar characteristics of received immigrants and integration challenges related to these characteristics. The UK, Germany, and France are typical, long-standing destination countries, with Germany and France receiving many low-educated immigrants (in contrast to the UK). Italy and Spain fall into the category of new receiving countries, and similar to Germany and France, a large share of immigrants has a rather low level of education. Of these European countries, academic achievement and sense of belonging of immigrant students is lower than nonimmigrants particularly in Germany, Italy, and Spain (OECD, 2015a), indicating a strong need to close the gap between these groups. Compared to the UK and France, immigrants who have been educated in Germany, Spain, and Italy are rarely among the highly educated (OECD, 2018), suggesting differential effects of school and classroom practices for immigrant and nonimmigrant students in these countries. School and classroom practices aiming at the enhancement of students' outcomes have the potential to narrow the gap between nonimmigrant and immigrant students and are particularly relevant in these countries.

Despite the similar challenges faced by Germany, Italy, and Spain to integrate mainly low-educated immigrants and similarities with regard to their integration policy climate (all three ranked as being slightly to halfway favorable for immigrants) (Huddleston, Bilgili, Joki, & Vankova, 2015), these countries differ with regard to their educational systems (e.g., OECD, 2019; Wößmann, 2009). In terms of between-school tracking, Germany starts tracking students when they are 10 years old in comparison to 14 and 16 years old in Italy and Spain, respectively. This produces different school compositions or profiles (e.g., immigrants being more densely clustered together in Germany than in Italy and Spain), posing different challenges for schools with a high proportion of immigrant students compared to schools with a comparably low proportion of immigrant students. Another difference is the readiness and competence of schools and teachers to manage cultural diversity: According to the 2018 Teaching and Learning International Survey (TALIS) of nationally representative secondary school teachers (OECD, 2019), only 52% of the Spanish teachers reported that they feel they can cope with the challenges of a multicultural classroom ‘quite a bit’ or ‘a lot’ in teaching a culturally diverse class, in comparison to the OECD average of 67% and 80% in Italy. These differences may result in a nuanced picture of the effects of school and classroom practices on students' outcomes across these three countries.

The dynamic model of educational effectiveness

Models of educational effectiveness to guide school improvement and better student outcomes have come a long way: built on an early atheoretical search of what accounts for student achievement, educational production models and organizational models have emerged to explain various student outcomes from student-, class-, school-, and system-level antecedents. The dynamic model developed by Creemers and Kyriakides (2006) maintains that student outcomes are influenced by multilevel factors (as in other educational production and organizational models), and it further proposes that relationships between factors of different levels are complex. Thus it is necessary to 1) broaden the scope of student outcomes, 2) acknowledge the nonlinear relationship with some factors, 3) investigate the interactive effects of factors at different levels, and 4) improve the measurement of these factors along five dimensions: frequency (i.e., the quantity that an activity is present in a system, school, or classroom), focus (i.e., the specificity and purpose of an activity), stage (i.e., the phase of an activity, with the assumption that the activity needs to take place for a long period of time to accumulate effects on student learning), quality (i.e., properties of the activity and its optimal use), and differentiation (i.e., the extent to which the activity is implemented for and has impact on all subjects in the same way). Thus, these dimensions aim to capture not only quantity but also quality and processes of educational effectiveness.

In the context of immigrant education, two aspects of the dynamic model are particularly important. First, the differentiation dimension of the measurement, referring to the extent to which effectiveness factors are implemented in the same way for all students irrespective of their background, points to certain school and classroom practices being differently applied to immigrant and nonimmigrant students. For instance, teachers' academic expectations towards students are associated with better student outcomes. Yet, teachers tend to have lower academic expectations towards immigrant students in comparison to nonimmigrant students, which can be detrimental to the former group's learning outcomes (Timmermans, Kuyper, & van der Werf, 2015). Subsequently, changing the biased expectations from teachers can be subjected to interventions. Second, the interaction of factors from different levels, especially interactions across levels (e.g., immigrant background and classroom/school-level practices), is important to reveal, given that individual differences in family background, prior knowledge, social-emotional skills, and culturally filtered ways of thinking and working can condition students' readiness to gain from school instructional practices (Creemers & Kyriakides, 2006, 2008). Even though generic quality-enhancing instruction targets all students, intended effects may be inconsistent for different student groups, and these effects may further be moderated by national contexts (e.g., Marks, McKenna, & Coll, 2018; Shapira, 2012). These two aspects imply differential effectiveness of instructional practices for immigrant and nonimmigrant students and the need to identify an optimal combination of factors to maximize the effectiveness for immigrant education.

School and classroom practices and student outcomes

Building on the dynamic model in the context of immigrant education, we are interested in unfolding the associations of generic school and classroom instructional practices and student outcomes with a special focus on the interactions between immigrant background and school factors.

Student outcomes

So far, the main outcome indicator for instructional quality and effectiveness is students' academic achievement. Yet, in response to the multiple nature and goals of education, general meta-cognitive skills and nonacademic outcomes (e.g., motivation) are increasingly recognized to be equally, if not even more important, to help students succeed in school and society in the long term (e.g., Kuger, Klieme, Jude, & Kaplan, 2016). It is acknowledged that multiple outcomes may be complementary or competitive, and that they are related to similar or different instructional practices (Sammons, 1996). In line with the belongingness hypothesis, *sense of belonging at school,* defined as the extent to which students feel accepted, respected, and connected to other students and teachers, can be perceived as a prerequisite for overall school functioning (Baumeister & Leary, 1995; Maslow, 1962). It is a powerful indicator of how immigrant students are integrating into their new surroundings, for both short- and long-term adaptations (C.-Y. Chiu, Walter, David, & Colleen, 2013). Students with a higher sense of belonging at school often show higher academic achievement, intrinsic motivation, wellbeing (Roeser, Midgley, & Urdan, 1996) as well as reduced rates of school dropouts and social rejection (M. M. Chiu, Pong, Mori, & Chow, 2012). However, compared to nonimmigrant students, immigrant students often are less likely to feel that they belong at school. Additionally, educational systems vary in their ability to strengthen immigrant student's feelings of belongingness at school. Hence, it is vital to consider sense of belonging at school as a nonacademic outcome besides achievement.

School and classroom practices

High-quality teaching has the potential to maximize student outcomes, yet, may not be equally effective for all students. With German panel data, Atlay, Tieben, Hillmert, and Fauth (2019) demonstrated that classroom management is positively associated with student performance regardless of socio-economic status (SES), while cognitive activation and supportive climate are more beneficial for students with a high SES than those from low SES families. Thus, SES even magnified the achievement gap. Yet, research to date has hardly investigated the associations between high-quality practices and immigrant background, which we seek to remedy. Among a myriad of school and classroom practices, we target within-school ability grouping, teachers grading practices based on hard and soft criteria, and provisions of extracurricular activities. We selected these practices as they reflect three common components of instructional quality, namely, grouping procedures, teacher behaviors, and teaching materials (Creemers & Kyriakides, 2006). They involve students directly and can be readily intervened at decentralized levels (i.e., school and classroom) and have the potential to narrow the gap between immigrant and nonimmigrant students' learning outcomes. In this section, we formulate hypotheses with regard to their expected associations with nonimmigrant and immigrant students' academic achievement and sense of belonging at school.

Within-school ability grouping

Ability grouping refers to policies and practices that sort students within classrooms as well as between classrooms and school types according to criteria such as previous grades, teachers' recommendations, or standardized tests (Ammermueller, 2005). Narrowing the range of students' abilities can help the teacher to align the level, pace, and practices of instruction to students' needs more closely, which fosters student's achievement, interest, and motivation (Bygren, 2016; Field, Kuczera, & Pont, 2007). To date, studies on ability grouping have mainly focused on between-school ability grouping and suggest differential effects for different student groups: Within high-ability groups, social comparisons with high-achieving peers can positively affect achievement; whereas less able students within low-achieving groups lose the opportunity to benefit from positive peer effects (Marsh, 1991). Additionally, varying instructional quality and differential expectations between different ability groups can widen achievement gaps. For instance, students' in high-achieving groups often experience a higher-quality learning environment (Hattie, 2009). Similarly, Belfi, Fraine, Goos, and van Damme (2012) reported a positive association of ability grouping and sense of belonging for high but not for low performing student groups. Ability grouping is often accompanied by stereotyping and student expectations, which can influence relationships with peers and teachers. Compared to students in high-achieving groups, students within low-achieving groups are often viewed more negatively, which can lead to a lower sense of school belonging for the latter group (Ireson & Hallam, 2005). Immigrant students are significantly more often placed into groups with lower average achievement levels than nonimmigrant students (even with similar achievement levels as nonimmigrant students and after controlling for the SES) (Caro, Lenkeit, Lehmann, & Schwippert, 2009). Thus, ability grouping often segregates immigrants within low-achieving groups, which increases the immigrant-native achievement gap (Bygren, 2016).

