universität wien

MASTERARBEIT / MASTER'S THESIS

Title der Masterarbeit / Title of the Master's Thesis

„Different Fish in the Pond. The Influence of Immigrant Background and Ethnic Classroom Composition on the Big-Fish-Little-Pond Effect"

verfasst von / submitted by

Ing. Katharina Bergen, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Marko Lüftenegger

Acknowledgements

First, I would like to thank Dr. Marko Lüftenegger for his support and guidance throughout the process of writing my master's thesis and his effort in answering all my questions as soon as possible. I would also like to thank Sophie Oczlon, BSc Msc for the opportunity to help collecting the data that were used in this study's analyses, as well as the advice she has given to me on several occasions. Furthermore, I would like to thank Dr. Takuya Yanagida for sharing his expertise concerning methodology and statistical analyses. Special thanks also go to my partner Martin for his continuous support throughout the years and the many hours he spent on helping me to work out solutions for the various problems encountered while working on this thesis. Thank you to all the people that supported me through giving constructive feedback, proofreading the thesis, or simply having an open ear. Finally, I would like to thank my family and friends who supported me during the entire course of my studies.

Abstract

Increasing multiculturality can be registered in Austria's population, leading to more heterogeneity in school classrooms, which presents both challenges and opportunities. It is of interest to investigate potential differences between students with and without immigrant background to gain insight about how all students can be promoted at best. The present study focuses on the Big-fish-little-pond effect (BFLPE), a very robust, pan-human effect that can have negative consequences for students' academic self-concept (ASC). The main goals were to investigate whether and how the ASC and the BFLPE differ for students with and without immigrant background and whether potential differences in the strength of the BFLPE were moderated by the ethnic classroom composition (the proportion of immigrant students in the classroom). Multilevel multiple regression analyses and multilevel multiple group analyses of cross-sectional data of a survey study of 1,649 students (698 of immigrant background) of 87 classrooms in 10 secondary schools were used to investigate the hypotheses. Results show that students of immigrant background report a significantly higher ASC than their native classmates if individual achievement was statistically controlled for. The BFLPE could not be replicated in the present study. No moderators could be found for the non-significant effect, supporting the generalizability of the BFLPE.

Keywords: academic self-concept, big-fish-little-pond effect, immigrant background, ethnic classroom composition, stereotypes

Zusammenfassung

In Österreich kann eine zunehmende Multikulturalität der Gesellschaft verzeichnet werden, die zu mehr Heterogenität in Österreichs Klassenzimmern führt, was sowohl Herausforderungen als auch Potenziale in sich birgt. Vor diesem Hintergrund ist es von Interesse, mögliche Unterschiede zwischen Schüler*innen mit und ohne Migrationshintergrund zu untersuchen, um alle Schüler*innen bestmöglich fördern zu können. Die vorliegende Masterarbeit beschäftigt sich mit dem Big-fish-little-pond effect (BFLPE), einem universalen, robusten Effekt, der das akademische Selbstkonzept (ASC) negativ beeinflussen kann. Es wurde untersucht, ob und wie sich Schüler*innen mit und ohne Migrationshintergrund in ihrem ASC voneinander unterscheiden, ob sie unterschiedlich stark vom BFLPE beeinflusst werden und ob dies abhängig vom Anteil an Schüler*innen mit Migrationshintergrund in der Klasse ist. Mehrebenen-Multiple-Regressionsanalysen und Mehrebenen-Mehrgruppen-Kausalanalysen wurden auf Basis der Fragebogendaten von 1.649 Schüler*innen (698 mit Migrationshintergrund) der Sekundarstufe I durchgeführt. Die Ergebnisse zeigen, dass Schüler*innen mit Migrationshintergrund unter statistischer Kontrolle der individuellen Leistung ein signifikant höheres ASC vorwiesen als ihre einheimischen Klassenkolleg*innen. Der BFLPE konnte in dieser Stichprobe nicht nachgewiesen werden. Die untersuchten Moderatoren zeigten keine Wirkung auf den nicht-signifikant Effekt, was die Generalisierbarkeit des BFLPE bestätigt.

Schlagworte: akademisches Selbstkonzept, big-fish-little-pond Effekt, Migrationshintergrund, ethnische Klassenzusammensetzung, Stereotype

Table of Content

List of Figures

List of Tables

List of Abbreviations

AHS ..... *allgemein bildende höhere Schule [academic secondary school]*
ASC ..... *academic self-concept*
BCAV ..... *between-class achievement variance*
BFLPE ..... *big-fish-little-pond effect*
BSAV ..... *between-school achievement variance*
I/E ..... *internal/external*
ICC ..... *intraclass correlation*
MLR ..... *maximum likelihood with robust standard errors*
NMS ..... *Neue Mittelschule [new secondary school]*
OECD ..... *Organization for Economic Co-operation and Development*
REM ..... *reciprocal effects model*
SES ..... *socioeconomic status*
SESSKO ..... *Skalen zur Erfassung des schulischen Selbstkonzepts [School-related self-concept scales]*
UNECE ..... *United Nations Economic Commision for Europe*

1 Introduction

The way humans perceive themselves and their abilities plays a big role in every aspect of their lives, which is why a positive self-concept has been identified as a highly desirable outcome in many different fields of psychology (Marsh & Craven, 2006). In the academic context, a positive academic self-concept (ASC) is especially important as it is associated with many different variables such as achievement (e.g., Huang, 2011; Marsh, 1990; Marsh & Craven, 2006; Marsh et al., 2018; Valentine et al., 2004), academic interest (e.g., Marsh & Yeung, 1997; Marsh et al., 2005), educational attainment (e.g., Guay et al., 2004; Guo et al., 2015; Marsh & O'Mara, 2008) and occupational aspirations (e.g., Eccles, 2009; Guo et al., 2015; Nagengast & Marsh, 2012).

One phenomenon that should be emphasized especially when talking about ASC is the big-fish-little-pond effect (BFLPE), which describes the negative relationship between the average achievement of a class or school and the ASC of a student. Said negative relationship has tremendous implications when deciding which school to attend, as attending a high-ability school, which is usually thought to give the best academic education, might result in a lower ASC compared with attending a low-ability school (e.g., Marsh, 1984; Marsh, 1987; Marsh & Parker, 1984). This is especially of relevance in countries with explicit school level tracking, where students are separated in high- or low-ability schools according to their achievement.

The BFLPE has been demonstrated to be temporally stable (e.g., Becker & Neumann, 2016, 2018; Marsh & O'Mara, 2010; Marsh et al., 2007; von Keyserlingk et al., 2019) and generalizable across different cultures and a wide range of individual characteristics, such as learning strategies, individual ability level, and anxiety (e.g., Seaton et al., 2009; Seaton et al., 2010). It has been criticized, though, that key variables associated with differences in the extent that individuals are influenced by the BFLPE have not been identified. However, to know when and where the BFLPE is likely to occur and who might be affected the strongest is important for developing counteracting interventions (Dai & Rinn, 2008). Two potential moderating variables that have not been investigated sufficiently are immigrant background and the ethnic composition of the classroom, in terms of the proportion of students of immigrant background. In light of the increasing population of people of immigrant background in Austria (Statistik Austria, 2019) and the resulting heterogeneity in the classrooms (Braunsteiner et al., 2018), these variables are considered to be of special interest for the BFLPE research.

The goal of this study is to investigate differences in the ASC and the BFLPE between students with and without immigrant background and examine the role of the ethnic class-

room composition. Cross-sectional data of a survey study in 10 Austrian secondary schools is used to investigate the research questions.

In the following chapters, an overview is provided of the situation in Austria, followed by a theoretical outlining of the ASC and the BFLPE, before describing cultural stereotypes and their effect on both students with and without immigrant background.

1.1 Immigration and Education in Austria

Since 1984, Austria is registering a consistently positive migrant balance (more people moving to, than away from it), a trend, that reached its peak in the European migrant crisis of 2015/2016, with more than 113,000 people moving to Austria in 2015, about 90,000 of whom were refugees (Oberwimmer et al., 2018; Statistik Austria, 2020b). In 2019, about 2.07 million people of immigrant background lived in Austria, accounting for approximately 23.7% of the total population (Statistik Austria, 2020a). In the definition of immigrant (or foreign) background, Austria follows the recommendations of the United Nations Economic Commission for Europe (UNECE), classifying people to be of immigrant background, if they or both their parents were born abroad (UNECE, 2015, p. 136).

Concerning education, there are a few considerable differences between the native population and people of immigrant background. For example, the mean level of educational attainment is lower for people of immigrant background than in the native population, which is related to a higher percentage of low paying occupations and, subsequently, low socioeconomic status (SES) (Oberwimmer et al., 2018). Students of immigrant background are less frequently attending schools that are graduated with *Matura* (the school-leaving examination that allows people in Austria to attend universities) and are more frequently ending their education without completing compulsory schooling. Furthermore, the proportion of adolescents that do not attend any further education programs after completing compulsory schooling is higher in students of immigrant background (Statistik Austria, 2019). The relationship between the social background and educational attainment is especially strong in Austria, therefore, most of these differences can be attributed to children of immigrant background being affected more strongly by social risk factors, such as low educational levels of their parents, low SES and speaking a non-German language in everyday life (Oberwimmer et al., 2018). For example, while 3% of the parents of native students have completed only compulsory schooling, the percentage is more than six times as high (19%) for students of immigrant background (Oberwimmer et al., 2018). This fact is problematic, as the “inheritance” of educational attainment (the phenomenon of children very frequently obtaining the same educa-

tional attainment as their parents) is especially high both for children of immigrant background and in lower levels of parents' education, resulting in an increased risk for students of immigrant background to not achieve high levels of educational attainment (Statistik Austria, 2019).

The increasing population of people with immigrant background consequently leads to more heterogeneity in Austria's classrooms, not only in culture, language, and religion (Braunsteiner et al., 2018), but also in performance (Bacher, 2010). This presents challenges, as the main educational mandate in Austria is to promote the personal and social development of each child in accordance with their individual needs, with the ultimate goal to help them become open-minded, independent, and sympathetic parts of the community (Braunsteiner et al., 2018). In Austria, these challenges have not been mastered yet, as is demonstrated by the existence of strong relationships between language, SES, parents' education, and the academic achievements of children, indicating that not every child is promoted at its best (Braunsteiner et al., 2018; Oberwimmer et al., 2018). It is to be noted though, that these relationships vary greatly over different schools, indicating that some schools are better in handling heterogeneity than others (Braunsteiner et al., 2018). Not only challenges but also opportunities arise as students with different ethnic backgrounds, experiences, interests, and abilities provide the ideal environment to learn from each other and further cultural understanding and solidarity (Vock & Gronostaj, 2017). Performance-heterogeneity also provides opportunities for the improvement of academically weaker students and those with special educational needs, as they are profiting from greater heterogeneity in the classroom. High-performing students have been found to perform well independently of performance heterogeneity (Braunsteiner et al., 2018).

There might be some direct advantages of immigration to Austria as well. People of immigrant background, especially children and adolescents, bring skills and potentials that could be an enrichment for the society (Bacher, 2010). As has been outlined already, these potentials are currently unlikely to unfold, as students of immigrant background are confronted with many disadvantages in Austria's school system (Bacher, 2010; Braunsteiner et al., 2018; Oberwimmer et al., 2018; Statistik Austria, 2019).

In light of these circumstances, it is important to identify and investigate differences between students with and without an immigrant background, in order to develop programs and interventions that enable all students to unfold their potentials, so that everyone can profit from the increasing multiculturality of our society. With focus on the ASC and the related BFLPE, the goal of this master thesis is to contribute to this line of research.

1.2 The Academic Self-Concept (ASC)

This chapter aims to give an overview of the ASC construct, its formation and development, outline its importance in the academic context, and point out differences between students with and without immigrant background.

1.2.1 Definition and Differentiation from Related Constructs

The term self-concept can be used rather broadly to label all perceptions a person has formed about themself through their experiences within their environment (Shavelson et al., 1976). As such, it is a relativistic psychological impression that needs some basis of comparison to be formed (Marsh, 1984). This definition has been provided by Shavelson and his colleagues (1976), who reviewed existing definitions of self-concept and formulated a model of self-concept that is still influential today. In their model, both hierarchy and multidimensionality of self-concept are proposed. The two components of self-concept, "academic self-concept" and "non-academic self-concept" can further be divided into different domains and are subordinate to a general self-concept (see Figure 1A). Besides being hierarchical and multidimensional, self-concept was proposed to be stable at the general level, becoming more unstable as one descends the hierarchy. Furthermore, it is posited to have both an evaluative and a descriptive component and to become more differentiated with increasing age (Shavelson et al., 1976). For example, it was assumed that children might first gain a general ASC, before developing ASCs for distinct subject domains (e.g., mathematics). ASC was therefore conceptualized as a single second-order factor model with a general ASC and several domain-specific ASCs (Shavelson et al., 1976).

Subsequent research found that even in early age children strongly differentiated between verbal and math ASCs, as could be seen by very weak correlations between the two self-concepts. Consequently, the two self-concepts could no longer be incorporated into one general ASC (Marsh & Shavelson, 1985). These weak correlations did not apply to verbal and math achievements which are usually highly correlated (for a meta-analysis see Möller et al., 2009). Further investigation of this phenomenon lead to the development of the internal/external (I/E) frame of reference model, according to which the formation of ASCs depends not only on the comparison of one's abilities to those of other people in the environment (external) but also to the comparison of a person's performance in different subjects (internal) (Marsh, 1986). These internal comparisons lead to a negative relationship between math achievement and verbal ASC, and vice-versa, as a good performance in one domain overshadows the performance in the other. External comparisons and their influence on ASC

are susceptible to frame-of-reference or context effects such as the BFLPE, which will be discussed in detail in a later section.

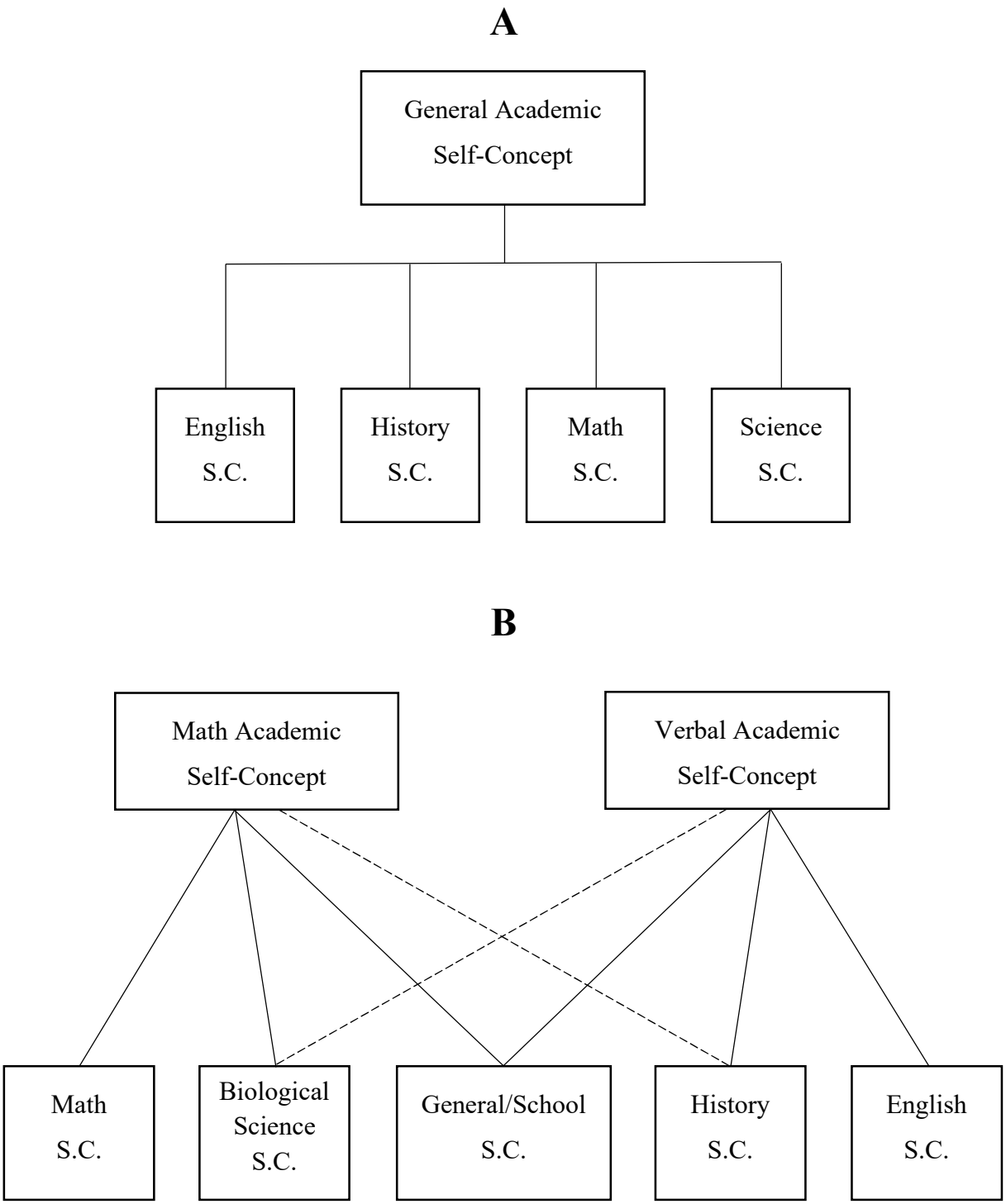

Figure 1. The academic portion of Shavelson's et al. (1976) ASC model (A) and Marsh and Shavelson's (1985) revision (B). Domain-specific S.C.s are not exhaustive. S.C. = self-concept. Adapted from "A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement" by H. Marsh, B. M. Byrne, and R. J. Shavelson, 1988, *Journal of Educational Psychology, 80*(3), p. 378. Copyright 1988 by the American Psychological Association.

These findings lead to a revision of the Shavelson et al. (1976) model (Marsh & Shavelson, 1985), where verbal and math ASC comprised two second-order factors of ASC that could account for almost all of the variance in domain-specific self-concept measures (Marsh et al., 1988, see Figure 1B). This also applies to what has been called school self-concept but can arguably also be called general or global ASC, as it is measured by scales that are relatively unidimensional and content free (Marsh et al., 1988). If not specified otherwise, all further mentioning of ASC in this thesis refers to such a general ASC. Although most of the variance of general ASC can be explained by verbal and math ASCs, there still is a small amount of unique variance left, which might be influenced by social aspects of the school, such as peer relations (Marsh et al., 1988).

In previous research, ASC has been found to be rather high at the beginning of elementary school, only to decrease until students have formed a realistic ASC towards the end of elementary school (Helmke, 1998; Marsh et al., 1998). Interestingly, there are gender differences in this process, with girls being quicker in developing a more realistic ASC than boys (Hasselhorn & Gold, as cited in Ehm et al., 2011). Gender differences can be found in ASC responses as well, with inconsistent findings concerning general ASC, where girls were found to report a higher ASC than boys (Marsh et al., 1998), and vice-versa (Nouwen & Clycq, 2019). Considering verbal and math ASCs explicitly, it is apparent that those self-concepts follow gender role-related stereotypes, with girls having a higher verbal ASC, and boys a higher math ASC (Ehm et al., 2011; Marsh et al., 1988), possibly even independent of the actual performance in the respective domains (e.g., Faber et al., 2011).

ASC is only one of many self-beliefs that have been studied in their relation to different outcome variables in academic and other contexts alike. Given the fact that researchers studying self-beliefs come from many different disciplines and have oftentimes different perspectives, there is some confusion about which self-belief terms describe the same constructs and which are distinct. One distinction that has been verified more than once is that between ASC and (academic) self-efficacy (Bong & Skaalvik, 2003; Marsh et al., 2019; Lent et al., 1997). The self-efficacy construct, first proposed by Bandura (1977, as cited by Bong & Skaalvik, 2003) refers to an individual's conviction that they can perform the actions required to accomplish given tasks. While many similarities have been found between ASC and academic self-efficacy (Bong & Skaalvik, 2003), key differences have been identified (Marsh et al., 2019): 1) ASC is largely based on past accomplishments and therefore retrospective, whereas academic self-efficacy is prospective in that it refers to beliefs about which future tasks can be accomplished in light of an individual's current ability. 2) ASC has a descriptive

and an evaluative function, due to it being often the result of social comparison processes and therefore external frames of reference, while self-efficacy is a solely descriptive construct based on dimensional comparison processes (comparing abilities in different domains) and internal frames of reference (Huang, 2011; Marsh & Yeung, 1997). 3) Because of the social comparison processes underlying ASC, it can be subject to frame-of-reference effects (oftentimes negative effects that arise from comparison with specific environmental factors). As self-efficacy is mostly descriptive, frame-of-reference effects are considerably attenuated (Marsh et al., 2019).

Another construct that has been studied often in educational psychology is outcome expectancy, or expectations of success (e.g., Eccles, 2009; Marsh et al., 2019). In the expectancy-value theory (e.g., Eccles, 2009) expectation of success has been identified as one of two sets of beliefs that fundamentally influence life-defining behavioral choices concerning, for example, education and occupation (Eccles, 2009). The terms outcome expectancy or expectations of success have often been used interchangeably with ASC (e.g., Guo et al., 2015). Empirically, a distinction between the two constructs has been difficult to demonstrate, which is why Marsh and his colleagues (2019) have deemed this distinction a "jangle fallacy", meaning that two apparently dissimilar terms label the same construct.

1.2.2 Research Support for the Importance of ASC in the Educational Context

ASC has been deemed "a 'hot' variable that makes good things happen" (Marsh & Craven, 2006, p. 158), for example, enhancing academic achievement or facilitating higher educational attainment. The most important relationships between ASC and educational outcome variables that have been demonstrated in previous studies are outlined below.

ASC and academic achievement. It is undoubted nowadays that a positive relationship between ASC and academic achievement exists, but about two decades ago, researchers still disagreed about this. Valentine et al. (2004) provided clarification with a meta-analysis where they synthesized findings of over 200 longitudinal studies and found a small, but existent correlation between ASC and academic achievement. The existence of this relationship being established, the question of what the causal ordering of these constructs was, remained. This question was deemed important, as it had relevant implications for interventions and educational programs (Marsh, 1990). Believing that ASC is merely a reflection of academic achievement (*skill-development model*), it can be argued that an improvement of achievement will result in an improved ASC without the necessity of further interventions. On the other hand, the assumption that ASC causally precedes academic achievement (*self-enhancement*

model) has been used to justify educational programs designed to enhance ASC (Marsh, 1990). The ASC researcher Herbert Marsh has argued that a reciprocal causal relationship between ASC and academic achievement is most likely (*reciprocal effects model* [REM], e.g., Marsh, 1990). Concerning interventions and educational programs, practitioners should aim at improving ASC and academic achievement simultaneously, as enhancing just one would likely result in only short-lived improvements (Marsh & Craven, 2006). Support for the REM has also been found in more recent studies (e.g., Huang, 2011; Marsh et al., 2018). According to these findings, the effects of the reciprocal relationship between ASC and academic achievement are robust and invariant throughout early and middle adolescence, meaning that the reciprocal relationship is stable throughout a considerable period of time (Marsh et al., 2018). Huang (2011) pointed out, however, that the explained variance of both variables accounted for by each other was small (between 4% and 7%) and that future studies should include more predictors to see if more variance can be explained.

ASC and academic interest. According to the expectancy-value theory, people will choose to engage in tasks, activities, and behaviors for which they have the highest expectations of success (e.g., Eccles, 2009). In line with this assumption, ASC has been found to be related to coursework selection specifically, and academic interest in general. Marsh and Yeung (1997), for example, have investigated to what extent ASC and academic achievement influence coursework selection and found that while achievement could not consistently predict coursework selection, ASC was positively related to *wanting* to take a course in all nine investigated subjects and was able to predict *actually taking* a course in all but one subjects. In another study, the relationship between ASC and academic interest was investigated under the statistical control of academic achievement. A reciprocal effect between ASC and academic interest was found, whereas the effect of ASC on subsequent interest was stronger than the other way around, although the effect size did not exceed 0.10 (Marsh et al., 2005).

ASC and educational attainment. In a ten-year longitudinal study of three age cohorts, Guay and his colleagues (2004) investigated the relationship between ASC and educational attainment under the statistical control of academic achievement, family structure (single parent vs. two-parent family) and SES, finding that ASC significantly predicted educational attainment in two of the three cohorts. Similar results were reported in a longitudinal study conducted to investigate the relationship between the components of the expectancy-value theory (ASC and task values) and educational outcomes that found that ASC predicted educational attainment as good as achievement and better than IQ and task values (Guo et al., 2015). Possible mechanisms underlying the positive relationship between ASC and educational attain-

ment have been proposed by Marsh and O'Mara (2008), who pointed out the behavioral implications of having a higher ASC such as a reduction in test anxiety, selecting more advanced courses or less chance of dropping out of school.

ASC and occupational aspirations. Research concerning the expectancy-value theory has identified ASC as a necessary prerequisite for occupational choices, although the equal importance of both components of the theory has been stressed (Eccles, 2009). The relationship between ASC and occupational aspirations was furthermore investigated in the study of Guo et al. (2015), finding that ASC was a stronger predictor of occupational aspirations than achievement and the only one consistent over all data waves. In a similar study, this relationship was also investigated when simultaneously considering academic achievement, finding that ASC almost completely mediated the relationship between achievement and career aspirations (Nagengast & Marsh, 2012).

Additionally, studies investigating possible negative effects of school-average ability on educational outcomes such as coursework selection (Marsh, 1991), educational attainment and occupational aspirations (Marsh, 1991; Marsh & O'Mara, 2010) have found that ASC mediates those negative effects, confirming once more the relationship between ASC and those outcomes.

1.2.3 Differences between Students with and without Immigrant Background

Many presumptions make it seem obvious that students of immigrant background should have a lower ASC than native students. The most apparent is that students of immigrant background very often show lower levels of academic achievement than their non-immigrant classmates (e.g., Organization for Economic Co-operation and Development [OECD], 2010). In accordance with the REM (Marsh, 1990), this should result in a lower ASC. More generally, students of immigrant background have a higher chance of having to repeat a year or being held back from starting school in the first place (Ehm et al., 2011). In short, they more often experience failure in the academic context, which is likely to negatively affect their ASC (Shajek et al., 2006). Other researchers assumed that the acculturative stress stemming from immigration, minority status, and sociocultural disadvantages might have a direct negative effect on students' confidence in their abilities (Roebers et al., 1998; Shajek et al., 2006).

Given these sound arguments, it is surprising that many studies have yielded results that contradict these presumptions. Concerning ASC in the mathematics domain, for example, it was consistently found that students of immigrant background have a higher ASC when

their usually lower academic achievement is controlled for (e.g., Ehm et al., 2011; Faber et al., 2011; Faber & Billmann-Mahecha, 2012; Marsh, 2016; Roebers et al., 1998; Shajek et al., 2006). Results concerning verbal ASC are less consistent, with some studies finding that students of immigrant background had lower verbal ASCs than their native classmates (Roebers et al., 1998; Shajek et al., 2006), while in some, ASC did not differ despite lower performances (Faber & Billmann-Mahecha, 2012) or was even higher when statistically controlling for achievement (Ehm et al., 2011; Faber, 2013). Studies investigating differences in general ASC are sparse. Seo et al. (2019) conducted a study where they did not focus on immigrants per se but compared the ASCs of minority (Black and Latinx) and White students. Besides math and verbal ASC, they considered a general ASC and found that Black students scored significantly higher in all three measures of ASC and Latinx students demonstrated similar ASCs as White students, while both minority groups showed lower levels of achievement (Seo et al., 2019). It can be expected that Latinx students also would have shown higher ASCs if the individual achievement levels had been statistically controlled for. In summary, while there are definite differences between students of immigrant background and their native classmates, it is notable that none of the above-mentioned effects exceed small effect sizes.