To summarize, research has mainly focused on between-school ability grouping, yet, we expect to find similar patterns for within-school ability grouping. As highlighted above, studies on ability grouping point towards negative effects for students within low-achieving groups compared to students within high-achieving groups. As immigrant students are more often grouped into low-achieving groups regardless of SES and prior achievement, we expect with*in-school* ability grouping to be more negatively associated with achievement (Hypothesis 1a) and sense of belonging at school (Hypothesis 1b) for immigrant compared to nonimmigrant students.

Teachers' grading practices based on "hard" and "soft" factors

Grading is the assignment of symbolic numbers, letters, or terms for making an end-point judgement about students' performance (Tomlinson, 2005). Grades can be used for academic purposes, such as making decisions on track placements and grade retention (e.g., van Ewijk, 2011), and can also serve as socio-emotional support by rewarding students' motivation, self-efficacy, and interest to foster long-term school outcomes (e.g., Betts & Grogger, 2003). Accordingly, teachers may base their grading on different criteria and sources of evidence (Bayer, Klieme, & Jude, 2017). Two types of grading can be distinguished: (1) grading based on "hard" factors such as pre-established, clearly stated, content-specific learning goals at the national or regional level and comparisons of academic performance among students, and (2) "soft" factors such as students' progress, efforts, or metacognitive outcomes (e.g. performance in collaborative problem solving). Yet, the differential effects of these two types of grading on learning outcomes are hardly researched to date.

As grading based on "hard" factors serves the excellence focus (for high stakes decisions and competitiveness), we expect grading based on "hard" factors to be especially associated with all students' academic achievement (Hypothesis 2a). In contrast, grading based on "soft" factors may be beneficial to enhance students' motivation and sense of belonging at school as not only academic achievement but also efforts and personal qualities and skills are rewarded (Hypothesis 2b). However, this does not indicate that each type of grading is only associated with one outcome as hypothesized; we investigate their effects on both outcomes.

Moreover, grading based on "soft" factors also allows taking different prerequisites of students (e.g., knowledge of the school system and language skills) into account, and thus, seems extremely important for nonacademic outcomes for immigrant students (Motti-Stefanidi & Masten, 2013; Schachner et al., 2018). Thus, we expect grading based on "soft" factors to be more positively associated with sense of belonging at school for immigrant compared to nonimmigrant students (Hypothesis 2c).

Extracurricular activities

School-based extracurricular activities are designed to promote a positive academic and mental development of students by providing physical and psychological safety, structure, supportive relationships, opportunities to belong, and skill building (Farb & Matjasko, 2012). Besides boosting academic achievement, the emphasis of extracurricular activities on facilitating high-quality social interactions, was found to be beneficial for nonacademic outcomes, such as sense of belonging, and to reduce problem behavior and the likelihood of dropping out (Brown & Evans, 2016; Fredricks & Eccles, 2006; Mahoney, Cairns, & Farmer, 2003; McNeal, 1998). Positive effects of extracurricular activities have been found for youth from diverse backgrounds (Jiang & Peguero, 2016). Thus, for all students, we expect the quantity and diversity of extracurricular activities to be positively associated with achievement (Hypothesis 3a) and sense of belonging at school (Hypothesis 3b). Yet, extracurricular activities are expected to be especially beneficial for immigrant students. As postulated in the Social Control Theory, repeated involvement in extracurricular activities can bridge the gap in social capital needed to succeed in school (Hirschi, 1972). Besides providing additional experiences that contribute to academic achievement (e.g., receiving knowledge of the school system which their parents might not have), interactions with peers, academic mentors, or the larger community can mitigate the relative lack of social capital and to help adjust and integrate into the new educational surroundings. This is particularly the case for first-generation immigrant youth, over and above self-selection effects (Camacho & Fuligni, 2015; Im, Hughes, Cao, & Kwok, 2016). Thus, the positive associations between quantity of extracurricular activities and student achievement (Hypothesis 3c) and sense of belonging at school (Hypothesis 3d) are expected to be stronger for immigrant than non-immigrant students.

We test all hypotheses for Germany, Italy, and Spain. Despite some differences in their educational systems, there is not sufficient evidence in the literature to formulate culture-specific hypotheses.

Individual background and school-compositional characteristics as covariates

Previous research has repeatedly reported correlates of students' outcomes with individual background variables including the *family SES* (e.g., SES is positively related to academic achievement and sense of belonging at school) and the students' *gender* (e.g., males tend to perform better than females in science) (e.g., Le, 2009). With regard to important school compositional characteristics, the *percentage of low SES intake* and *language diversity* showed negative associations with students' achievement (e.g., Aikens & Barbarin, 2008; Langenkamp & Carbonaro, 2018). Their association with sense of belonging is rather unclear (e.g., Opdenakker & Van Damme, 2000). We include these covariates in our analyses as less malleable factors with an expected impact on learning outcomes.

Method

We made use of data from the Programme for International Student Assessment (PISA) in 2015 to investigate the associations of practices at school and classroom level with learning outcomes of immigrant and nonimmigrant students in Germany, Italy, and Spain.

Sample

PISA assesses competencies of 15-year-old students in reading, mathematics, and science in various countries and economies triennially. With a two-stage random sampling of schools (stage 1) and students (stage 2) in PISA, selected school principals fill out a school questionnaire which provides information on school context, leadership, management, and practices. Selected students take a context questionnaire and a subset of the cognitive test of different combinations. In 2015, there was an optional module to involve teachers who teach in the modal grade of 15-year-old students to be surveyed on their beliefs, attitudes, and practices in participating schools. A total number of 25 teachers were randomly sampled within each school, and science and general teachers were distinguished to answer partially overlapping questions. This module was taken up by 19 countries. The questionnaires, data, manual, and the assessment frameworks are publically available for research use (OECD, 2015b, 2017).

We combined data from students, teachers, and principals in our study. In Germany, the sample included 5691 students (with 967 immigrants: 215 first-generation and 752 seconds-generation), 3359 teachers, and 256 school principals. In Italy, there were 11,232 students (899 immigrants: 526 first-generation and 373 seconds-generation), 4300 teachers, and 474 school principals. In Spain, there were 6577 students (681 immigrants: 556 first-generation and 125 seconds-generation), 2264 teachers, and 201 school principals.

Measures

Targeted constructs were from student, teacher, and principal responses.

Student outcomes

Achievement in Science in the cognitive assessment served as the academic outcome. In PISA, each student was administered only a subtest of the overall cognitive test to minimize test burden. By systematically varying items across student groups and using item response theory (IRT), these cognitive data were then scaled in a generalized partial credit model and student ability was estimated as plausible values. Plausible values are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured. Ten plausible values of science achievement for each student were produced and standard analyses with science achievement are to be performed on each of the plausible values.

Sense of Belonging at School is a self-reported six-item measure (sample item: “I make friends easily at school”) with a four-point Likert scale with response options ranging from “*strongly agree*” to “*strongly disagree*”. The internal consistency was adequate in all three countries (with values of Cronbach's Alpha of 0.85, 0.81, and 0.88 in Germany, Italy, and Spain, respectively). The construct was scaled in the PISA international calibration, with the generalized partial credit modeling across countries. The invariance of item parameters was checked and country- and language-specific item parameters freed in case of non-

Table 1
Model fit of the measurement invariance testing.

Scale	Model	χ^2	df	CFI	RMSEA
Sense of belonging (6 groups across all three countries)	Configural	1025.39**	36	0.97	0.08
	Metric	1189.42**	61	0.96	0.07
	Scalar	2970.49**	86	0.90	0.09
Sense of belonging (2 groups in Germany)	Configural	389.92**	12	0.95	0.11
	Metric	409.36**	17	0.95	0.09
	Scalar	485.56**	22	0.94	0.09
Sense of belonging (2 groups in Italy)	Configural	542.74**	12	0.96	0.09
	Metric	556.84**	17	0.96	0.08
	Scalar	636.52**	22	0.95	0.07
Sense of belonging (2 groups in Spain)	Configural	89.06**	12	0.99	0.04
	Metric	97.41**	17	0.99	0.04
	Scalar	116.22**	22	0.99	0.04
Teacher Grading (across 3 countries)	Configural	1574.72**	78	0.92	0.08
	Metric	1907.33**	92	0.90	0.08
	Scalar	4011.69**	106	0.79	0.11

Note. CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation.
** $p < 0.01$.

invariance. The IRT scaling produced weighted least square estimates for the latent dimension, which subsequently were transformed to scales with a mean of 0 and a standard deviation of 1 across OECD countries (OECD, 2017). A higher score indicates a higher level of sense of belonging at school. To ensure that mean scores are comparable across the immigrant and nonimmigrant groups in each country, we checked its measurement invariance in a multigroup confirmatory factor analysis. Three levels of measurement invariance, namely configural (i.e., the same construct being measured across groups), metric (i.e., same factor loadings across groups), and scalar invariance (i.e., same factor loadings and item intercepts across groups) were specified and tested (van de Vijver & Leung, 1997). Model fit was evaluated by the Comparative Fit Index (CFI: above 0.90 considered acceptable) and the Root Mean Square Error of Approximation (RMSEA: below 0.06 considered acceptable) (Hu & Bentler, 1999). The acceptance of a more restrictive model was based on the change of CFI and RMSEA values (change within 0.01 acceptable) (Cheung & Rensvold, 2002). All items loaded on the sense of belonging at school factor and the residuals of all negatively-worded items were correlated to control for the effects of item keying. The model fit summary is presented in Table 1. Within each country, scalar invariance was supported, whereas collapsing the countries for a six-group comparison lacked scalar invariance. This warranted valid mean comparisons of nonimmigrant and immigrant students within each country, but not across countries.[2]

Student background

Gender was recoded with 0 as male and 1 as female. Students' economic, social, and cultural status (*ESCS*) was a composite index scaled by the OECD, which consists of three sub-components: the parents' highest occupational status, the parents' highest educational level (in years of education) and the index of home possessions. *Immigrant background* was recoded with 0 for nonimmigrants and 1 for immigrants (both first-and second-generation[3]).