Another interesting result that has been found in several studies is that the relationship between ASC and achievement is not as strong for students of immigrant background, as it is for their native classmates (Faber et al., 2011; Nouwen & Clycq, 2019; Roebers et al., 1998; Shajek et al., 2006). It has been argued that this might even be the reason for the discrepancy between the high ASCs in contrast to the actual achievement of students of immigrant background, as they are, for some reason, not able to exploit the maximum potential of their positive ASCs (Seo et al., 2019). Other authors take this weakened relationship as an indicator that students of immigrant background or other stigmatized groups psychologically disengage their ASC from their achievement to protect themselves from negative stereotypes (Nouwen & Clycq, 2019).

Other attempts to explain the counterintuitive findings of high ASCs of students of immigrant background consider compensation effects that are in line with the I/E model, i.e., students of immigrant background comparing their good or average math achievements to their low verbal achievements, which results in a higher math ASC (Roebers et al., 1998; Shajek et al., 2006). However, considering that also a higher verbal and general ASC of students of immigrant background have been demonstrated (Ehm et al., 2011; Faber, 2013; Seo et al., 2019), this explanation is unlikely. Indeed, in their study about the differences in math

ASC between students with and without immigrant background, Faber et al. (2011) explicitly tested the I/E model and found that it was only significant for the native students.

Two more common explanations of immigrant students' high ASCs are the *selective devaluation hypothesis*, proposing that students of immigrant background place less value in education and consider schoolwork to be less important, and the *external attribution hypothesis*, which suggests that students of immigrant background attribute their (low) achievement to external factors to protect their ASC (Seo et al., 2019). In their study, Seo et al. (2019) did not find convincing support for either of these explanations and proposed some hypotheses of their own, e.g., that positive ASCs are a manifestation of protective cultural resources that are acquired through resilience, sense of control and collective identity, which, in turn, are promoted through exposure to negative stereotypes and discrimination. Another hypothesis is that students of immigrant background (in their study they mostly focused on Black American students) are more likely to attend low-achieving schools than native students and therefore might profit from the BFLPE, in that they might be surrounded by less able classmates than equally able students attending high-achieving schools, which positively affects their ASC.

1.3 The Big-Fish-Little-Pond Effect (BFLPE)

It is conventional wisdom that schools, where the average ability level of students is high, are the most beneficial and provide the best education for students (Marsh, 1991). The BFLPE describes the negative relationship between the school-average ability and the ASC of students, consequently challenging this belief. The following chapter provides an overview of the BFLPE, its implications, underlying processes, and the characteristics and circumstances that might influence its size.

1.3.1 Definition and Implications of the BFLPE

ASC is a relativistic construct that is based on the evaluation of accomplishments in relation to frames of reference (Marsh et al., 2008). This presents the fundamental theoretical premise underlying the BFLPE, in that "perceptions of the self cannot be adequately understood if the role of frames of reference is ignored" (Marsh et al., 2008, p. 321). The frame of reference for the BFLPE is the school- or class-average ability and as such, it occurs when a student compares themself with more able students and forms a lower ASC than an equally able student who compares themself with less able students (Marsh, 1987) (see Figure 2). In Austria, such situations arise easily due to the explicit school-level tracking, where students are separated into either the new secondary school (*Neue Mittelschule* [NMS], low-ability

track), or the lower cycle of the academic secondary school (*Allgemein bildende höhere Schule* [AHS], high-ability track), according to their achievement at the age of 10 (special needs education excluded) (Salchegger, 2016). As the BFLPE is dependent on differences in school- or class-average abilities, it can be concluded that it will be stronger the greater the variance of achievement is between schools or classes (between-school achievement variance [BSAV] / between-class achievement variance [BCAV], Salchegger, 2016). Following the revised model of ASC proposed by Marsh and Shavelson (1985), most researchers investigated the influence of the BFLPE on domain-specific (mainly math and verbal) ASCs. There have been studies that demonstrated the BFLPE on general ASC, though (e.g., Hoferichter et al., 2019; Marsh et al., 1988), which is the focus of the present study.

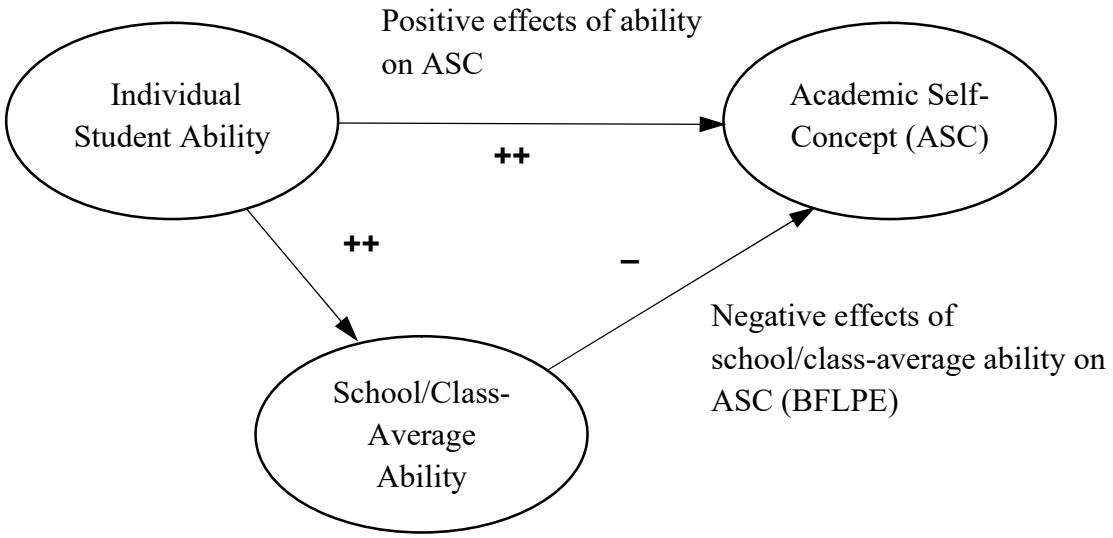

Figure 2. Theoretical predictions of the Big-Fish-Little-Pond Effect (BFLPE). Adapted from "Big-fish–little-pond effect on academic self-concept: A cross-cultural (26-Country) test of the negative effects of academically selective schools" by H. Marsh and K.-T. Hau, 2003, *American Psychologist, 58*(5), p. 369. Copyright 2003 by the American Psychological Association.

The BFLPE entails important practical implications, as the ASC is positively related to many educational outcomes, which is why it can reasonably be argued that it is important for students to form a positive self-image about their academic abilities. Although high-ability schools have been found to have slightly positive effects on academic achievement (Marsh, 1984; Marsh & Parker, 1984), according to the BFLPE, they negatively affect ASC (e.g., Marsh, 1984; Marsh, 1987; Marsh & Parker, 1984), posing the question of whether or not possible benefits of high-ability schools outweigh the disadvantages of having a poorer ASC. Furthermore, it has been argued that the benefits of attending high-ability schools have been overestimated because pre-existing differences between the students of high- and low-ability

school had not been taken into account (Marsh, 1991). If students' individual abilities are statistically controlled for, former positive effects of the school-average ability on academic achievement, educational and occupational aspirations become negative, whereby these effects are mediated in part by ASC (Marsh, 1991; Marsh & O'Mara, 2010). In conclusion, contrary to conventional wisdom, placing children in high-ability schools might not always be best for them.

1.3.2 Underlying Social Comparison Processes and the Generalized Other

One of the theoretical foundations of BFLPE theory is Festinger's social comparison theory (1954). Accordingly, social comparison processes are proposed to be the mechanisms underlying the BFLPE (e.g., Marsh et al., 2008). This has been subject to critique by Dai and Rinn (2008), who argued that even though using social comparison to explain the contrast effect, BFLPE theory does not heed all insight gathered by social comparison research. Social comparison processes do not necessarily lead to negative contrast effects but can also entail positive assimilation effects. The *reflected glory effect*, for example, posits that if a student perceives themself as similar to a slightly more able comparison target, they will profit from that target's perceived ability and develop a more positive ASC. Research has shown that social comparison is an active process and it is therefore unlikely that students would choose a comparison target that negatively affects their ASC (Dai & Rinn, 2008). In a reply to this critique, Marsh and his colleagues (2008) stated that they did not assume school- or class-average ability to be the only possible foundation for social comparison processes and acknowledged that contrast and assimilation effects can coexist. They pointed out, however, that there were crucial differences in the sources of social comparison information between social comparison research and BFLPE research: while in social comparison research students are typically asked to state a specific person to be their comparison target, the BFLPE is based on a situationally imposed comparison with a group or a *generalized other.* This generalized other can be operationalized as the school- or class-average ability, which serves as a relatively objective normative basis of comparison, that will, according to social comparison theory replace active social comparison strategies (Festinger, 1954, as cited in Marsh et al., 2008), which is the reason why the BFLPE will prevail over active social comparison processes (Marsh et al., 2008). Validation for these assumptions was provided by Huguet et al. (2008) who asked students in their study about their perceived relative standing in class (whether they felt to be better, worse, or average in math compared to their classmates). If the BFLPE was indeed based on comparison with the class-average ability, it should not be demonstrable

when students' perceived relative standing in class was statistically controlled for. This is exactly what they could establish, providing evidence that comparison with a generalized other is the underlying mechanism under the BFLPE (Huguet et al., 2008).

One of the first researchers to propose the concept of the generalized other was Mead (1934). He described the generalized other as the attitudes of others toward a person that are integrated by this person "into a single attitude or standpoint which may be called that of the 'generalized other'" (Mead, 1934, p. 90). By this conceptualization, it is likely that the generalized other will not only consist of the average ability of a student's classmates but that the (perceived) attitudes of their classmates about themself will also play a role. Mead further posits that "[t]he attitude of the generalized other is the attitude of the whole community" (Mead, 1934, p. 154), which allows the conclusion that attitudes or stereotypes that are prevalent in the society (cultural stereotypes) will reflect in the generalized other. It can be argued that in a society that is characterized by increasing multiculturality, stereotypes that concern ethnic groups may play a role in the formation of the generalized other, especially for students of immigrant background. Following this line of reasoning, it becomes apparent that it matters *who* a student goes to class with, i.e., which other people and their experiences and attitudes form their generalized other. In other words, the ethnic composition of the classroom might very well play a role in the formation of the generalized other and thus, for the BFLPE.

1.3.3 Stability, Generalizability, and Moderators of the BFLPE

There had been some doubt about whether the BFLPE was a stable, long-lasting effect or whether it was ephemeral (Dai & Rinn, 2008), but as of today, sufficient evidence that the BFLPE is a temporarily stable effect (Becker & Neumann, 2016, 2018; Marsh & O'Mara, 2010; Marsh et al., 2007; von Keyserlingk et al., 2019) and that it might even prevail in cross-context situations (Becker & Neumann, 2018; Marsh & O'Mara, 2010; Marsh et al., 2007) has been provided. Marsh et al. (2007), for example, conducted two studies to investigate whether the BFLPE was still demonstrable two and four years after high school graduation. They found that the negative effects of school-average ability on ASC did indeed prevail after graduation, whereby these effects were largely mediated by the ASC at the end of high school. Two years after graduation, they even found a small direct negative effect of school-average ability on ASC that could not be explained by previous ASC (Marsh et al., 2007). Similar results were obtained by Marsh and O'Mara (2010), who explored five data waves spanning eight years and found that ASC was still negatively affected by school-average ability one year after high school graduation. These findings suggest that the BFLPE is a long-

term effect that does not simply dissipate after leaving an academic context. Partial support for this has also been found by Becker and Neumann (2018), who investigated whether or not the negative effect of the school-average ability of elementary school on ASC would remain parallel to a new BFLPE after the transition from elementary to secondary school. They considered general, verbal, and math ASCs and found that the BFLPE from primary school persisted only for the general ASC, while for verbal and math it was replaced by a new BFLPE that used the secondary school-average ability as a frame of reference. In a similar study of the same authors, such parallel BFLPE effects could not be established but instead, the BFLPE of elementary school was only significant at the very beginning of secondary school, while the new BFLPE could be observed at the end of the first year of secondary school (Becker & Neumann, 2016). Von Keyserlingk and her colleagues (2019) also did not find any support for the persistence of the BFLPE through contextual changes when they investigated the transition from 10th to 12th Grade in German high schools, which goes in hand with enrolment in different course levels. Instead, they too found that the new environment was used as a frame of reference for the BFLPE. One possible reason for this discrepancies might be that in the studies of Marsh et al. (2007) and Marsh and O'Mara (2010) the contextual change was not necessarily of academic nature, meaning that it was not further specified whether students attended colleges, vocational schools or even entered the job market, while the other studies specifically investigated the transition from one classroom context into another (Becker & Neumann, 2016).

In addition to temporal stability, the cross-cultural generalizability of an effect is of special interest. For the BFLPE, this had first been demonstrated by Marsh and Hau (2003) who investigated data from 26 countries and found a significant negative effect of school-average ability on ASC in all but two. Seaton and her colleagues (Seaton et al., 2009) extended these findings by including even more countries that were selected with care to include both individualistic and collectivistic cultures, as well as economically developing and already developed nations. Of these 41 countries, the relationship between school-average ability and ASC was negative in all of them and statistically significant in 38. In short, overwhelming support was found for the cross-cultural generalizability of the BFLPE, leading to Seaton and her colleagues (2009) to label it a pan-human effect.