Teachers' grading practices

In the general teacher survey, criteria for grading were measured with 11 items (nine cross-culturally common items, two country-specific items) with response options ranging from 1 "*not at all*" to 4 "*to a large extent*". A principal component analysis of the nine common items revealed the distinction of *grading based on "hard" factors* (three items including "assign final grade based on written national or regional standards"), and *grading based on "soft" factors* (six items including "I recognize students' effort"). Items for these two factors were subjected to a three-group multigroup confirmatory factor analysis (i.e., teachers in the three countries) to check similar structure and metrics of the grading practices from the three countries (see Table 1). The configural model fit well, the metric model fit slightly worse than the configural model (delta CFA = 0.02, delta RMSEA = 0.00), and the scalar model fit significantly worse than the metric model. This indicated that these two types of grading practices and their similar structure were largely supported across countries. We estimated and saved the factor scores of the two factors in the metric invariance model, and computed the school-level mean reliability for each construct in each country. The values of ICC2 for "hard" grading were 0.36, 0.17, and 0.15 in Germany, Italy, and Spain, respectively, and those for "soft" grading were 0.50, 0.73 and 0.41, respectively. These teacher-level factor scores were then aggregated at the school level to indicate the school-level grading practices.

School background and practices

Two within-school *ability grouping practices* were measured in the principal questionnaire. One item concerns school policies to group students into different classes according to ability, and the other concerns grouping by subjects within classes. Both items were answered on a response scale with three response categories 1 "*not for any class (subject)*", 2 "*for some classes (subjects)*", and 3 "*for all classes (subjects)*".

The *provision of extracurricular activities* was measured with a set of activities the school offers to 15-year-old students with responses of 1"*Yes*" or 0 "*No*". As it was not a latent factor-based construct, an index was computed measuring the total number of the activities that occurred at school including i) band, orchestra or choir; ii) school play or school musical; and iii) art club or art activities (OECD, 2017). This index quantified the amount and diversity of extracurricular activities provided at school.

Two compositional characteristics of students on the school level were included. The *school SES intake* was measured by principals' estimated percentage of students from socioeconomic disadvantaged homes. Principals' estimated percentage of students with a heritage language being different from the test language was a proxy of school language diversity.

Results

We report the results in two parts: We first describe the mean differences of the outcomes between immigrant and nonimmigrant students per country, and provide the descriptives of the school-level variables of interest. Then, we report the differential associations of school variables with student outcomes in multiple group multilevel analysis.

Descriptives of target constructs

With the complex sampling scheme of PISA, the final student sampling weights were rescaled to have a population total equal to the sample size and were applied in all analyses. Using the final sampling weights represents the national student population more accurately than the unweighted data, which assumed the student sample is a random sample of the population. The rescaling of the weights aims for proper hypothesis testing, so that the significance is not affected by the much larger population size compared to the sample size. When plausible values of science achievement were modeled as the outcome, multiple imputations were sought to obtain a final set of unbiased estimates (i.e., analyses were performed for each plausible value and combined based on Rubin's rule) (Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

We compared the mean scores of science achievement between immigrant and nonimmigrant students in each country by regressing the plausible values on the dummy-coded immigrant background with multiple imputations in Mplus 7.3 (Muthen & Muthen, 1998-2012). In all three countries, there were significant differences between the two groups, with immigrant students scoring 43.72, 36.18, and 36.29 points lower than nonimmigrant students in Germany (559.22), Italy (522.23), and Spain (534.32), respectively. Given that scalar invariance of *sense of belonging at school* was only achieved for the two groups in each country but not across countries, an ANOVA was performed per country with immigrant background as the grouping variable. In Germany, there was no mean difference between the two groups: $M_{immigrant} = 0.26$ ($SD = 1.12$), and $M_{nonimmigrant} = 0.30$ ($SD = 1.06$), F

[2] We also estimated the factor scores from the metric and scalar invariance models in the six-group confirmatory factor analysis, and both sets of factor scores correlated with the PISA internationally scaled scores at 0.90, indicating a very high convergence. For the sake of greater cross-country comparability and generalizability beyond the targeted groups in this study, we used the PISA internationally scaled scores of sense of belonging.

[3] We acknowledge that effects might vary for first and second generation immigrants. Yet, for our study, we combined first and second generation immigrants based on the following reasons: 1) statistical reasons (too small sample sizes for both groups, leading to non-convergence of models), 2) the effects we found were rather similar for both groups, 3) school can be seen as public and social domain in acculturation, thus, both generations of immigrants need to adapt (given their minority status as immigrants).

Table 2
Descriptives of school-level variables.

	Germany		Italy		Spain	
	Mean	SD	Mean	SD	Mean	SD
Percentage Low SES intake	26.59	24.69	14.04	16.61	18.25	20.68
Proportion of students with a different heritage language	19.83	22.716	10.83	15.01	19.62	26.86
Ability Grouping by Class	1.38	0.636	1.18	0.512	1.45	0.60
Ability Grouping by Subject	1.53	0.60	1.54	0.735	1.41	0.56
Grading on "Hard" Factors	−0.01	0.18	0.00	0.24	0.00	0.24
Grading on "Soft" Factors	−0.01	0.15	0.00	0.18	0.00	0.14
Provision of extracurricular activities	2.05	0.98	1.28	0.95	1.1	0.91

$(1, 5630) = 1.05$, $p = .31$. In Italy, nonimmigrant students ($M = 0.07$, $SD = 0.86$) showed a significantly higher sense of belonging at school than immigrant students ($M = -0.16$, $SD = 0.82$), $F(1, 11{,}373) = 59.25$, $p < .01$, $\eta^2 = 0.005$. Similarly, nonimmigrants in Spain ($M = 0.51$, $SD = 1.16$) reported a higher sense of belonging at school than immigrant students ($M = 0.15$, $SD = 1.09$), $F(1, 6408) = 59.36$, $p < .01$, $\eta^2 = 0.01$. The more adverse outcomes of immigrants in comparison to nonimmigrants were confirmed (except for sense of belonging at school in Germany).

Tables 2 and 3 present the descriptive statistics and the correlations of targeted school-level variables for each country, respectively. Schools in all three countries showed rather diverse compositional characteristics, in-school ability grouping frequencies, grading practices, and the provision of extracurricular activities. Among the targeted school practices, only the two grading practices correlated moderately to strongly with each other, and other practices were weakly or nonsignificantly related, therefore they were modeled separately (see below).

Differential associations with students' outcomes

Modeling strategy

The differential associations of the school and classroom practices with student outcomes were modeled in multiple group multilevel analyses in Mplus 7.3. For each country, the analysis was performed for each outcome (i.e., achievement and sense of belonging at school) separately. The basic setup to predict an outcome was a two-level model (i.e., students nested in schools, the multilevel part) with immigrant background as a grouping variable (i.e., the multigroup part). As this grouping variable is a student-level discrete variable, schools can have students from more than one group (i.e., immigrant and nonimmigrant students) and these two groups of students are not completely independent from each other when school effects are considered. This modeling approach ensures to model outcomes of the two groups of students in their intact ecology of schools by taking the dependence at school level into consideration. Specifically, the differential effects of school practices on immigrant and nonimmigrant students (i.e., school-level random effects) can be modeled properly in multilevel structural equation modeling, where two latent variables of the same outcome (one latent variable for each group) are specified. These two latent variables are allowed to correlate with each other to account for the dependency of these two groups, and they are then regressed on the school-level predictors (Asparouhov & Muthen, 2012). Furthermore, whether the differential school-level effects are significantly different for immigrant and nonimmigrant students can be tested in the Wald test. Cross-country comparisons are not part of the statistical modeling, given the already complex models, and given that scales are not scalar invariant across countries (collapsing data across countries would result in biased estimates).

We first estimated the variance at the school level for achievement (a) and sense of belonging (b) in each country in the conventional multilevel model. In the baseline model (Model 1), we extended the conventional multilevel model to the multigroup multilevel model with immigrant background as a grouping variable, and estimated the effects of student background (ESCS and gender) and differential effects of the school compositional background (low SES intake and proportion of students with a different heritage language). The differential regression coefficients of the two school compositional predictors for the immigrant and nonimmigrant student group were compared in the Wald test (degrees of freedom equal 2: df = 2) in order to statistically test the significance of the differences. Subsequently, we added school-level predictors to the baseline model one at a time (to check the net effects of each practice) and tested whether the differential school-level effect was significantly different for immigrants and nonimmigrants for the target practice in the Wald test (df = 1). In Model 2 and Model 3 we

Table 3
Correlations of school-level variables.

Lower		1	2	3	4	5	6	7
		Percentage Low SES intake	Proportion of students with a different heritage language	Ability grouping by class	Ability grouping by subject	Grading on "Hard" Factors	Grading on "Soft" Factors	Provision of extra-curricular activities
Germany	2	0.56**						
	3	0.27**	0.16*					
	4	0.26**	0.13	0.09				
	5	−0.01	0.14	0.13	0.03			
	6	0.03	0.18*	0.05	0.04	0.61**		
	7	−0.26**	−0.11	−0.10	−0.02	0.12	−0.01	1
Italy	2	0.29**						
	3	0.13*	0.14*					
	4	0.02	−0.07	0.16**				
	5	−0.02	0.00	0.00	0.06			
	6	0.09	−0.20**	−0.11	0.04	−0.16**		
	7	−0.018	−0.04	0.15**	0.01	0.03	−0.11	1
Spain	2	0.30**						
	3	0.10	0.31**					
	4	0.19*	0.14	0.13				
	5	0.07	0.01	0.03	0.07			
	6	0.05	−0.29**	−0.08	−0.03	0.31*		
	7	−0.17*	−0.03	0.00	0.01	0.06	0.06	1

* $p < .05$.
** $p < .01$.

Table 4
Unstandardized and standardized regression coefficients the science achievement in the multigroup multilevel analysis.

Model	Predictor	Germany				Italy				Spain			
		B	B_s.e.	β	β_s.e.	B	B_s.e.	β	β_s.e.	B	B_s.e.	β	β_s.e.
1a	ESCS	**32.240**	2.506	**0.334**	0.020	**23.890**	2.159	**0.255**	0.020	**24.935**	1.299	**0.342**	0.016
	Female	**−19.555**	3.011	**−0.110**	0.020	**−14.235**	3.507	**−0.084**	0.020	**−9.229**	2.726	**−0.054**	0.016
	School Low SES-C1	**−1.397**	0.199	**−0.908**	0.070	**−1.351**	0.216	**−0.977**	0.050	**−0.362**	0.138	**−1.030**	0.067
	School Low SES-C2	**−2.094**	0.286	**−0.921**	0.070	**−1.088**	0.262	**−0.615**	0.130	**−0.639**	0.186	**−0.688**	0.188
	School Language-C1	−0.259	0.188	−0.154	0.110	−0.151	0.319	−0.071	0.150	0.036	0.081	0.151	0.361
	School Language-C2	−0.331	0.267	−0.134	0.110	**−1.702**	0.411	**−0.628**	0.130	**−0.336**	0.142	**−0.546**	0.212
	Wald Test	χ^2 (2) = 11.02**				χ^2 (2) = 14.24**				χ^2 (2) = 11.69**			
2a	ESCS	**3.086**	2.494	**0.317**	0.020	**23.399**	2.218	**0.250**	0.020	**24.346**	1.322	**0.336**	0.016
	Female	**−19.956**	2.932	**−0.114**	0.020	**−13.829**	3.571	**−0.082**	0.020	**−9.217**	2.747	**−0.054**	0.016
	School Low SES-C1	**−1.353**	0.214	**−0.794**	0.090	**−1.383**	0.233	**−0.956**	0.070	**−0.455**	0.142	**−0.980**	0.100
	School Low SES-C2	**−1.436**	0.307	**−0.688**	0.120	**−0.708**	0.242	**−0.512**	0.160	**−0.456**	0.202	**−0.683**	0.227
	School Language-C1	−0.339	0.189	−0.182	0.100	−0.248	0.345	−0.113	0.160	−0.018	0.087	−0.060	0.281
	School Language-C2	−0.294	0.254	−0.130	0.110	−0.908	0.387	−0.431	0.180	−0.042	0.154	−0.094	0.351
	Ability Group Class-C1	**−15.033**	5.965	**−0.236**	0.090	**−1.544**	7.482	**−0.030**	0.150	5.490	3.530	0.391	0.245
	Ability Group Class-C2	**−35.767**	7.435	**−0.458**	0.100	**−28.711**	7.998	**−0.593**	0.140	**−1.975**	5.017	**−0.548**	0.246
	Wald Test	χ^2 (1) = 15.85**				χ^2 (1) = 23.85**				χ^2 (1) = 14.55**			
3a	ESCS	**3.387**	2.557	**0.319**	0.020	**23.377**	2.154	**0.250**	0.020	**24.692**	1.309	**0.340**	0.016
	Female	**−19.404**	3.002	**−0.110**	0.020	**−14.731**	3.526	**−0.087**	0.020	**−9.135**	2.734	**−0.054**	0.016
	School Low SES-C1	**−1.458**	0.219	**−0.851**	0.080	**−1.362**	0.225	**−0.931**	0.070	**−0.406**	0.146	**−1.012**	0.087
	School Low SES-C2	**−1.542**	0.312	**−0.760**	0.110	**−0.987**	0.244	**−0.584**	0.130	**−0.441**	0.193	**−0.569**	0.230
	School Language-C1	−0.296	0.188	−0.159	0.100	−0.201	0.342	−0.092	0.160	0.025	0.083	0.093	0.323
	School Language-C2	−0.231	0.235	−0.106	0.110	−1.299	0.412	−0.514	0.150	−0.196	0.175	−0.383	0.328
	Ability Group Subject-C1	−1.819	6.087	−0.149	0.090	**−7.435**	4.318	**−0.233**	0.140	−0.463	3.771	−0.036	0.281
	Ability Group Subject-C2	**−32.636**	6.956	**−0.380**	0.090	**−16.194**	5.289	**−0.439**	0.140	**−12.932**	6.165	**−0.494**	0.236
	Wald Test	χ^2 (1) = 20.15**				χ^2 (1) = 4.37*				χ^2 (1) = 5.67*			
4a	ESCS	**31.928**	2.478	**0.330**	0.020	**23.472**	2.198	**0.251**	0.020	**24.902**	1.307	**0.342**	0.016
	Female	**−19.682**	3.036	**−0.110**	0.020	**−15.064**	3.541	**−0.089**	0.020	**−9.276**	2.719	**−0.054**	0.016
	School Low SES-C1	**−1.387**	0.204	**−0.894**	0.080	**−1.335**	0.213	**−0.966**	0.060	**−0.359**	0.137	**−1.008**	0.095
	School Low SES-C2	**−2.019**	0.294	**−0.892**	0.090	**−1.067**	0.272	**−0.612**	0.140	**−0.663**	0.181	**−0.685**	0.175
	School Language-C1	−0.282	0.195	−0.171	0.120	−0.198	0.328	−0.093	0.160	0.036	0.080	0.150	0.354
	School Language-C2	−0.416	0.285	−0.172	0.120	**−1.639**	0.448	**−0.612**	0.140	**−0.323**	0.130	**−0.504**	0.193
	“Hard” Grading-C1	8.812	18.306	0.045	0.100	−8.421	15.836	−0.080	0.150	−4.796	8.349	−0.179	0.311
	“Hard” Grading-C2	17.296	35.390	0.061	0.120	17.872	24.412	0.132	0.180	−16.800	19.921	−0.228	0.254
	Wald Test	χ^2 (1) = 0.67				χ^2 (1) = 1.10				χ^2 (1) = 0.35			
5a	ESCS	**31.820**	2.462	**0.329**	0.020	**23.320**	2.110	**0.252**	0.020	**24.924**	1.315	**0.342**	0.016
	Female	**−19.747**	3.029	**−0.111**	0.020	**−15.306**	3.407	**−0.092**	0.020	**−9.227**	2.724	**−0.054**	0.016
	School Low SES-C1	**−1.379**	0.206	**−0.892**	0.080	**−1.267**	0.210	**−0.796**	0.090	**−0.362**	0.140	**−1.025**	0.083
	School Low SES-C2	**−2.045**	0.292	**−0.899**	0.080	**−1.077**	0.260	**−0.619**	0.140	**−0.638**	0.187	**−0.685**	0.182
	School Language-C1	−0.297	0.195	−0.180	0.120	−0.237	0.351	−0.097	0.140	0.033	0.091	0.139	0.398
	School Language-C2	−0.408	0.284	−0.168	0.120	**−1.652**	0.460	**−0.617**	0.140	**−0.349**	0.140	**−0.567**	0.213
	“Soft” Grading-C1	17.113	25.408	0.068	0.100	**−82.607**	17.363	**−0.506**	0.100	−1.425	17.704	−0.029	0.376
	“Soft” Grading-C2	27.908	45.547	0.075	0.120	21.840	29.107	0.121	0.160	−26.084	33.041	−0.209	0.258
	Wald Test	χ^2 (1) = 0.61				χ^2 (1) = 12.43**				χ^2 (1) = 0.47			
6a	ESCS	**3.004**	2.507	**0.313**	0.020	**22.134**	2.201	**0.239**	0.020	**24.337**	1.323	**0.335**	0.016
	Female	**−19.737**	2.970	**−0.112**	0.020	**−17.260**	3.546	**−0.103**	0.020	**−9.255**	2.771	**−0.054**	0.016
	School Low SES-C1	**−1.409**	0.228	**−0.799**	0.090	**−1.422**	0.224	**−0.916**	0.070	**−0.412**	0.143	**−1.008**	0.093
	School Low SES-C2	**−1.833**	0.278	**−0.983**	0.090	**−0.781**	0.243	**−0.613**	0.170	**−0.592**	0.170	**−0.660**	0.178
	School Language-C1	−0.335	0.204	−0.177	0.110	−0.084	0.313	−0.035	0.130	0.026	0.086	0.092	0.318
	School Language-C2	−0.116	0.259	−0.059	0.130	**−1.143**	0.375	**−0.584**	0.160	−0.191	0.119	−0.313	0.201
	Extracurricular –C1	**1.113**	4.053	**0.234**	0.090	**9.603**	3.494	**0.339**	0.120	0.350	2.135	0.045	0.269
	Extracurricular –C2	−3.431	4.685	−0.075	0.100	−8.356	4.648	−0.358	0.200	**−12.515**	3.715	**−0.713**	0.155
	Wald Test	χ^2 (1) = 19.36**				χ^2 (1) = 22.55**				χ^2 (1) = 12.35**			