Although establishing the existence of the BFLPE in 41 countries, Seaton et al. (2009) also found a small but significant random effect on the country level, meaning that the size of the BFLPE varied across countries. This finding was replicated in a meta-analysis of 33 studies yielding 56 effect sizes (Fang et al., 2018), indicating the highest BFLPE for Asian stu-

dents. These differences, however, could not be explained through the cultural orientation (individualism vs. collectivism) or economic development (Seaton et al., 2009). Alternative reasons for the cross-country variation might be found in the difference of school systems, as was investigated by Salchegger (2016), who proposed that explicit school-level tracking would go in hand with stronger BFLPE and therefore explain the differences. The hypothesis was confirmed, establishing school-level tracking as a significant moderator of the BFLPE (Salchegger, 2016). Many more moderators have been investigated by different researchers throughout BFLPE research, most yielding only small or non-significant effects (e.g., Fang et al., 2018; Marsh, 2016; Schwabe et al., 2019; Seaton et al., 2010). Seaton et al. (2010), for example, investigated 16 individual student characteristics that might moderate the BFLPE. Those characteristics included social relationships, which are viewed as especially interesting for the present study, as increased multiculturality is believed to be reflected in social relationships in school. Specifically, Seaton et al. (2010) investigated *sense of belonging to school* (a construct that describes how connected a student feels to other students) and student-teacher relationships. While the latter was not found to be a significant moderator of the BFLPE, the former was, although the effect size was deemed too small to be of further interest (.041). Four of the other investigated moderators had large enough effect sizes to be of substantial value, revealing that the BFLPE was greater for highly anxious students, those who used memorization as a learning strategy or reported a high cooperative orientation. Higher ability levels also were associated with a stronger BFLPE, however, as there had been inconsistent findings regarding individual ability level in the past, and as the effect size was small (<.10), the authors concluded that the BFLPE generalizes well across ability levels (Seaton et al., 2010). Schwabe and her colleagues (2019) also investigated social relationships as potential moderators of the BFLPE and found contradicting results to those of Seaton and her colleagues (2010), where peer relationships had no significant interaction with school-average ability, but students with positive teacher relations were significantly less affected by the BFLPE. Gender had also been considered a potential moderator of the BFLPE, but has been found to only slightly and inconsistently interact with school-average achievement, leading the authors of the study to conclude that the BFLPE generalized well over gender (Marsh et al., 2007). Further significant moderators have been found in the meta-analysis of Fang et al. (2018), namely student age, with students in high school (Grades 9-12) experiencing a higher BFLPE than middle school (Grades 5-8) and college students, while elementary students reported the smallest effects. Variations within a school type have not been investigated. Another significant moderator was the ASC domain, where the BFLPE was strongest when

verbal ASCs were measured, followed by so-called STEM (Science, Technology, Engineering, Mathematics) ASCs and general ASC showed the weakest effect (Fang et al., 2018). An especially interesting finding for the present investigation was reported in Marsh's (2016) study, which focus was a different frame of reference effect (the *relative-year-in-school effect*). While investigating both effects to prove that they were independent of each other, Marsh found that there was a small but significant interaction between the BFLPE and immigrant background, in that students of immigrant background were slightly less affected by the BFLPE than their native classmates (Marsh, 2016).

In summary, there have been found quite a few individual and school-level characteristics that influence the strength of the BFLPE, although none were able to reverse the effect. While this provides very strong support for the generalizability of the BFLPE and its right as being labelled a pan-human effect, it is still of interest to examine possible moderation effects, as the knowledge about *which individual* under *what circumstances* (Dai & Rinn, 2008) is especially affected is necessary to develop possibly counteracting interventions (Dai & Rinn, 2008; Seaton et al., 2010). In light of the increasing multiculturality in the society and its possible implications for the generalized other immigrant background is considered an interesting candidate for a moderating variable on the individual level, especially as it had already been found to moderate the BFLPE (Marsh, 2016). For unknown reasons, this finding has never been further investigated. Referring to the possibility of social relationships affecting the BFLPE (Schwabe et al., 2019; Seaton et al., 2010), the ethnic composition of the classroom, specifically, the proportion of students of immigrant background in the class, is proposed as a potential moderator on the class level. The main goal of the present study is to examine whether these two variables influence the BFLPE.

1.4 Cultural Stereotypes

Stereotypes are the cognitive component of attitudes towards people and mean specific attributes or traits that are generalized over a group of people without consideration for the individuality of the group members (Aronson et al., 2017). Cultural stereotypes label stereotype content that results from shared public views of groups or societies (Fiske et al., 2002). Stereotypes about immigrants are often negative, for example, immigrants are thought to be less educated and more often unemployed than natives (Alesina, Miano, & Stantcheva, 2018) and are perceived as incompetent and untrustworthy (Lee & Fiske, 2006). In the stereotype content model (Fiske et al., 2002), groups of people (e.g., elderly, businesswomen, rich people, migrant workers, etc.) can be rated on two dimensions: warmth, which assesses the intent

of an out-group toward the in-group, and competence, which refers to the capability of the out-group in realizing this intent and entails beliefs about competence and respectability. The generic immigrant or refugee is rated as lacking both warmth and competence (Kotzur et al., 2019; Lee & Fiske, 2006). If immigrants or refugees are considered as separate groups according to their country of origin or their flight motives, a more differentiated image about their warmth and competence is formed, with, for example, Asian immigrants being perceived as more competent than Latinx immigrants (Lee & Fiske, 2006) and economic refugees more competent than war refugees (Kotzur et al., 2019). However, the majority of immigrant groups are still perceived as less competent than natives (Froehlich & Schulte, 2019).

1.4.1 Implications for Students of Immigrant Background

Stereotypes about immigrants can affect students of immigrant background in many ways. Alesina and his colleagues (Alesina, Carlana, et al., 2018), for example, have found that teachers with negative stereotypes about immigrants give lower grades to students of immigrant background compared to their native classmates that performed equally as well on standardized, blindly-graded tests. Similarly, Froehlich et al. (2018) have found that negative stereotypes significantly predicted preservice teachers' (teachers that are still in training but already have teaching experience) attribution of underperformance to characteristics of immigrant students instead of to the educational system. This discrimination, that students of immigrant background are at risk to suffer, has grievous consequences besides the obvious effect on their grades. Students of immigrant background might feel less self-confident about their abilities or are discouraged to put effort into their schoolwork, which in turn likely affects their actual performance. This leads to a self-fulfilling prophecy that might impact immigrant students' educational and occupational chances (Alesina, Carlana, et al., 2018).

Not only the actual but also the fear of meeting such negative expectations can negatively affect students of immigrant background, as has been proposed by *stereotype threat theory* (e.g., Steele, 1997). Stereotype threat means the fear of specific group members to confirm a stereotype about their group and, as a consequence, being reduced to that stereotype (Steele, 1997). Immigrant background as a stereotyped group constitutes somewhat of a special case, firstly because the group membership is not as obvious as it is for sex or race, and secondly because the valence and content of stereotypes varies across different immigrant groups (Appel et al., 2015). In a meta-analysis specifically concerning immigrants, Appel and his colleagues (2015) found conclusive support that people of immigrant background are affected by stereotype threat as much as other minority or gender groups. Individuals usually

are highly sensitive to cues that indicate negative stereotypes about their social group, which is why these cues do not need to be especially explicit or blatant, but indirect and subtle cues are sufficient to induce stereotype threat (Spencer et al., 2015; Walton & Cohen, 2003), which is why it can be called "a threat in the air" (Steele, 1997). If a stereotype threat becomes chronic, it can lead to disidentification, i.e., to "a reconceptualization of the self and one's values" (Steele, 1997, p. 614), which usually entails the removal of the domain in which the stereotype is settled, e.g., the academic context, as a basis of self-evaluation. An example given by Steele (1997) describes a competitive, male-oriented math environment as a potential situation in which the stereotype threat might become chronic for women. Following this example, it can be argued that a classroom full of non-stereotyped, i.e., native peers, can constitute such a chronic stereotype threat situation for students of immigrant background. Disidentification might be one of the reasons for the weakened correlation between ASC and academic achievement that has been described for students of immigrant background, which might explain their relatively high ASCs. On the other hand, stereotype threat might directly affect ASC, as it is likely to diminish a person's perception of their abilities in the stereotyped domain (Spencer et al., 2015).

Evidence that stereotypes which disparage immigrants' intelligence or competence might negatively affect the ASC of students of immigrant background has also been found by Lehman (2012). He investigated under what circumstances ethnic minorities (Asian, Black, & Latinx) in the USA would report a decreased ASC and what role perceived discrimination and the heterogeneity of students' friendship groups would play. While neither perceived discrimination nor characteristics of friendship groups influenced ASC for Asian and Latinx students, Black students reported a higher ASC when they had homogeneous friendship groups, supporting the hypothesis that students with similar experiences provide social support for each other. Perceived discrimination as well as heterogeneous friendship groups lead to decreases in the ASC of Black students, suggesting that negative stereotypes become more salient with increased friendship heterogeneity (Lehman, 2012). In other words, having more White friends in their friendship group made negative stereotypes more salient for Black students. Expanding this finding to the classroom context, it can be argued that negative stereotypes about minority groups will become more salient, the more majority group members attend the same class.

Further research about the impact of negative stereotypes on ASC seems to be sparse, however, it is assumed that cultural stereotypes are internalized by students of immigrant

background in a way that influences their intellectual abilities and motivations (Diehl & Fick, 2016), which likely includes their ASC.

1.4.2 Implications for Students without Immigrant Background

Stereotypes diminishing the intellectual abilities of out-groups can not only negatively affect these group's members, but they can also positively affect in-group members - a phenomenon called *stereotype lift effect* (Walton & Cohen, 2003). In a meta-analysis of 43 studies, Walton and Cohen (2003) found that if negative stereotypes were linked to intellectual tests, members of non-stereotyped groups performed significantly better than when no such link was made. In line with findings of stereotype threat research, negative stereotypes seemed to be linked to evaluate tasks automatically, meaning that no explicit cue was necessary to induce the stereotype lift effect. The underlying mechanism of the stereotype lift effect is proposed to be downward social comparison with negatively stereotyped group members, resulting in an elevation of self-efficacy or self-esteem, which boosts the performance of the non-stereotyped group members (Walton & Cohen, 2003). ASC is not explicitly mentioned as a potential mediator of the stereotype lift effect, but as it is based on social comparison processes, it is likely to be influenced by downward comparison. Also, the relationship between ASC and academic performance has been established many times. It can be therefore argued that ASC is at least as likely to mediate the relationship between negative stereotypes about out-groups and the performance boost of in-group members, as is self-efficacy. In conclusion, ASC is likely to be positively affected by the stereotype lift effect.

Empirical evidence about students of immigrant background serving as a base for the stereotype lift effect has been found by several studies, whereby individual characteristics such as political orientation (Chatard, Selimbegović, & Konan, 2008), prejudice and stereotype endorsement (Chatard, Selimbegović, Konan, & Mugny, 2008) seem to moderate the effect. In a study concerning stigmatized native students (females, when considering math performance and students with low SES, when considering general intelligence), it was found that those students did indeed use students of immigrant background as a basis for downward comparison, in that they performed significantly better when comparing themselves to immigrant female or low SES students, respectively, than to a native pendant (Konan et al., 2011).

In summary, cultural stereotypes are likely to affect the ASC both of students with and without immigrant background due to the described stereotype threat and lift effects. Whether or not they also influence students' susceptibility to the BFLPE due to their reflection in ethnically diverse classrooms shall be investigated in the present study.

1.5 The Present Investigation

Increasing multiculturality goes hand in hand with many societal and institutional challenges and opportunities that also concern students' education. The changing environment in terms of growing ethnic diversity in the classroom is likely to affect both students with and without immigrant background. It is therefore of special interest to look further into how students are affected and whether there are differences between students with and without immigrant background.

The focus of the present study lies on students' ASC, as it has been established as a tremendously important predictor for diverse educational outcomes. In the light of cultural diversity research, it is especially interesting to consider a general ASC, as this domain of ASC might be uniquely influenced by peer relation or other social aspects of the school environment. While studies considering math or verbal ASCs have consistently found differences between the ASC of students of immigrant background and their native peers, studies investigating differences in general ASC are sparse. The only study that could be found to report differences in general ASC investigated Black and Latinx minority groups and not students of immigrant background per se (Seo et al., 2019). The first goal of the present study is to investigate whether general ASC differs between students with and without immigrant background (*Research question 1*). Considering that students of immigrant background have been found to report higher ASC in different domains in the past, it is expected that students of immigrant background will report a higher general ASC than their native peers if students' individual achievement levels are statistically controlled for (*Hypothesis 1*).

When investigating ASC, it is especially interesting to focus on the BFLPE, as it has been established as a pan-human phenomenon that negatively affects ASC and other educational outcomes. Most of the studies investigating the BFLPE have focused on domain-specific (mainly math or verbal) ASCs. The present study's focus lies on general ASC, as it is believed that this will be susceptible to influences of peer relations. Although the generalizability of the BFLPE has been demonstrated over many different individual and school-level characteristics, it is argued that it is still relevant to investigate *which individual* under *what circumstances* is especially susceptible to the BFLPE, as this knowledge is valuable when developing interventions (Dai & Rinn, 2008). The main goal of the present study is therefore to consider two more possible moderators of the BFLPE: students' immigrant background (*which individual*) and the proportion of students of immigrant background in the classroom (*what circumstances*). These variables are of special interest for several reasons: The key underlying process of the BFLPE is posited to be social comparison with a generalized other. As

the generalized other is formed through students' classmates, it will be influenced by *who* attends the class, i.e., whether students have an immigrant background and what their attitudes are. It is proposed that cultural stereotypes will be reflected in these attitudes. Negative stereotypes about the intelligence and competence of immigrant students will likely be more salient, the more native classmates they have. It is therefore apparent that students of immigrant background and students in classrooms with different proportions of students of immigrant background will form different generalized others and this is likely to influence the size of the BFLPE. Furthermore, social relationships have been investigated as potential moderators of the BFLPE before, with inconsistent findings. The consideration of the ethnic composition of the classroom (proportion of immigrant students) as a moderator could contribute insights to this line of research. Additionally, immigrant background has been identified as a significant moderator of the BFLPE before (Marsh, 2016), without being investigated further. The following research questions arise: Does the size of the BFLPE differ for students with and without immigrant background (*Research question 2*) and does it vary with different proportions of students of immigrant background in the classroom (*Research question 3*)? As Marsh (2016) has found that students of immigrant background were slightly less affected by the BFLPE than their native classmates, the same direction is expected to be found in the present research (*Hypothesis 2*). Since negative stereotypes are likely to be more salient when there is a high proportion of native students in the classroom, which could lead to a stronger stereotype threat effect, it is hypothesized that students of immigrant background will experience a stronger BFLPE, the more native students are in the classroom (*Hypothesis 3a*). For native students, on the other hand, a high proportion of students of immigrant background in the classroom might enhance the stereotype lift effect, which is proposed to diminish the BFLPE (*Hypothesis 3b*). In other words, for both students with and without immigrant background, it is expected that the higher the proportion of students of immigrant background in the classroom is, the smaller the BFLPE will be (*Hypothesis 3*).