C1 = nonimmigrants; C2 = immigrants.The unstandardized coefficients in **bold** indicate signficance at $p < .05$.
* $p < .05$.
** $p < .01$.

evaluated the effects of ability grouping across classes and across subjects within classes, respectively. In Model 4 and Model 5 variables for grading based on “hard” and on “soft” factors were added, respectively. In Model 6, the provision of extracurricular activities was targeted.

Achievement as outcome

In the null model, the intraclass correlations were 0.47, 0.44, and 0.14 for achievement in Germany, Italy, and Spain, respectively. Table 4 presents the unstandardized and standardized regression coefficients and the Wald test results for Model 1a to Model 6a in each country.

In all three countries, the baseline model (Model 1a) showed a consistent, positive association of students' ESCS with achievement, and that females underperformed males in science. School compositional characteristics were significantly, differently associated with achievement for immigrants and nonimmigrants. Specifically, low SES student intake at school level was consistently, negatively related to achievement for all student groups. In Italy and Spain, the proportion of students with a different heritage language was negatively related to achievement for immigrants, but not for nonimmigrants.

With student background and school composition controlled for, ability grouping across classrooms (Model 2a) was significantly,

Table 5
Unstandardized and standardized regression coefficients the sense of belonging at school in the multigroup multilevel analysis.

Model	Predictor	Germany				Italy				Spain			
		B	B_s.e	β	β_s.e	B	B_s.e.	β	β_s.e	B	B_s.e	β	β_s.e
1b	ESCS	**0.061**	0.021	**0.052**	0.018	0.026	0.015	0.027	0.015	**0.061**	0.014	**0.064**	0.015
	Female	**−0.136**	0.039	**−0.064**	0.018	−0.045	0.029	−0.026	0.016	0.010	0.034	0.005	0.015
	School Low SES-C1	**−0.004**	0.001	**−0.720**	0.288	0.000	0.001	−0.008	0.119	0.002	0.002	0.188	0.126
	School Low SES-C2	**−0.011**	0.002	**−0.754**	0.174	**−0.005**	0.002	**−0.365**	0.172	−0.001	0.002	−0.072	0.124
	School Language-C1	0.002	0.001	0.321	0.236	−0.003	0.002	−0.184	0.132	**−0.004**	0.001	**−0.467**	0.114
	School Language-C2	**0.007**	0.002	**0.522**	0.182	−0.006	0.004	−0.276	0.217	**−0.008**	0.002	**−0.748**	0.166
	Wald Test	χ^2 (2) = 4.95				χ^2 (2) = 12.02**				χ^2 (2) = 14.97**			
	ESCS	**0.058**	0.021	**0.050**	0.018	0.022	0.015	0.023	0.015	**0.057**	0.014	**0.060**	0.015
	Female	**−0.139**	0.039	**−0.065**	0.018	−0.041	0.029	−0.023	0.016	0.000	0.033	0.000	0.015
	School Low SES-C1	**−0.003**	0.001	**−0.523**	0.237	−0.001	0.001	−0.093	0.117	0.002	0.002	0.130	0.137
	School Low SES-C2	**−0.011**	0.002	**−0.791**	0.203	−0.003	0.002	−0.328	0.322	0.000	0.002	0.031	0.139
	School Language-C1	0.001	0.001	0.251	0.216	−0.003	0.002	−0.209	0.138	**−0.004**	0.001	**−0.505**	0.121
	School Language-C2	**0.006**	0.002	**0.505**	0.188	−0.002	0.005	−0.118	0.360	**−0.006**	0.002	**−0.646**	0.232
	Ability Group Class-C1	**−0.065**	0.030	**−0.352**	0.176	0.059	0.038	0.181	0.116	−0.004	0.039	−0.010	0.110
	Ability Group Class-C2	−0.032	0.061	−0.075	0.140	−0.100	0.061	−0.317	0.326	**−0.120**	0.058	**−0.314**	0.152
	Wald Test	χ^2 (1) = 0.36				χ^2 (1) = 6.82**				χ^2 (1) = 5.50*			
3b	ESCS	**0.064**	0.022	**0.054**	0.018	0.023	0.015	0.024	0.016	**0.057**	0.014	**0.059**	0.015
	Female	**−0.143**	0.039	**−0.067**	0.018	−0.038	0.028	−0.022	0.016	0.017	0.033	0.008	0.015
	School Low SES-C1	**−0.003**	0.002	**−0.593**	0.297	−0.001	0.001	−0.066	0.116	0.002	0.002	0.162	0.137
	School Low SES-C2	**−0.012**	0.003	**−0.818**	0.190	−0.004	0.002	−0.329	0.188	0.001	0.002	0.074	0.131
	School Language-C1	0.002	0.001	0.323	0.243	−0.003	0.002	−0.215	0.137	**−0.004**	0.001	**−0.470**	0.112
	School Language-C2	**0.006**	0.003	**0.481**	0.189	−0.002	0.005	−0.126	0.245	**−0.006**	0.001	**−0.622**	0.194
	Ability Group Subject-C1	−0.043	0.034	−0.237	0.199	0.029	0.018	0.142	0.089	**−0.098**	0.042	**−0.242**	0.103
	Ability Group Subject-C2	0.020	0.062	0.041	0.124	−0.041	0.041	−0.144	0.156	**−0.242**	0.060	**−0.509**	0.136
	Wald Test	χ^2 (1) = 1.50				χ^2 (1) = 3.03				χ^2 (1) = 8.71**			
4b	ESCS	**0.064**	0.021	**0.054**	0.018	0.022	0.015	0.023	0.015	**0.062**	0.014	**0.065**	0.015
	Female	**−0.131**	0.039	**−0.062**	0.018	−0.048	0.029	−0.027	0.016	0.012	0.034	0.005	0.015
	School Low SES-C1	**−0.004**	0.001	**−0.641**	0.260	0.000	0.001	0.004	0.116	0.002	0.002	0.188	0.123
	School Low SES-C2	**−0.012**	0.002	**−0.767**	0.172	**−0.005**	0.002	**−0.352**	0.168	−0.001	0.002	−0.062	0.127
	School Language-C1	0.001	0.001	0.284	0.230	−0.002	0.002	−0.159	0.125	**−0.004**	0.001	**−0.467**	0.108
	School Language-C2	0.008	0.002	0.583	0.182	−0.005	0.004	−0.229	0.219	**−0.008**	0.002	**−0.749**	0.169
	"Hard" Grading-C1	0.023	0.109	0.038	0.183	**0.153**	0.075	**0.214**	0.111	**0.223**	0.102	**0.246**	0.110
	"Hard" Grading-C2	**−0.573**	0.240	**−0.355**	0.185	−0.013	0.210	−0.011	0.185	0.124	0.239	0.102	0.197
	Wald Test	χ^2 (1) = 5.37*				χ^2 (1) = 0.54				χ^2 (1) = 0.15			
5b	ESCS	**0.064**	0.021	**0.054**	0.018	0.023	0.015	0.023	0.015	**0.061**	0.014	**0.064**	0.015
	Female	**−0.131**	0.039	**−0.061**	0.018	−0.048	0.028	−0.027	0.016	0.010	0.034	0.004	0.015
	School Low SES-C1	**−0.004**	0.001	**−0.657**	0.280	0.000	0.001	0.040	0.122	0.002	0.002	0.189	0.126
	School Low SES-C2	**−0.011**	0.003	**−0.797**	0.177	**−0.005**	0.002	**−0.302**	0.156	−0.001	0.002	−0.076	0.127
	School Language-C1	0.001	0.001	0.266	0.245	−0.003	0.002	−0.207	0.133	**−0.004**	0.001	**−0.467**	0.118
	School Language-C2	**0.008**	0.002	**0.599**	0.184	−0.006	0.004	−0.254	0.206	**−0.008**	0.002	**−0.719**	0.185
	"Soft" Grading-C1	0.112	0.141	0.149	0.188	**−0.505**	0.109	**−0.549**	0.094	−0.006	0.170	−0.003	0.106
	"Soft" Grading-C2	**−0.588**	0.277	**−0.296**	0.130	**0.477**	0.233	**0.295**	0.156	0.326	0.538	0.151	0.235
	Wald Test	χ^2 (1) = 4.70*				χ^2 (1) = 12.58**				χ^2 (1) = 0.34			
6b	ESCS	**0.068**	0.021	**0.058**	0.018	0.023	0.015	0.024	0.016	**0.050**	0.014	**0.052**	0.015
	Female	**−0.137**	0.039	**−0.064**	0.018	**−0.058**	0.028	**−0.033**	0.016	0.007	0.035	0.003	0.015
	School Low SES-C1	**−0.004**	0.001	**−0.701**	0.322	0.000	0.001	−0.034	0.116	0.002	0.002	0.159	0.133
	School Low SES-C2	**−0.012**	0.003	**−0.810**	0.170	−0.004	0.002	−0.315	0.190	0.000	0.002	0.009	0.136
	School Language-C1	0.002	0.001	0.373	0.250	−0.003	0.002	−0.197	0.136	**−0.004**	0.001	**−0.444**	0.112
	School Language-C2	**0.006**	0.003	**0.449**	0.174	−0.003	0.004	−0.178	0.237	**−0.007**	0.002	**−0.734**	0.198
	Extracurricular –C1	−0.027	0.021	−0.262	0.224	0.027	0.018	0.159	0.106	0.050	0.027	0.225	0.112
	Extracurricular –C2	0.025	0.028	0.085	0.097	−0.044	0.041	−0.189	0.202	−0.080	0.048	−0.299	0.179
	Wald Test	χ^2 (1) = 3.38				χ^2 (1) = 2.97				χ^2 (1) = 7.00**			

C1 = nonimmigrants; C2 = immigrants.The unstandardized coefficients in **bold** indicate signficance at $p < .05$.
* $p < .05$.
** $p < .01$.

differently associated with achievement of the two groups of students in all three countries. The negative associations for immigrants were stronger than those for nonimmigrants. Model 3a showed similar significant differential results for grouping by subjects: ability grouping by subjects was more negatively related to achievement of immigrants than nonimmigrants. Hypothesis 1a, which stated a more negative association between in-school ability grouping and achievement among immigrant compared to nonimmigrant students, was supported.

Grading based on "hard" criteria (Model 4a) did not show any significant prediction on achievement in the three countries and the Wald test showed nonsignificant differences in its effects for immigrant and nonimmigrant students. Hypothesis 2a, which stated a positive association between grading based on "hard" criteria and achievement, was not supported. Grading based on "soft" factors (Model 5a) did not predict achievement of either group of students in Germany and Spain. In Italy, a negative effect for nonimmigrant students and a nonsignificant effect for immigrant students was found. All in all, there was no consistent patterning across student groups or countries regarding associations between grading practices and achievement.

The provision of diverse extracurricular activities (Model 6a) was differently associated with achievement across the two groups of students in all three countries: a positive effect was found for nonimmigrant students in Germany and Italy (while it was non-significant for immigrants in these two countries), and a negative effect was found for immigrants in Spain (while it was non-significant for nonimmigrants). Hypothesis 3a (extracurricular activities are positively associated with achievement) was largely supported among the non-immigrant groups but not among immigrant students, Hypothesis 3c suggesting a stronger effect for immigrants than nonimmigrants, was not supported.

Sense of belonging at school as outcome

The intraclass coefficients were 0.02, 0.04, and 0.04 in Germany, Italy, and Spain, respectively. The same model building strategy as for achievement was employed. The unstandardized and standardized regression coefficients and the Wald test results are presented in Table 5. In Model 1b, we found a consistent positive association of students' ESCS with students' sense of belonging at school in Germany and Spain, but not in Italy. There was a significant gender difference in Germany: females felt less sense of belonging at school than males. School compositional characteristics exhibited differential associations with sense of belonging in Italy and Spain but not in Germany. In Italy, school intake of students of low SES was negatively associated with only immigrants' sense of belonging, whereas the proportion of students with a different heritage language did not play a role. In Spain, the proportion of students with a different heritage language was a negative predictor for both immigrants and nonimmigrants, and its effect seemed stronger for immigrants than nonimmigrants, whereas school SES intake did not show any significant effects.

Ability grouping across classes (Model 2b) was differentially related to the sense of belonging at school for the two groups of students in Spain. Particularly, it was more strongly, negatively related to immigrant student' sense of belonging in comparison to nonimmigrants in Spain. Similar, ability grouping by subjects (Model 3b) had a significantly negative effect on sense of belonging at school in Spain, and more so for immigrant compared to nonimmigrant students. Hypothesis 1b (a more negative association between ability grouping and sense of belonging among immigrants) was supported in Spain.

In Model 4b, “hard” grading had significantly different effects for the two groups of students in Germany, but not in Italy or Spain. Immigrant students' sense of belonging at school was more negatively associated with “hard” grading than for nonimmigrants in Germany (in Italy and Spain, this practice seemed to benefit nonimmigrants' sense of belonging at school). In Model 5b, “soft” grading showed significantly different associations with the two student groups in Germany and Italy, yet in different ways. In Germany, it had a negative association with immigrants' sense of belonging, whereas in Italy the effect was positive for immigrant students and negative for nonimmigrant students. Hypothesis 2c, which referred to a stronger effect of “soft” grading on sense of belonging among immigrants compared to nonimmigrants, was supported only in Italy.

Provision of diverse extracurricular activities (Model 6b) did not predict sense of belonging of either group of students in these countries (Hypothesis 3b and 3d were not supported).

Discussion

There is a widespread interest in fostering immigrant students' learning experience and outcomes by identifying “malleable” school and classroom factors that have the potential to enhance their achievement and sense of belonging at school. We set out to test the differential associations of school-level practices of different supportive focus (academic excellence versus caring and expressive support) with students' achievement and sense of belonging at school with immigrant background factored in. Using data of three European countries from the 2015 PISA, we found that (1) there were largely consistent associations of student background (ESCS and gender) and school compositional characteristics (school intake of low SES students and proportion of students with a different heritage language) with science achievement, and to a lesser extent with sense of belonging at school; these school composition effects tended to differ for immigrant and nonimmigrant students (except for sense of belonging at school in Germany); (2) with student background and school compositional characteristics controlled for, in-school ability grouping for classes and subjects (as academic excellence focused practices) were in general negatively related to all students' learning experience, and they were more negatively related to the achievement of immigrant than non-immigrant students; (3) “hard” grading (as an academic excellence focused practice) was not related to achievement in any student group in the three countries; (4) practices of caring and expressive support (“soft” grading, provision of diverse extracurricular activities) showed mixed results in their associations with achievement and school sense of belonging across student groups, and they did not fulfil their hypothesized potential in promoting immigrants' sense of belonging at school. In the following, we discuss these findings and their implications.

Multidimensionality of outcomes

Before discussing the differential associations, it should be noted that, in comparison to nonimmigrant students, immigrant students indeed had a lower achievement in all three countries, and a lower level of sense of belonging at school in Italy and Spain. Thus, there is a need to close the gap between immigrant and nonimmigrant students for both outcomes. As advocated by the dynamic model of educational effectiveness (Creemers & Kyriakides, 2006), evaluation of educational effectiveness should go beyond academic achievement. Sense of belonging serves as an important nonacademic outcome, because it is part and parcel of human capital that can be intervened more successfully than basic cognitive skills for youth and adults (Heckman, 2000). In this study, science achievement and sense of belonging at school were found to be weakly related, representing distinctive outcomes of students' learning that do not always go hand in hand (Sammons, 1996). Thus, it stands to reason that they are predicted by different practices to different degrees.

School-and classroom level practices

As hypothesized, in-school ability grouping invariably showed a more negative association with immigrant students' achievement in all countries (H1a). With student and school background controlled for, the more negative association of ability grouping with immigrants' sense of belonging at school (H1b) was only supported in Spain. The lack of support in Germany and Italy may be partially due to the low school-level variations of sense of belonging. Ability grouping seems to be beneficial for specific student groups only, and it introduces more separation and educational inequality by often placing immigrant students in lower ability groups. Immigrant students might be trapped into low level educational environments, before they even had the chance to acquire linguistic skills or knowledge of the destination country (Nusche, 2009). Thus, caused by external barriers, immigrants might not be able to demonstrate their actual academic potential. Furthermore, research and policies advocating the delay of school tracking should also be coupled with reducing or eliminating within-school ability grouping for equity considerations, as this is especially negative for immigrant students' learning outcomes.