2 Methods

2.1 Participants and Sample Size

Participants were 1,649 students from 87 classrooms (class size seven to 28 students, M = 18.95, SD = 4.04, see Table A in Appendix A) in 10 secondary schools in eastern Austria (in the federal states Vienna, Lower Austria and Burgenland). The students attended the Grades 6-8, their ages ranged from 10 to 16 years (M = 12.41, SD = 1.06). The sample was a convenience sample of schools selected with care to get an even distribution of high and low

immigrant proportions. Of the participants, 50.6% stated their gender as female, 47.8% as male, with 1.6% missing values. The percentage of students of immigrant background was 42.3 with 0.8% missing values. Most of the students of immigrant background were born in Austria (59%), 40.4% were born abroad and are therefore first-generation immigrants, and 0.6% of the values were missing. In total, the students represented 74 countries of origin (see Table B in Appendix A), with the largest proportion being from Turkey (14.6%), followed by Serbia (9.5%), and Slovakia (6.9%). The two main secondary school tracks in the Austrian school system were represented, with 58.8% of all participants (80.5% of those with immigrant background) attending a NMS and 41.2% (57.6% of the native students) attending an AHS (see Table C in Appendix A).

2.2 Procedure and Design

The present study uses cross-sectional data that were collected in October 2019 as part of another study investigating the role of friendships in acculturation and development of adolescents with immigrant background in Austria. Formal approval from the educational departments in Vienna, Lower Austria and Burgenland was obtained and schools in these states were contacted and asked for participation. After obtaining permission from the school principals, written informed consent was asked of students' parents or legal representatives. The final decision to participate was left with the students. Data were collected via paper and pencil questionnaires in the respective classrooms during regular class hours. The instruction language was German, which is the national language of Austria. The questionnaire consisted of two parts, a general part to be filled out by all students and a specific part consisting of items to assess ethnic identity and acculturation for students of immigrant backgrounds. The whole questionnaire took about 30 min to complete, whereas the actual completion time varied across classrooms from approximately 25 to 50 min. If possible, two trained research assistants were present per class to ensure a smooth procedure with one giving instructions and the other one distributing and collecting the questionnaires and assisting in answering questions that arose. The data were transferred manually into digital form for further use.

2.3 Measures

The whole questionnaire was comprised of 13 scales with a maximum of 112 items. The following measures were used in the present study:

2.3.1 Academic Self-Concept

General ASC was measured via the absolute ASC scale of the students version of the school-related self-concept scales (SESSKO, Schöne et al., 2002). The SESSKO was developed to measure four facets of ASC (three facets concerning different norms for comparison and one absolute scale) in German speaking countries and could therefore be used in its original form. All four scales of the SESSKO have successfully been used before to demonstrate the BFLPE (Hoferichter et al., 2018). The absolute ASC scale consists of five items measured on a 5-point semantic differential scale and decidedly foregoes comparison to any given criterion (e.g., "I am … not intelligent (1) / very intelligent (5)). Validation on 216 and 438 students (Grades 5-10) resulted in the absolute ASC scale showing an internal consistency of .81 and .84, respectively (Dickhäuser et al., 2002). In the present sample the absolute ASC scale showed an equally good reliability of α=.84.

2.3.2 Academic Achievement

Students were asked to state the grades of their latest school report (June 2019) in mathematics, German and English (as a foreign language). In Austria, a 5-point grading scale is used, where 1 ("very good") represents the highest achievement and 5 ("unsatisfactory") the lowest. To simplify interpretation, the grades in each of the subjects were inverted so that higher numbers indicate better achievements. Grades show a high correlation with ASC and are easily accessible for social comparison (Marsh et al., 2007; Vogl et al., 2018), therefore providing a more direct source of feedback for students than standardized achievement tests (Marsh et al., 2007). A disadvantage is that grades do not vary along a metric that is common to all students in all classes and schools, as they are subject to the *grading-on-a-curve effect* (Marsh et al., 2007). This describes the phenomenon that teachers tend to use classroom-related social frames of reference for grading, in that they compare each student to the highest and/or lowest achieving students in their classes and grade them accordingly. This results in a low variation of grades between classes, which consequently leads to the possibility of the BFLPE being underestimated (Marsh et al., 2007; Vogl et al., 2018). Grades are the most suitable achievement indicator for investigating the BFLPE if they are corrected for the grading-on-a-curve effect by using objective indicators of achievement to rescale class-average grades (Vogl et al., 2018). No objective indicators of achievement are available in the present study, which makes a correction impossible. As previous studies were able to demonstrate the BFLPE also when using uncorrected grades (e.g., Marsh et al., 2007; Vogl et al., 2018), it is argued that they will suffice in the present study.

2.3.3 Immigrant Background

Immigrant background was represented in a dichotomous variable with the value 0 representing native students and 1 students of immigrant background. Students were asked to state their own and their parents' countries of birth to obtain information about their immigrant background. As some students were still young, the possibility of them not knowing where their parents were born was taken into consideration and additional information about their immigrant background was gained by asking what language they spoke most often at home. In line with the official definition of Austria, which follows the recommendations given by the UNECE, students were classified as being of immigrant background if they or both their parents were born abroad (UNECE, 2015, p. 136). If students did not state the country of birth of one of their parents, they were still defined as being of immigrant background, if the other parent was stated to have been born abroad and the language that was most often spoken at home was not German. If information of both parents' countries of birth were missing and the student's country of birth was Austria or not stated, immigrant background was declared as unknown (missing).

2.3.4 Ethnic Classroom Composition

The interesting characteristic of ethnic classroom composition in the present study is the proportion of students of immigrant background (subsequently referred to as immigrant proportion). This was calculated by dividing the number of immigrant students through the total number of students in a class, resulting in a value between 0 and 1, with a higher number indicating a higher immigrant proportion.

2.3.5 Control Variables

There were some variables that were not of specific interest in the present study, but that have previously been found to influence ASC and/or the BFLPE. These variables were tested for statistical significance in the present sample to control for possible effects that could confound the investigated correlations.

Age. Students were asked to state their age at the beginning of the questionnaire. ASC has been found to vary with age but is thought to reach a stable level at the beginning of secondary school. Age has further been found to significantly moderate the BFLPE. It could be argued that students age would not matter in the present sample, as they all attend a school in the secondary level 1 (equating middle schools), but as it cannot be ruled out that there might be variation within a school type, both main and interaction effects of age were investigated

for significance. The main effect of age on ASC proved to be significant under the statistical control of gender, individual achievement, and class-average achievement (b = .043, p = .024). The interaction with class-average achievement (moderation of the BFLPE) was not significant (b = -.039, p = .496). Only the main effect of age on ASC was included in further analyses.

Gender. Students were asked to state their gender at the beginning of the questionnaire. The responses were coded as 0 for female, 1 for male. There was no "diverse" option for students' who might not identify as either male or female. Differences concerning gender have been found for ASC. BFLPE research has found no conclusive support for gender as potential moderator. Thus, only the main effect of gender was tested in the present sample. Gender was found to significantly predict ASC under the statistical control of age, individual achievement, and class-average achievement (b = .193, p < .001) and was included in further analyses.

2.4 Statistical Analyses and Missing Data

The BFLPE is a context effect[1] and as such it requires the simultaneous modelling of relationships among variables located at the individual and the classroom level, which can be achieved by a multilevel modelling approach (Marsh et al., 2008), which also takes into account the dependency of data. A dependency occurs whenever data is nested in a hierarchical structure, for example, students in a classroom as they tend to be more similar to each other than to students from different classrooms. Non considering this dependency would lead to an underestimation of standard errors which, in turn, would result in inflated levels of Type I errors (falsely rejecting the null hypothesis) (Hox, 2010).

The doubly-manifest model proposed by Marsh et al. (2009) was used to estimate the BFLPE. The disadvantage of this model is that it does not account for measurement and sampling errors, but it is likely to be more stable than doubly-latent models that do control for this errors. The decision to use the doubly-manifest model was made because the small amounts of bias, that are likely to result, are probably less worrisome than the potentially unstable estimates of the doubly-latent model (Marsh et al., 2009). As a manifest variable, general ASC was calculated by averaging across the scores of the five SESSKO items. Individual academic achievement was calculated by averaging across the three grades for each student. Corre-

[1] A context effect occurs, when a group level variable (class-average achievement) that is formed through aggregation of an individual level variable (individual achievement) significantly affects an outcome variable (ASC), even after the effect on the individual level is controlled for (Raudenbush & Bryk, 2002).

spondingly, class-average achievement was calculated by averaging across the three grades of all the students in a class.

To test *Hypotheses 2* and *3*, multiple group analyses with immigrant background as group variable were conducted. In multiple group analysis, two models are calculated: In the first one, parameters are estimated separately for each group, i.e., they are allowed to vary between groups (unconstrained model). In the second, parameter estimates are held invariant across the groups (constrained model). To test whether differences in the parameter estimates between the groups are significant, a chi-square difference test is conducted to compare the model fits of the constrained and unconstrained model. If the constrained model fits the data better or equally as well as the unconstrained model, no significant differences exist between the groups (Weiber & Mühlhaus, 2014). The loglikelihood is used in the present study to conduct the chi-square difference test, using the following formula:

$$\chi^2 = -2 * (L_0 - L_1) / cd$$

where L_0 is the loglikelihood value for the constrained model, L_1 is the loglikelihood value for the unconstrained model and *cd* is the difference test scaling correction that is calculated by the formula:

$$cd = (p_0 * c_0 - p_1 * c_1) / (p_0 - p_1)$$

with p_0 and p_1 being the numbers of parameters and c_0 and c_1 being the scaling correction factors in the constrained and unconstrained models, respectively ("Mplus", 2020). The resulting chi-square score was tested for significance with the online-calculator of Social Science Statistics ("Social Science Statistics", 2020).

The standard procedure of multiple group analysis assumes that the data of groups are independent to each other. Using a level-1 variable as a group variable raises two problems: (1) Level-2 data like class-average achievement or immigrant proportion are constant for all students who attend the same class, and thus are not independent between immigrant and native students in the same classroom. (2) The dependency of data due to cluster membership (increased similarity of students in a classroom) is not accounted for when calculating parameter estimates separately for each level-1 group (Ryu, 2015). Therefore, using a level-1 variable as group variable will likely result in an underestimation of standard errors which, as has been mentioned at the beginning of this chapter, will result in a higher probability of producing false positives (Hox, 2010). As there is no easily available solution to this problem, this must be borne in mind when interpreting the results of the analyses.

Effect sizes for level-2 predictors in multilevel models are calculated with the following formula that has first been proposed by Tymms (2004) and was used in several studies that investigated the BFLPE (e.g., Marsh et al., 2009; Seaton et al., 2010; Vogl et al., 2018):

$$\Delta = 2 * b * SD_{predictor} / \sigma_e$$

where b is the unstandardized regression coefficient in the multilevel model, $SD_{predictor}$ is the standard deviation of the level-2 predictor variable, and σ_e is the residual standard deviation at level 1. The effect size is comparable to Cohen's *d* (Marsh et al., 2009; Seaton et al., 2010; Vogl et al., 2018), with effect sizes of .20 being considered small, .50 medium, and .80 large (Cohen, 1988). As effect size for level-1 predictors, the standardized regression coefficient β is used, which is equal to the partial correlation *r* between the criterion and the predictor variables, i.e., it is the correlation between a certain predictor and the criterion, when all other predictors are controlled for. For *r*, the values of .10, .24, and .37 are considered small, medium, and large effect sizes, respectively (Fritz et al., 2012). Both the unstandardized regression coefficient (*b*) and the effect size (Δ or β) are reported in the results section.

The percentage of missing data on the seven variables used in analyses ranged from 0.0% to 1.6% (see Table D in Appendix A). Sixty-eight cases (4.12%) were incomplete, accounting for a total of 0.68% missing values. All analyses were conducted with Mplus Version 8.4 (Muthén & Muthén, 1998-2019), which uses full information maximum likelihood (FIML) to estimate models under missing data theory using all available data (Muthén & Muthén, 1998-2017). The FIML procedure automatically excludes cases with missing values on predictor variables from analyses.

2.4.1 Centering method

Regression parameters are interpreted as the value of the outcome variable *y*, when all predictor variables x_i are 0 (intercept) and the relationship between the outcome variable *y* and a predictor *x*, when all other predictors are 0 (slope). This interpretation is not plausible if variables do not have an intrinsically meaningful value 0 (usually metric variables such as age, achievement, etc.). These variables are recommended to be centered, as this allows the regression parameters to be interpreted at the mean value of predictor variables (Dawson, 2014). In multilevel analyses, two forms of centering are possible for level-1 variables: centering at the grand mean of the whole sample (grand-mean centering) or centering at the mean of the cluster *j* to which a case *i* belongs (group-mean centering). As level-2 variables do not deviate within a cluster, only grand-mean centering is possible for these variables (Enders &

Tofighi, 2007). Criterion (outcome) variables should not be centered, as this would lead to results that would fail to reflect the true variation of the variable (Dawson, 2014). Binary variables do have a meaningful value of 0 but are still recommended to be centered to avoid confusion in the interpretation of parameter estimates (Enders & Tofighi, 2007). Grand-mean and group-mean centering yield different parameter estimates and imply different interpretations of these parameters. Which centering method should be chosen depends on the particular research questions. For an extensive overview over the difference between the two centering methods and their application see Enders and Tofighi (2007). If level-1 predictors are the focus of interest, group-mean centering can be recommended. If the variable of interest is a level-2 predictor, grand-mean centering will be the better choice. It is possible that different forms of centering are required to answer different research questions in the same study (Enders & Tofighi, 2007). Context effects (such as the BFLPE) are one of the few exceptions were both centering methods produce equivalent parameter estimates in the sense that the parameter estimates can be algebraically equated (for details see Enders & Tofighi, 2007). Concerning the BFLPE, the estimated effect of the level-2 variable academic achievement on ASC is not a direct estimate of the BFLPE if group-mean centering of individual achievement is used (Marsh et al., 2009), but it is when grand-mean centering is used. To simplify interpretation, grand-mean centering is chosen to investigate the BFLPE in the present study.

In summary, the following centering methods are used in the present study: For *Research question 1*, where the main effect of immigrant background (level-1 predictor) on ASC is investigated, group-mean centering is used for all level-1 predictor variables. To replicate the BFLPE and answer *Research questions 2* and *3* (whether the BFLPE differs for students of immigrant background and whether this is moderated by immigrant proportion), grand-mean centering is used for all predictor variables.

The Mplus syntaxes used to investigate the research questions can be found in Appendix B.

3 Results

3.1 Preliminary Analyses

Bivariate correlations, descriptive statistics, and intraclass correlations (ICCs) of the investigated variables are shown in Table 1. ICCs were calculated to determine whether cluster effects were present in the data. The ICC for ASC was .04, meaning that approximately 4% of the variance can be attributed to differences on the class-level. While this seems to be little, similar ICCs have been reported by previous studies. Vogl et al. (2018), for example,

reported an ICC for ASC of .03. Similarly, Hoferichter et al. (2018), who used the SESSKO to measure ASC, found an ICC of .08 for the absolute ASC scale and ICCs of .02-.03 for the remaining three scales of the SESSKO. Both research groups were able to demonstrate the BFLPE in their samples. The variance of class-average achievement (BCAV) was calculated, as it is one of the preconditions for the BFLPE (Salchegger, 2016). The BCAV was 0.15, which is rather small compared to BSAVs in different countries reported by Salchegger (2016), which ranged from 0.04 - 0.62. However, the BFLPE was significant in all investigated countries of Salchegger's study, leading to the conclusion that the BFLPE can be present even if there are only small achievement difference between the investigated groups.

Table 1. Bivariate correlations, descriptive statistics, and ICCs.

	1	2	3	4	5	6	7
1. ASC							
2. Individual achievement	.312**						
3. Immigrant background	-.046	-.243**					
4. Age	-.052*	-.254**	.230**				
5. Sex	.027	-.232**	.142**	.123**			
6. Class-average achievement	.049**	.151**	-.166**	-.155**	-.078**		
7. Immigrant proportion	-.026**	-.066**	.158**	.067**	.042**	-.066**	
M	3.598	3.247	0.420	12.409	0.468	3.243	0.428
SD	0.752	0.883	0.494	1.063	0.500	0.387	0.261
Actual range	1 - 5	1 - 5	0 / 1	10 - 16	0 / 1	2.50 - 4.08	0.05 - 0.92
ICC	.040	.144	.233	.675	.052		

Note. *$p \leq .01$, **$p \leq .001$.

3.2 ASC - Differences between Students with and without Immigrant Background

A one-way ANOVA yielded that students of immigrant background reported a similar ASC compared to their native classmates ($F(1, 1{,}630) = 3.83$, $p = .051$, $\eta^2 = .002$, 95% CI [.000; .009]), while showing lower levels of achievement ($F(1, 1{,}608) = 79.38$, $p < .001$, $\eta^2 = .047$, 95% CI [.029; .069]; for further comparisons of students with and without immigrant background see Table C in Appendix A). To test whether they reported a higher ASC when individual achievement was controlled for (*Hypothesis 1*), a multilevel multiple regression analysis was conducted, with ASC as dependent, immigrant background as independent and individual achievement, age, and gender as control variables. The default estimator for multi-level regression models was used (maximum likelihood with robust standard errors, [MLR] Muthén & Muthén, 1998-2017). The sample size for this analysis was N = 1,581. Students of immigrant background reported a significantly higher ASC than their native classmates, confirming *Hypothesis 1*. The standardized regression coefficient β for immigrant background is .172, indicating a small effect size. Interestingly, age did no longer significantly predict ASC when immigrant background was included in the analysis. All parameter estimates are depicted in Table 2.

Table 2. Level-1 predictors of ASC.

Variable	*B*	β
Constant	3.600 (0.024)**	
Individual achievement	.445 (0.023)**	.481 (0.023)**
Immigrant background	.126 (0.045)*	.172 (0.06)*
Age	.006 (0.035)	.005 (0.022)
Sex	.195 (0.033)**	.266 (0.046)**
R^2	.226 (0.021)**	

Note. Standard errors in parentheses; *$p \leq .01$, **$p \leq .001$.

3.3 Replication of the BFLPE

Before testing *Hypotheses 2* and *3*, it was investigated whether the BFLPE could be replicated in the present sample. A multilevel multiple regression analysis was conducted that included all variables but the those of interest for *Hypothesis 2* and *3*, i.e., immigrant background and immigrant proportion (Model 0). The sample size for this analysis was N = 1,590. The regression coefficient of class-average achievement was small and not significant (-.067, p = .813, Δ = -.017, see Table 3 for standardized and Table E in Appendix C for unstandardized parameter estimates). The BFLPE could not be replicated in the present sample. Nevertheless, analyses were conducted to investigate *Research questions 2* and *3*, as it is possible that the inclusion of further variables could change the relationship between class-average achievement and ASC, leading to different results in more complex models.

3.4 Immigrant Background and Immigrant Proportion as Moderators of the BFLPE

To test *Hypothesis 2*, a multilevel multiple group analysis was conducted with immigrant background being the group variable, ASC being the dependent, class-average achievement the independent and individual achievement, age and gender being control variables (Model 1). The default estimator for multilevel multiple group models is MLR (Muthén & Muthén, 1998-2017). The sample size for this analysis was N = 1,581. The chi-square difference test of model fit yielded a non-significant result ($\chi^2(4, 1{,}581) = 6.740$, p = .150, see Table E in Appendix C), indicating that there were no significant differences in parameter estimates between students with and without immigrant background. Standardized parameter estimates for Model 1 are depicted in Table 3 (for unstandardized parameter estimates see Table F in Appendix C). The BFLPE was slightly positive and again not significant (.125, p = .612, Δ = .037). *Hypothesis 2* could not be confirmed.

Hypothesis 3 was also tested through multilevel multiple group analysis with the addition of immigrant proportion as a moderating variable (Model 2). The sample size for this analysis was N = 1,581. The chi-square difference test of model fit yielded a significant result ($\chi^2(6, 1{,}581) = 8.691$, $p < .001$, see Table E in Appendix C), indicating that there were significant differences in parameter estimates between students with and without immigrant background. Parameter estimates for the unconstrained Model 2 are depicted in Table 3.

Table 3. Standardized parameter estimates for models of the BFLPE.

Parameter	Model 0	Model 1 (constrained)		Model 2 (unconstrained)	
		Native	I. B.	Native	I. B.
Intercept	3.604 (0.018)***	3.568 (0.028)***	3.656 (0.027)***	3.540 (0.33)**	3.660 (0.034)**
Level-1 (individual)					
Individual achievement	.523 (0.024)***	.535 (0.024)***	.491 (0.023)***	.552 (0.027)***	.460 (0.036)***
Age	.058 (0.026)*	.055 (0.026)*	.064 (0.030)*	.097 (0.030)***	.018 (0.044)
Sex	.258 (0.045)***	.258 (0.044)***	.265 (0.045)***	.305 (0.058)***	.198 (0.064) **
Level-2 (classroom)					
Class-average achievement	-.067 (0.284)	.125 (0.247)	.125 (0.247)	-.051 (0.307)	-.488 (0.426)
Immigrant proportion				-.754 (0.370)*	-.415 (0.342)
Interaction				-.041 (0.256)	.094 (0.311)

Note. Standard errors in parentheses; I.B.= immigrant background; Interaction = interaction of class-average achievement and immigrant proportion; a chi-square difference test yielded no significant differences between the parameter estimates of native and students of immigrant background for Model 1; *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

Again, the BFLPE could not be established in the present sample (-.051, p = .868, Δ = -.018, and -.488, p = .252, Δ = -.133, for students with and without immigrant background, respectively). The main effect of immigrant proportion was significant only for students without immigrant background (-.754, p = .005, Δ = -.265), the interaction between immigrant proportion and class-average achievement (moderation of the BFLPE) was not significant for either group (-.041, p = .874, Δ = -.015, and .094, p = .762, Δ = .026, for students with and without immigrant background, respectively). *Hypothesis 3* could not be confirmed. Age significantly predicted ASC only for students without immigrant background (.097, p = .001).

4 Discussion

The purpose of the present study was to investigate differences in the general ASC of students with and without immigrant background (*Research question 1*) and to examine potential moderators of the negative relationship between class-average achievement and ASC (the BFLPE, *Research questions 2* and *3*). ASC is an especially relevant variable in the educational context, as it is related to many desirable outcomes such as academic achievement, coursework selection, educational attainment, and occupational aspirations (e.g., Marsh, 1991; Marsh & O'Mara, 2010). The BFLPE has been established as a pan-human phenomenon (Seaton et al., 2009) and has possibly wide-ranging practical implications (e.g., Marsh, 1991; Marsh & O'Mara, 2010). In the context of increasing multiculturality in Austria's classrooms (Braunsteiner et al., 2018), immigrant background and ethnic composition of the classroom (in terms of proportion of immigrant students) have been identified as interesting potential moderators of the BFLPE. Multilevel multiple regression analyses and multilevel multiple group analyses of cross-sectional data from 87 classrooms were used to investigate the hypotheses. Students of immigrant background were found to have a higher ASC than their native classmates. No significant moderators of the BFLPE could be confirmed. Below, findings are discussed in more detail before turning to potential limitations of the study and possible directions for future research.

4.1 Discussion of Results

Research question 1 of the present study was whether general ASC differed between students with and without immigrant background. The results of the one-way ANOVA yielded that students of immigrant background reported a similar ASC as their native classmates while showing lower levels of achievement, which is in line with findings of Seo et al. (2019), who investigated the ASCs of minority (Black and Latinx) students. Results of the multilevel

multiple regression analysis indicate that students of immigrant background report a higher general ASC than their native classmates, if individual achievement is statistically controlled for, confirming *Hypothesis 1*. Similar results have also been found in previous research concerning math (Ehm et al., 2011; Faber et al., 2011; Faber & Billmann-Mahecha, 2012; Marsh, 2016; Roebers et al., 1998; Shajek et al., 2006) and verbal ASCs (Ehm et al., 2011; Faber, 2013), specifically. Seo et al. (2019) proposed that the BFLPE might be an explanation for immigrant students' higher ASC, as they are more likely to attend low-achieving schools than native students and, consequently, have a higher chance of being surrounded by less able classmates. This pattern can be confirmed in the present sample, with 80.5% of students of immigrant background attending a low-achieving school (NMS) compared to only 42.4% of native students. As the BFLPE could not be replicated, this explanation cannot be verified, though.

The preconditions of finding the BFLPE in Austria are generally good, which is why it is unexpected that the BFLPE could not be demonstrated in the present sample. Firstly, in Austria an explicit school-level tracking system is implemented, which has been found to significantly explain cross-country variance in the size of the BFLPE, with countries with explicit school-level tracking showing higher BFLPEs (Salchegger, 2016). Secondly, considerably sized BFLPEs have been found in previous studies, with regression coefficients of -.23 (Marsh & Hau, 2003) and -.48 (Seaton et al., 2009). However, there are several possible explanations for the failure of replicating the BFLPE:

Sample. The present sample is much smaller (N = 1,649) than those of previous studies, who used data from PISA tests, resulting in sample sizes of N = 4,444 (Marsh & Hau, 2003) and N = 4,378 (Seaton et al., 2009), respectively. Furthermore, PISA data comprise a nationally representative sample (Marsh & Hau, 2003), while the present sample was a convenience sample with participants of only three out of Austria's nine federal states.

BCAV. The BCAV of the present sample is 0.15, which differs substantially from the BSAV for Austria that has previously been reported to be 0.55 (Salchegger, 2016). As the BFLPE has been demonstrated in countries with a BSAV of 0.05 (e.g., Finland, see Salchegger, 2016), the small BCAV cannot be the sole explanation for failing to demonstrate the BFLPE, but might have contributed to it.

Grading-on-a-curve effect. A likely reason for why the BCAV is so small in the present sample is that grades, which were uncorrected for the grading-on-a-curve effect, have been used to assess individual and class-level achievement. This probably results in similar distributions of grades over all classrooms and, consequently, in little variation between clas-

ses, which, in turn, leads to an underestimation of the BFLPE (Marsh et al., 2007; Vogl et al., 2018). While this underestimation will likely contribute to the failure of replicating the BFLPE, it has to be noted that previous studies were able to demonstrate the BFLPE even when using uncorrected grades (Marsh et al., 2007; Vogl et al., 2018).

General ASC. Another possible reason for failing to replicate the BFLPE might be the investigation of general ASC, instead of domain-specific ASCs. Although the BFLPE has been demonstrated to affect general ASC before (e.g., Fang et al., 2018; Hoferichter et al., 2018; Marsh et al., 1988), this effect has been significantly weaker than those concerning verbal and math ASCs (Fang et al., 2018). Indeed, it has been argued that due to the weak correlations between verbal and math ASCs, it might be inappropriate to aggregate verbal and math ASCs into a general ASC (Marsh et al., 1988).

ICC. Yet another aspect that might hamper the demonstration of the BFLPE is the small ICC of .04. The less variance can be attributed to differences at the class-level, the less likely it is to find statistically significant evidence that a variable of interest predicts this variance. As previous studies that reported similar ICCs did manage to demonstrate the BFLPE (Hoferichter et al., 2018; Vogl et al., 2018), it can be assumed that the small ICC cannot be the sole explanation for failing to replicate the BFLPE.