“Hard” grading was hypothesized to have a positive association with achievement (H2a), but in all three countries, it did not show any significant nor differential association with achievement of immigrant and nonimmigrant students. It seems that this academic excellence focused practice is necessary but not sufficient to promote achievement. Instead, “hard” grading should be combined with personalized feedback and the adaptation of teaching practices according to students' needs in order to produce desirable outcomes (Harks, Rakoczy, Hattie, Besser, & Klieme, 2014; Rakoczy, Klieme, Bürgermeister, & Harks, 2008).

The expectation that grading based on "soft" factors should benefit sense of belonging of all students (H2b) and especially immigrant students (H2c) was not confirmed in all countries. Instead, we found "soft" grading to exhibit differential effects across countries. We found a positive association with immigrant students' sense of belonging in Italy only (compared to a negative association for nonimmigrant students). Yet, the reverse was true for Germany, where grading based on "soft" factors was negatively associated with immigrant students' sense of belonging at school. Unexpectedly, "soft" grading was found to be more negatively related to achievement of nonimmigrants than immigrants in Italy. These mixed findings raise the question on the subjectivity and bias that teachers may exhibit while grading based on "soft" factors. There is more flexibility in teachers' "soft" than "hard" grading, and "soft" grading may be more vulnerable to biased judgements (e.g., caused by prejudices relating to ethnic background) (Archer & McCarthy, 1988). For instance, if teachers (unconsciously) believe that immigrants as a group tend to perform worse than nonimmigrants, these expectations can have a negative impact on grades assigned to immigrants, even though objectively their performance does not differ compared to nonimmigrants (Bonefeld & Dickhäuser, 2018). Furthermore, immigrant students may perform worse as a reaction to the teacher's behaviors and unfair treatment which might lower their sense of belonging. This mechanism is worth exploring in different cultural contexts. Empirical data on possible prejudiced perceptions and grading among teachers and subsequent response mechanisms of students in different countries should be collected to further elucidate why associations of "soft" grading and learning outcomes for immigrant and nonimmigrant students differ. All in all, it is expected that grading practices considering students efforts and nonacademic abilities which are fairly implemented by teachers (i.e., without bias against immigrants), might be beneficial for immigrant students, as they allow more flexibility to consider different prerequisites of students.

The provision of extracurricular activities was mainly beneficial to nonimmigrant students' achievement (H3a partially supported) but not sense of belonging, and it did not benefit immigrants' achievement nor sense of belonging at school (H3b, c, and d rejected). One reason might be that immigrant students are less likely to participate in extracurricular activities than nonimmigrants (Jiang & Peguero, 2016). Social and economic barriers experienced by immigrant students (e.g., lack of knowledge of the school system or discrimination) may impede their participation in extracurricular activities at school. Another explanation for the negative or missing impact on immigrant students' achievement might be that immigrant students tend to participate in extracurricular activities with no clear relation to academics (e.g., Turkish immigrants or Hispanics are likely to attend sports) (Peguero, 2011). Our analysis is based on principals' self-reports with regard to the provision of diverse extracurricular activities and not actual participation. Thus, the provision of extracurricular activities is not sufficient: immigrant students should be encouraged and guided to actively involve themselves in extracurricular activities.

The importance of the context

Across all three countries, immigrant students showed a lower achievement compared to nonimmigrant students. Yet, the mechanisms are not all the same. Besides similarities (i.e., negative effects of school low SES intake and in-school ability grouping on achievement), we also found quite some differences across countries. We highlight two main findings that capitalize on the impact of national and school contexts. First, the role of the proportion of students with a different heritage language played a different role: with school low SES intake controlled for, it was not predictive of achievement in Germany, but it was still negatively related to immigrant students' achievement in Italy and Spain (Model 1a). For sense of belonging at school, the proportion of students with a different heritage language was positively related to immigrants' sense of belonging, whereas it played a negative role in Spain. Secondly, immigrant students in Spain, in comparison to immigrant students in Germany and Italy, seem to be in a more adverse situation, as in-school ability grouping and provision of extracurricular activities were negatively associated with their achievement (these effects were weaker or nonsignificant for immigrants in Germany and Italy).

Admittedly, these differences can be attributed to a complex set of factors, such as different effects of instructional quality (Bellens, Van Damme, Van Den Noortgate, Wendt, & Nilsen, 2019) or different main immigrant groups within each destination country and the social and ethnic hierarchy associated with them (e.g., Verkuyten & Kinket, 2000). The national and school contexts are inescapably important. One explanation for the cross-country differences in the effects of the proportion of students with a different heritage language may be linked to the different between-school tracking systems. Early between-school tracking in Germany may result in clustering immigrant students in certain school types, which is not the case in Italy and Spain. This is indirectly confirmed by the proportion of schools with more than 30% immigrant students in our samples: 20% in Germany, but only 7% each in Italy and Spain. In line with this, the proportion of students with a different heritage language can serve as a protective factor for immigrants' sense of belonging, as it prevents further segregation of immigrant students in their social life at school (students of different minority ethnic backgrounds tend to cluster and they are not singled out as the only minority group). Its negative effect on sense of belonging in Spain may further be due to the country's multilingual history, where language represents different conflicting group identities (Enesco, Navarro, Paradela, & Guerrero, 2005).

Similarly, the more severe inequity experienced by immigrant students in Spain compared to Germany and Italy can have multiple causes. One possibility is the readiness and competence of schools and teachers to manage multicultural classrooms, and help level up immigrant students' learning. The higher proportion of teachers in Spain feeling not adequate in coping with a multicultural classroom compared to the OECD average in the TALIS report seems to be supportive of this assumption (OECD, 2019). A national culture to enhance intercultural competence for principals, teachers, and students may work towards equity between immigrant and nonimmigrant students.

Limitations and further directions

Our study has a few limitations that further research can remedy: we made use of secondary data from the PISA where only the student and school levels are factored in; yet a lot happens at the classroom level which could not be captured, and the aggregation of teacher reports on the school level did not always provide reliable estimates (e.g., "hard" grading had a relatively low reliability). Future research with a finer distinction and synchronization of analysis levels helps pinpoint the intervention points at individual, classroom, and school level. Secondly, we did not distinguish first-and second-generation immigrants and information on the immigrant's culture of origin was missing, which prevents us from unpacking the associations according to culture of origin and generational status, and in part might have contributed to our mixed results across countries. However, the large, national representative samples, rich data from various sources and the fitting modeling approach enable us to draw robust conclusions and unfold the complexity of immigrant education in three European countries. Further research targeting immigrant education should gather background information and study immigration issues in a more reflective manner (Motti-Stefanidi & Salmela-Aro, 2018).

Conclusions

To identify beneficial school and classroom practices for better learning experiences for immigrants, we investigated the associations of within-school ability grouping, grading practices, and the provision of extracurricular activities with achievement and sense of belonging at school for immigrant and nonimmigrant students in Germany, Italy, and Spain. We selected these countries based on their similar challenges to integrate low-educated immigrants and immigration politics. We demonstrated the universality of the negative impact of in-school ability grouping especially for immigrant students in all countries, and found much cross-cultural and cross-group variation on other factors. Our results point towards the importance of the national and school contexts in which immigrants are educated. As different practices differently affect immigrant students' learning outcomes compared to those of nonimmigrants, targeted interventions should build upon this knowledge to promote student learning matched to the needs of the different student groups in specific contexts.

Funding

This work was partially supported by the Marie Sklodowska-Curie Individual Fellowship European program [grant number 748788]

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appdev.2019.101089.

References

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*, 235–251. https://doi.org/10.1037/0022-0663.100.2.235.

Ammermueller, A. (2005). Educational opportunities and the role of institutions. *ZEW Discussion Papers* (pp. 05–44). . https://doi.org/10.2139/ssrn.753366.

Archer, J., & McCarthy, B. (1988). Personal biases in student assessment. *Educational Research, 30*, 142–145. https://doi.org/10.1080/0013188880300208.

Asparouhov, T., & Muthen, B. (2012). Multiple group multilevel analysis. *Mplus Web Notes NO. 16*. Retrieved from https://www.statmodel.com/examples/webnotes/webnote16.pdf.

Atlay, C., Tieben, N., Hillmert, S., & Fauth, B. (2019). Instructional quality and achievement inequality: How effective is teaching in closing the social achievement gap? *Learning and Instruction, 63*, 101211. https://doi.org/10.1016/j.learninstruc.2019.05.008.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497–529. https://doi.org/10.1037/0033-2909.117.3.497.

Bayer, S., Klieme, E., & Jude, N. (2017). Assessment and evaluation in educational contexts. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.). *Assessing contexts of learning: An international perspective*. Springer.

Belfi, B., Fraine, B.d., Goos, M., & van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review, 7*, 62–74. https://doi.org/10.1016/j.edurev.2011.09.002.

Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in Flanders (Belgium), Germany, and Norway. [journal article]. *Large-scale Assessments in Education, 7*, 1. https://doi.org/10.1186/s40536-019-0069-2.