One final explanation for the failure of replicating the BFLPE is, that it simply does not exist in the present sample, but as the existence of BFLPE has been demonstrated very convincingly (e.g., Seaton et al., 2009) and it has been found in Austria before (e.g., Marsh & Hau, 2003; Seaton et al., 2009), this explanation seem to be the least likely. It is more likely that a combination of the above discussed points is the reason that the BFLPE could not be demonstrated in the present sample.

Even though the BFLPE could not be demonstrated, *Research questions 2* and *3* were still investigated in further analyses. Both subsequent multilevel multiple group analyses yielded no significant BFLPE for either group. Since both *Hypotheses 2* and *3* proposed changes in the size of the BFLPE, the failed replication of the BFLPE entailed that neither hypothesis could be indisputably confirmed. Nevertheless, it was still of interest to test the invariance of parameter estimates between students with and without immigrant background, as a significant result could have indicated whether the hypotheses would hold true if the BFLPE was present. The chi-square difference test for Model 1 (*Research question 2* – differences in the BFLPE between students with and without immigrant background) was non-significant, while it was significant for Model 2 (*Research question 3* - differences in the BFLPE depending on the proportion of students of immigrant background in the classroom),

indicating that significant differences between students with and without immigrant background only existed in Model 2.

Concerning *Hypothesis 2*, this means that the moderating effect of immigrant background that was found by Marsh (2016) could not be replicated. One explanation for this could be that Marsh used data of PISA tests, resulting in a total sample size of $N = 276{,}165$ students. In the light of this large sample size, the regression coefficient of the interaction between immigrant background and the BFLPE was very small (.012), leading Marsh to conclude that the moderating effect was substantively trivial. Considering the comparably small sample size of the present study, it is likely that there was too little statistical power to identify such a small effect. It must be noted, though, that both in Marsh's and in the present study, students of immigrant background were treated as a homogeneous group, which does not correctly reflect reality (e.g., Bacher, 2010), for example, because stereotypes differ between different immigrant groups (Appel et al., 2015; Kotzur et al., 2019; Lee & Fiske, 2006). This will likely result in different experiences for students of different immigrant backgrounds, which can manifest in a varying susceptibility to the BFLPE. Ignoring these variations might be the reason for not finding significant differences in the overall BFLPE between students with and without immigrant background.

Regarding *Research question 3*, the significance of the chi-square difference test means that a potential interaction between immigrant proportion and class-average achievement would have affected students with and without immigrant background differently. Since this interaction is not significant, the ethnic composition of the classroom, in terms of the proportion of students of immigrant background, seems to be no significant moderator of the BFLPE. Potential reasons for this finding are:

Homogenization of students of immigrant background. The neglection of the heterogeneity of students of immigrant background, that has been proposed as possible explanation for why the BFLPE did not differ between groups, might also explain why immigrant proportion is no significant moderator of the BFLPE. The variance in stereotype valence and content (Appel et al., 2015; Kotzur et al., 2019; Lee & Fiske, 2006) might lead to positive stereotypes being salient for some students of immigrant background and negative stereotypes being salient for others. Accumulated, these differences might lead to an overall non-significant effect of immigrant proportion as a moderator. Similarly, the stereotype lift effect will probably vary for native students depending on whether stereotypes about their classmates are negative or positive, confounding the overall effect of immigrant proportion.

Moderators of the stereotype lift effect. Previous studies have identified individual characteristics such as political orientation (Chatard, Selimbegović, & Konan, 2008) or prejudice and stereotype endorsement (Chatard, Selimbegović, Konan, & Mugny, 2008) as moderators of the stereotype lift effect. These characteristics have not been controlled for in the present study, which possibly lead to the non-significance of the overall stereotype lift effect (i.e., the interaction of immigrant proportion and class-average achievement), as it is not accounted for that not all native students are affected in the same way.

Operationalization and salience of stereotypes. Another possible explanation for the non-significance of the moderating effect is that the proportion of students of immigrant background alone is not enough to make negative stereotypes sufficiently salient. Following the theoretical considerations that stereotyped people are especially sensitive to cues that indicate these stereotypes (Spencer et al., 2015; Walton & Cohen, 2003), it was assumed that the mere presence of students of immigrant background would suffice to induce stereotype threat or lift effects. As there was no form of direct measurement of the presence of stereotypes in the present study, it can neither be confirmed nor denied if stereotypes were indeed salient enough to affect the students in their responses.

Finally, the reason for the non-significance of both investigated moderators of the BFLPE might simply be that they indeed do not play a role. Considering the impressive evidence about the generalizability of the BFLPE (e.g., Fang et al., 2018; Seaton et al., 2009) and the continuous demonstration that moderating effects, if significant at all, are usually very small (e.g., Marsh, 2016; Seaton et al., 2010), this explanation seems likely.

The main effect of immigrant proportion on ASC was significant for students without immigrant background only, indicating that native students reported a significantly lower ASC, the more students of immigrant background were in the classroom. This is an interesting result, as it implies direct negative consequences of the increased heterogeneity in Austria's classrooms, which is in contrast with the supposed benefits of classroom heterogeneity (e.g., Braunsteiner et al., 2018; Vock & Gronostaj, 2017). One possible explanation might be a negative relationship between the proportion of immigrant students and academic achievement that has been found in previous studies (e.g., Peetsma et al., 2006; Walter & Stanat, 2008). Peetsma and her colleagues (2006), for example, found that the proportion of students of immigrant background in the classroom negatively affected learning gains in mathematics, whereby this was partly compensated for students of immigrant background. While reanalysing PISA data, Walter and Stanat (2008) found that the proportion of immigrant students as a whole did not have a negative effect on the reading performance of students, but the propor-

tion of students with Turkish background had a significant negative effect on the reading performance. A significant, albeit very small, negative correlation between immigrant proportion and academic achievement has been found in the present study (r=-.066, see Table 1). As the relationship between ASC and academic achievement is reciprocal (REM, Marsh, 1990), potentially lower levels of achievement in classrooms with higher proportions of immigrant students might explain the lower ASCs. As the relationship between academic achievement and ASC is weaker for students of immigrant background (Faber et al., 2011; Nouwen & Clycq, 2019; Roebers et al., 1998; Shajek et al., 2006), this would also explain why only students without immigrant background are negatively affected by the proportion of immigrant students in the classroom.

A finding that is not the focus of the present investigation, but that is deemed interesting nonetheless, is the predictive value of age for ASC. When immigrant background was included in the multilevel multiple regression analysis, age lost its significance as a positive predictor of ASC. A possible explanation for this might be that age does not predict ASC for students with and without immigrant background in the same way. Indeed, as could be demonstrated in Model 2, age was only a significant predictor of ASC for native students. There have been previous studies that found that the relationship between ASC and other variables differs between students with and without immigrant background, for example, students of immigrant background reporting a weaker correlation between ASC and individual achievement than their native classmates (Faber et al., 2011; Nouwen & Clycq, 2019; Roebers et al., 1998; Shajek et al., 2006). The non-significant relationship between age and ASC for students of immigrant background in the present study might be another indicator that the formation of ASC differs substantially between students with and without immigrant background.

4.2 Limitations

Concerning the methodology, there are some potential limitations that have already been discussed in the previous section: the investigation of a general ASC compared to a domain-specific ASC, as well as the lack of a correction for the grading-on-a-curve affect, both of which might have led to an underestimation of the BFLPE. Furthermore, the operationalization of the salience of stereotypes through the proportion of students of immigrant background in the classroom might have been insufficient. Thus, the lack of explicit measure of stereotypes probably limits the present study's potential to find differences between students with and without immigrant background that can be attributed to these stereotypes. Addition-

ally, a level-1 group variable has been used in the multilevel multiple group analyses, which probably lead to an underestimation of standard errors that may provoke the incorrect rejection of the hypothesis (Hox, 2010). As none of the relationships of interest were significant in the present study, i.e., the null hypotheses were not rejected in the first place, this limitation has no impact.

Additionally, language barriers might constitute a limitation. This impression resulted from several instances, where students (presumably of immigrant background) asked for the meaning of specific words or questions used on the questionnaire. As there is no exhaustive documentation of these instances, the students that inquired and the words and questions that were concerned, no statistical analyses were possible, which is why it can neither be confirmed nor denied that language barriers were indeed a limitation.

Further limitations can be identified in the sample of the present study. Foremost, students of immigrant background were investigated as a whole, which does not adequately represent reality, as students of immigrant background are not a homogeneous group. On the contrary, not only will each individual be influenced by their cultural background in a unique way, previous research has also demonstrated that different groups of immigrants are perceived very differently by members of the host country (Appel et al., 2015; Kotzur et al., 2019; Lee and Fiske, 2006). A more differentiated consideration of students of immigrant background was not possible, due to the relatively small sample size of each immigrant group (see Table B in Appendix A), which made a simple juxtaposition of students of immigrant background and students without immigrant background necessary. A final limitation is the fact that a convenience sample was used, meaning that it cannot be excluded that there might be biases that skew the findings of the present study.

4.3 Conclusion and Directions for Future Research

Students of immigrant background have been found to report significantly higher ASC than their native classmates. This finding contributes to ASC research, as it is the first time that differences in *general* ASC were investigated regarding immigrant background, not minority (e.g., Black, Latinx) students (see Seo et al., 2019). Furthermore, the investigation of immigrant background and immigrant proportion as moderators of the BFLPE contribute to the BFLPE research, as these moderators have not been considered as potential moderators before. The non-significant results suggest that these variables do not influence the size of the BFLPE, which further demonstrates the generalizability of the BFLPE.

It is possible that the non-significant findings were caused by methodological inadequacies. Therefore, it might be of interest to replicate the present study with an improved study design. The following suggestions can be made:

Usage of domain-specific measures of ASC. General ASC has some unique variance that might be explained by social aspects of the school such as peer relations (Marsh et al., 1988). As it was intended to investigate the influence of the ethnic composition of the classroom on the BFLPE, general ASC has been chosen as measure for ASC. It has been argued, however, that a general measure of ASC might be inappropriate, as it does not account for the weak correlations between domain-specific (e.g., verbal and math) ASCs (Marsh et al., 1988). Future researchers should consider investigating the present research questions using domain-specific ASCs.

Correction for the grading-on-a-curve effect. Uncorrected grades were used to assess individual and class-average achievement, as no other achievement indicators were available. As the grading-on-a-curve effect results in a low variation of grades between classes, which consequently leads to the possibility of the BFLPE being underestimated (Marsh et al., 2007; Vogl et al., 2018), it is suggested that objective achievement measures are obtained in future research, which can be used to correct grades for the grading-on-a-curve effect.

Manipulation of stereotype salience. Stereotypes about students of immigrant background were assumed to be the more salient, the higher the proportion of native students in a classroom was. This might be insufficient to test whether stereotypes influence the size of the BFLPE for students with and without immigrant background. A more suitable approach might be to manipulate stereotype salience to ensure that they are on the students' minds when completing measures that assess their ASC.

Differentiated investigation of immigrant background. Students of immigrant background were treated as a homogeneous group, which does not adequately reflect reality (Bacher, 2010), as stereotype content and valence differ between different immigrant groups (Appel et al., 2015; Kotzur et al., 2019; Lee & Fiske, 2006). The influence of stereotypes will likely vary with the particular immigrant background a person has, which means that investigating students of immigrant background as a whole might yields a confounded result. In future research, focusing on one or a few distinct groups of students of immigrant background (e.g., students with Turkish background, who constitute the largest group in the present sample) might be preferable.

Finally, the non-significant relationship between age and ASC for students of immigrant background might be an indicator that the formation of ASC differs substantially be-

tween students with and without immigrant background, a possibility that might be interesting to investigate in future research.

5 References

Alesina, A., Carlana, M., La Ferrara, E., & Pinotti, P. (2018). Revealing stereotypes: Evidence from immigrants in schools. *NBER Working Paper Series*, Article 25333. https://doi.org/10.3386/w25333

Alesina, A., Miano, A., & Stantcheva, S. (2018). Immigration and redistribution. *NBER Working Paper Series*, Article 24733. https://doi.org/10.3386/w24733

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers In Psychology, 6*, Article 900. https://doi.org/10.3389/fpsyg.2015.00900

Aronson, E., Wilson, T., Akert, R., & Sommers, S. (2017). *Social psychology* (9th ed.). Pearson.

Bacher, J. (2010). Bildungschancen von Kindern mit Migrationshintergrund: Ist-Situation, Ursachen und Maßnahmen [Educational opportunities of children with immigrant background: current situation, causes, and interventions]. *WISO, 33*(1), 29–48.