Betts, J. R., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review, 22*, 343–352. https://doi.org/10.1016/s0272-7757(02)00059-6.

Bonefeld, M., & Dickhäuser, O. (2018). (Biased) grading of students' performance: Students' names, performance level, and implicit attitudes. *Frontiers in Psychology, 9*, 481. https://doi.org/10.3389/fpsyg.2018.00481.

Brown, R., & Evans, W. P. (2016). Extracurricular activity and ethnicity. *Urban Education, 37*, 41–58. https://doi.org/10.1177/0042085902371004.

Bygren, M. (2016). Ability grouping's effects on grades and the attainment of higher education. *Sociology of Education, 89*, 118–136. https://doi.org/10.1177/0038040716642498.

Camacho, D. E., & Fuligni, A. J. (2015). Extracurricular participation among adolescents from immigrant families. *Journal of Youth and Adolescence, 44*, 1251–1262. https://doi.org/10.1007/s10964-014-0105-z.

Caro, D. H., Lenkeit, J., Lehmann, R., & Schwippert, K. (2009). The role of academic achievement growth in school track recommendations. *Studies in Educational Evaluation, 35*, 183–192. https://doi.org/10.1016/j.stueduc.2009.12.002.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. https://doi.org/10.1207/s15328007sem0902_5.

Chiu, C.-Y., Walter, J. L., David, M., & Colleen, W. (2013). Cross-cultural competence: Theory, research, and application. *Journal of Cross-Cultural Psychology, 44*, 843–848. https://doi.org/10.1177/0022022113493716.

Chiu, M. M., Pong, S.-l., Mori, I., & Chow, B. W.-Y. (2012). Immigrant students' emotional and cognitive engagement at school: A multilevel analysis of students in 41 countries. *Journal of Youth and Adolescence, 41*, 1409–1425. https://doi.org/10.1007/s10964-012-9763-x.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modeling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*, 347–366. https://doi.org/10.1080/09243450600697242.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* London, UK: Routledge.

Enesco, I., Navarro, A., Paradela, I., & Guerrero, S. (2005). Stereotypes and beliefs about different ethnic groups in Spain. A study with Spanish and Latin American children living in Madrid. *Journal of Applied Developmental Psychology, 26*, 638–659. https://doi.org/10.1016/j.appdev.2005.08.009.

van Ewijk, R. (2011). Same work, lower grade?: Student ethnicity and teachers' subjective assessments. *Economics of Education Review, 30*, 1045–1058. https://doi.org/10.1016/j.econedurev.2011.05.008.

Farb, A. F., & Matjasko, J. L. (2012). Recent advances in research on school-based extracurricular activities and adolescent development. *Developmental Review, 32*, 1–48. https://doi.org/10.1016/j.dr.2011.10.001.

Field, S., Kuczera, M., & Pont, B. (2007). *No more failures: Ten steps to equity in education.* Paris: OECD Publishing.

Fredricks, J. A., & Eccles, J. S. (2006). Is extracurricular participation associated with beneficial outcomes? Concurrent and longitudinal relations. *Developmental Psychology, 42*, 698–713. https://doi.org/10.1037/0012-1649.42.4.698.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*, 269–290. https://doi.org/10.1080/01443410.2013.785384.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics, 54*, 3–56. https://doi.org/10.1006/reec.1999.0225.

Hirschi, T. (1972). *Causes of delinquency.* 1. Ed., 2. Print ed.Berkeley: University of California Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indixes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*. https://doi.org/10.1080/10705519909540118.

Huddleston, T., Bilgili, O., Joki, A.-L., & Vankova, Z. (2015). *Migrant Integration Policy Index 2015*.

Im, M. H., Hughes, J. N., Cao, Q., & Kwok, O.-M. (2016). Effects of extracurricular participation during middle school on academic motivation and achievement at grade 9. *American Educational Research Journal, 53*, 1343–1375. https://doi.org/10.3102/0002831216667479.

Ireson, J., & Hallam, S. (2005). Pupils' liking for school: Ability grouping, self-concept and perceptions of teaching. *The British Journal of Educational Psychology, 75*, 297–311. https://doi.org/10.1348/000709904x24762.

Jiang, X., & Peguero, A. A. (2016). Immigration, extracurricular activity, and the role of family. *Education and Urban Society, 49*, 314–340. https://doi.org/10.1177/0013124516643759.

Kuger, S., Klieme, E., Jude, N., & Kaplan, D. (Eds.). (2016). *Assessing contexts of learning: An international perspective.* Cham: Springer.

Langenkamp, A. G., & Carbonaro, W. (2018). How school socioeconomic status affects achievement growth across school transitions in early educational careers. *Sociology of Education, 91*, 358–378. https://doi.org/10.1177/0038040718802257.

Le, L. T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing, 9*, 122–133. https://doi.org/10.1080/15305050902880769.

Mahoney, J. L., Cairns, B. D., & Farmer, T. W. (2003). Promoting interpersonal competence and educational success through extracurricular activity participation. *Journal of Educational Psychology, 95*, 409–418. https://doi.org/10.1037/0022-0663.95.2.409.

Marks, A. K., McKenna, J. L., & Coll, C. G. (2018). National mmigration receiving contexts. *European Psychologist, 23*, 6–20. https://doi.org/10.1027/1016-9040/a000311.

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal, 28*, 445–480. https://doi.org/10.3102/00028312028002445.

Maslow, A. (1962). *Towards a psychology of need.* Princeton, NJ: D Van Nostrand.

McNeal, R. B. (1998). High school extracurricular activities: Closed structures and stratifying patterns of participation. *The Journal of Educational Research, 91*, 183–191. https://doi.org/10.1080/00220679809597539.

Migration Policy Institute (2015). *Trends in international migrant stock: Migrants by destination and origin.* United Nations database: POP/DB/MIG/Stock/Rev2015.

Motti-Stefanidi, F., & Masten, A. S. (2013). School success and school engagement of immigrant children and adolescents. *European Psychologist, 18*, 126–135. https://doi.org/10.1027/1016-9040/a000139.

Motti-Stefanidi, F., & Salmela-Aro, K. (2018). Challenges and resources for immigrant youth positive adaptation. *European Psychologist, 23*, 1–5. https://doi.org/10.1027/1016-9040/a000315.

Muthen, L. K., & Muthen, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthen & Muthen.

Nusche, D. (2009). *What works in migrant education? A review of evidence and policy options.*

Paris: OECD Publishing.
OECD (2015a). *Indicators of immigrant integration 2015: Settling in.* Paris: OECD Publishing.
OECD (2015b). *PISA 2015 assessment and analytical framework.* Paris: OECD Publishing.
OECD (2017). *PISA 2015 technical report.* Paris: OECD Publishing.
OECD (2018). *Settling in 2018: Indicators of immigrant integration.* Paris: OECD Publishing.
OECD (2019). *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners.* Paris: OECD Publishing.
Opdenakker, M.-C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondaryeducation: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*, 165–196. https://doi.org/10.1076/0924-3453(200006)11:2;1-Q;FT165.
Peguero, A. A. (2011). Immigrant youth involvement in school-based extracurricular activities. *The Journal of Educational Research, 104*, 19–27. https://doi.org/10.1080/00220670903468340.
Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction grading and feedback in mathematics classrooms. *Zeitschrift für Psychologie, 216*, 111–124. https://doi.org/10.1027/0044-3409.216.2.111.
Roeser, R. W., Midgley, C., & Urdan, T. (1996). Perceptions of the school psychological environment and early adolescents' self-appraisals and academic engagement. *Journal of Educational Psychology, 88*, 408–422. https://doi.org/10.1037/0022-0663.88.3.408.
Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*, 142–151. https://doi.org/10.3102/0013189x10363170.
Sammons, P. (1996). Complexities in the judgement of school effectiveness. *Educational Research and Evaluation, 2*, 113–149. https://doi.org/10.1080/1380361960020201.
Schachner, M. K., Juang, L., Moffitt, U., & van de Vijver, F. J. R. (2018). Schools as acculturative and developmental contexts for youth of immigrant and refugee background. *European Psychologist, 23*, 44–56. https://doi.org/10.1027/1016-9040/a000312.
Shapira, M. (2012). An exploration of differences in mathematics attainment among immigrant pupils in 18 OECD countries. *European Educational Research Journal, 11*, 68–95. https://doi.org/10.2304/eerj.2012.11.1.68.
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*, 459–478. https://doi.org/10.1111/bjep.12087.
Tomlinson, C. A. (2005). Grading and differentiation: Paradox or good practice? *Theory Into Practice, 44*, 262–269. https://doi.org/10.1207/s15430421tip4403_11.
Verkuyten, M., & Kinket, B. (2000). Social distances in a multi ethnic society: The ethnic hierarchy among Dutch preadolescents. *Social Psychology Quarterly, 63*, 75–85. https://doi.org/10.2307/2695882.
van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis of comparative research.* Thousand Oaks, CA: Sage.
Wößmann, L. (2009). International evidence on school tracking: A review. *CESifo DICE Repor, 7*, 26–34.