Becker, M., & Neumann, M. (2016). Context-related changes in academic self-concept development: On the long-term persistence of big-fish-little-pond effects. *Learning and Instruction, 45*, 31-39. https://doi.org/10.1016/j.learninstruc.2016.06.003

Becker, M., & Neumann, M. (2018). Longitudinal big-fish-little-pond effects on academic self-concept development during the transition from elementary to secondary schooling. *Journal of Educational Psychology, 110*(6), 882-897. https://doi.org/10.1037/edu0000233

Bong, M., & Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1-40. https://doi.org/10.1023/A:1021302408382

Braunsteiner, M.-L., Fischer, C., Kernbichler, G., Prengel, A., & Wohlhart, D. (2018). Erfolgreich lernen in Klassen mit hoher Heterogenität [Successful learning in classes with high heterogeneity]. In S. Breit, F. Eder, K. Krainer, C. Schreiner, A. Seel, & C. Spiel (Eds.), *Nationaler Bildungsbericht Österreich 2018, Band 2* (pp. 19-62). Leykam. https://doi.org/10.17888/nbb2018-2-8

Chatard, A., Selimbegović, L., & Konan, P. (2008). Leftists' and rightists' IQ as a function of stereotype salience. *Journal of Research in Personality, 42*(6), 1602-1606. https://doi.org/10.1016/j.jrp.2008.05.004

Chatard, A., Selimbegović, L., Konan, P., & Mugny, G. (2008). Performance boosts in the classroom: Stereotype endorsement and prejudice moderate stereotype lift. *Journal of Experimental Social Psychology, 44*(5), 1421-1424. https://doi.org/10.1016/j.jesp.2008.05.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Dai, D., & Rinn, Y. (2008). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review, 20*(3), 283-317. https://doi.org/10.1007/s10648-008-9071-x

Dawson, J. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology, 29(*1), 1-19. https://doi.org/10.1007/s10869-013-9308-7

Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept [The academic self-concept scales: Construction and evaluation of a new instrument]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 23*(4), 393-405. https://doi.org/10.1024//0170-1789.23.4.393

Diel, C., & Fick, P. (2016). Ethnische Diskriminierung im deutschen Bildungssystem [Ethnic discrimination in the German educational system]. In C. Diehl, C. Hunker, & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten* (pp. 243-286). Springer VS. https://doi.org/10.1007/978-3-658-04322-3

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78-89. https://doi.org/10.1080/00461520902832368

Ehm, J., Duzy, D., & Hasselhorn, M. (2011). Das akademische Selbstkonzept bei Schulanfängern: Spielen Geschlecht und Migrationshintergrund eine Rolle? [The academic self-concept of first graders: Do gender and immigration background play a role?]. *Frühe Bildung*, 37-45. https://doi.org/10.1026/2191-9186/a000008

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138. https://doi.org/10.1037/1082-989X.12.2.121

Faber, G. (2013). Klassenzusammensetzung als Kontextfaktor für die Leistungs- und Selbstkonzeptentwicklung [Classroom composition as context factor for achievement and self-concept development]. In R. Becker & A. Schulze (Eds.), *Bildungskontexte: Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (pp. 325-351). Springer Fachmedien. https://doi.org/10.1007/978-3-531-18985-7_12

Faber, G., & Billmann-Mahecha, E. (2012). Schulfachliche Leistungen und Selbsteinschätzungen in Abhängigkeit von Familiensprache und Geschlecht: Befunde aus einer Studie in dritten und vierten Grundschulklassen [Achievement and self-assessment in school subjects depending on family language and gender: Findigs of a study of third and fourth elementary school grades]. In F. Hellmich, S. Förster, & F. Hoya (Eds.), *Bedingungen des Lehrens und Lernens in der Grundschule. Bilanz und Perspektiven. Jahrbuch Grundschulforschung, Bd. 16* (pp. 271–274). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-19137-9_52

Faber, G., Tiedemann, J., & Billmann-Mahecha, E. (2011). Selbstkonzept und Lernfreude in der Grundschulmathematik: Die Bedeutung von Migration und Geschlecht. Längsschnittliche Ergebnisse aus der Hannoverschen Grundschulstudie [The impact of migration background and gender on elementary students' achievement and self-beliefs in the mathematics domain]. *Heilpädagogische Forschung 37*, 127–143.

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology, 9*. Article 1569. https://doi.org/10.3389/fpsyg.2018.01569

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878-902. https://doi.org/10.1037/0022-3514.82.6.878

Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18. https://doi.org/10.1037/a0024338

Froehlich, L. Y., Mok, S. E., Martiny, S., & Deaux, K. (2018). Stereotype threat-effects for Turkish-origin migrants in Germany: Taking stock of cumulative research evidence. *European Educational Research Journal*. https://doi.org/10.1177/1474904118807539

Froehlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *PLoS ONE, 14*(9), Article E0223103. https://doi.org/10.1371/journal.pone.0223103

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity, 3*(1), 53-68. https://doi.org/10.1080/13576500342000040

Guo, J., Marsh, H., Morin, A., Parker, P., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. *American Educational Research Journal, 52*(2), 371-402. https://doi.org/10.3102/0002831214565786

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzepts vom Kindergarten bis zur 6. Klassenstufe [From optimism to realism? On the development of the ability-related self-concept from preschool to 6th grade]. In W. Schneider, & F. E. Weinert (Eds.), *Entwicklung im Kindesalter* (pp. 116 – 132). Beltz.

Hoferichter, F., Latsch, A., Lazarides, R., & Raufelder, D. (2018). The big-fish-little-pond effect on the four facets of academic self-concept. *Frontiers in Psychology, 9*, Article 1247. https://doi.org/10.3389/fpsyg.2018.01247

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed., Quantitative methodology series). Routledge.

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

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505-528. https://doi.org/10.1016/j.jsp.2011.07.001

Konan, P., Chatard, N., Selimbegović, A., Mugny, L., & Moraru, G. (2011). Deflecting stereotype threat through downward comparison: When comparison with immigrants boosts the performance of stigmatized native students. *Social Justice Research, 24*(2), 191-205. https://doi.org/10.1007/s11211-011-0134-7

Kotzur, P., Friehs, M., Asbrock, F., & Zalk, M. (2019). Stereotype content of refugee subgroups in Germany. *European Journal of Social Psychology, 49*(7), 1344-1358. https://doi.org/10.1002/ejsp.2585

Lee, T., & Fiske, S. (2006). Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations, 30*(6), 751-768. https://doi.org/10.1016/j.ijintrel.2006.06.005

Lehman, B. (2012). The impacts of friendship groups' racial composition when perceptions of prejudice threaten students' academic self-concept. *Social Psychology of Education, 15*(3), 411-425. https://doi.org/10.1007/s11218-012-9190-6

Lent, R. W., Brown, S. D., & Gore, P. A. (1997). Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy. *Journal of Counseling Psychology, 44*(3), 307-315. https://doi.org/10.1037/0022-0167.44.3.307

Marsh, H. (1984). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal, 21*(4), 799-806. https://doi.org/10.3102/00028312021004799

Marsh, H. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23(*1), 129-149. https://doi.org/10.3102/00028312023001129

Marsh, H. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280-295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*(4), 646-656. https://doi.org/10.1037/0022-0663.82.4.646

Marsh, H. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal, 28*(2), 445-480.

Marsh, H. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*(2), 256-273. https://doi.org/10.1037/edu0000059

Marsh, H., Byrne, B., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*(3), 366-380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133-163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion Study. *Child Development, 69*(4), 1030-1053. https://doi.org/10.1111/j.1467-8624.1998.tb06159.x

Marsh, H., & 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. *The American Psychologist, 58*(5), 364-376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H., & O'Mara, A. J. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*(4), 542-552. https://doi.org/10.1177/0146167207312313

Marsh, H., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie, 24*(1), 51-72. https://doi.org/10.1024/1010-0652/a000004

Marsh, H., & Parker, J. (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. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H., Pekrun, R., Murayama, K., Arens, A., Parker, P., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology, 54*(2), 263-280. https://doi.org/10.1037/dev0000393

Marsh, H., Pekrun, R., Parker, P., Murayama, K., Guo, J., Dicke, T., & Arens, A. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331-353. https://doi.org./10.1037/edu0000281

Marsh, H., Seaton, W., Trautwein, M., Lüdtke, U., Hau, O., O'Mara, K., & Craven, T. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*(3), 319-350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20,* 107-125. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397-416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after

graduation. *American Educational Research Journal, 44*(3), 631-669. https://doi.org/10.3102/0002831207306728

Marsh, H., & Yeung, A. (1997). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal, 34*(4), 691-720. https://doi.org/10.3102/00028312034004691

Mead, G. H. (1934). *Mind, self, and society from the standpoint of a social behaviorist.* Chicago Univ. Press.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129-1167. https://doi.org/10.3102/0034654309337522

Mplus. (2020). *Chi-square difference testing using the Satorra-Bentler scaled chi-square.* https://www.statmodel.com/chidiff.shtml

Muthén, B.O., & Muthén, L.K. (1998–2019). *Mplus (Version 8.4)*.

Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus user's guide* (8th ed.).

Nagengast, B., & Marsh, H. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033-1053. https://doi.org/10.1037/a0027697

Nouwen, W., & Clycq, N. (2019). The role of teacher–pupil relations in stereotype threat effects in flemish secondary education. *Urban Education*, *54*(10), 1551–1580. https://doi.org/10.1177/0042085916646627

Oberwimmer, K., Baumegger, D., & Vogtenhuber, S. (2018). Indikatoren A: Kontext des Schul- und Bildungswesens [Indicators A: Context of the school and education system]. In S. Breit, F. Eder, K. Krainer, C. Schreiner, A. Seel, & C. Spiel (Eds.), *Nationaler Bildungsbericht Österreich 2018, Band 1* (pp. 25-48). Leykam. https://doi.org/10.17888/nbb2018-2-8

Organisation for Economic Co-operation Development (OECD). (2010). *Closing the Gap for Immigrant Students: Policies, Practice and Performance*. (OECD Reviews of Migrant Education). https://doi.org/10.1787/9789264075788-en

Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class composition influences on pupils' cognitive development. S*chool Effectiveness and School Improvement, 17*(3), 275-302. https://doi.org/10.1080/13803610500480114

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Roebers, C., Mecheril, A. & Schneider, W. (1998). Migrantenkinder in deutschen Schulen [Migrant children in German schools]. *Zeitschrift für Pädagogik, 44*(5), 723 – 736.

Ryu, E. (2015). Multiple group analysis in multilevel structural equation model across Level 1 groups. *Multivariate Behavioral Research, 50*(3), 300-315. https://doi.org/10.1080/00273171.2014.1003769

Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish–little-pond effect across cultures. *Journal of Educational Psychology, 108*(3), 405-423. https://doi.org/10.1037/edu0000063

Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO)* [School-related self-concept scales]. Hogrefe.

Schwabe, F., Korthals, R., & Schils, T. (2019). Positive social relationships with peers and teachers as moderators of the big-fish-little-pond effect. *Learning and Individual Differences, 70*, 21-29. https://doi.org/10.1016/j.lindif.2018.12.006

Seaton, M., Marsh, H., & Craven, R. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101*(2), 403-419. https://doi.org/10.1037/a0013838

Seaton, M., Marsh, H., & Craven, R. (2010). Big-fish-little-pond effect: Generalizability and moderation—two sides of the same coin. *American Educational Research Journal, 47*(2), 390-433. https://doi.org/10.3102/0002831209350493

Seo, E., Shen, Y., & Benner, A. (2019). The paradox of positive self-concept and low achievement among Black and Latinx youth: A test of psychological explanations. *Contemporary Educational Psychology, 59*. Article 101796. https://doi.org/10.1016/j.cedpsych.2019.101796

Shajek, A., Lüdtke, O., Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund [Immigrant students‘ academic self-concepts]. *Unterrichtswissenschaft 34*(2), 125-145.

Shavelson, R., Hubner, J., & Stanton, G. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407-441. https://doi.org/10.3102/00346543046003407

Social Science Statistics (2020). *P value from chi-square calculator.* https://www.socscistatistics.com/pvalues/chidistribution.aspx

Spencer, S., Logel, C., & Davies, P. (2016). Stereotype threat. *Annual Review of Psychology, 67*(1), 415-437. https://doi.org/10.1146/annurev-psych-073115-103235

Statistik Austria. (2019). *Migration und Integration. Zahlen. Daten. Indikatoren 2019* [Migration and integration. Numbers. Data. Indicators 2019].

Statistik Austria (2020a). *Population with foreign background 2019.* http://www.statistik.at/web_en/statistics/PeopleSociety/population/population_change_by_demographic_characteristics/population_with_foreign_background/036095.html

Statistik Austria. (2020b). *International migration to and from Austria by citizenship 1961-2019.* http://www.statistik.at/web_en/statistics/PeopleSociety/population/migration/028953.html

Steele, C. (1997). A threat in the air. *American Psychologist, 52*(6), 613-629. https://doi.org/10.1037/0003-066X.52.6.613

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). National Foundation for Educational Research.

United Nations Economic Commission for Europe (UNECE). (2015). *Conference of European Statisticians Recommendations for the 2020 censuses of population and housing.* https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf

Valentine, J.C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review, *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

Vock, M., & Gronostaj, A. (2017). *Umgang mit Heterogenität in Schule und Unterricht* [Handling of heterogeneity in school and teaching]. Friedrich-Ebert-Stiftung.

Vogl, K., Schmidt, I., & Preckel, F. (2018). The role of academic ability indicators in big-fish-little-pond effect research: A comparison study. *The Journal of Educational Research, 111*(4), 429-438. https://doi.org/10.1080/00220671.2017.1291485

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

Walter, O., & Stanat, P. (2008). Der Zusammenhang des Migrantenanteils in Schulen mit der Lesekompetenz: Differenzierte Analysen der erweiterten Migrantenstichprobe von PI-

SA 2003 [The relationship between the migrant proportion in schools and reading: differentiated analyses of the extended migrant sample of PISA 2003]. *Zeitschrift Für Erziehungswissenschaft, 11*(1), 84-105. https://doi.org/0.1007/s11618-008-0005-7

Walton, G., & Cohen, G. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*(5), 456-467. https://doi.org/10.1016/S0022-1031(03)00019-2

Weiber R., Mühlhaus D. (2014). Mehrgruppen-Kausalanalyse (MGKA) [Multiple-group causal analysis]. In R. Weiber & D. Mühlhaus (Eds.), *Strukturgleichungsmodellierung* (2nd ed., pp. 285-322). Springer-Gabler. https://doi.org/10.1007/978-3-642-35012-2_14

6 Appendices

A. Complementary Information about the Sample

Table A. Sample sizes at the individual-, classroom-, and school-level.

School	Number of classrooms	Students/Classroom (range)
School 1	10	11- 24
School 2	12	7- 22
School 3	9	10 - 25
School 4	5	13 - 19
School 5	14	9 - 25
School 6	5	14 - 23
School 7	5	12 - 20
School 8	4	14 - 19
School 9	18	16 - 28
School 10	5	14 - 17
M (SD)	8.70 (4.76)	18.95 (4.04)

Note. $N_{schools}$ = 10; $N_{classrooms}$ = 87; $N_{students}$ = 1,649.

Table B. Descriptive statistics for each immigrant background.

Immigrant Background (A - K)	*N*	%	Immigrant Background (L - Z)	*N*	%
Afghanistan	30	3.64	Lebanon	2	0.24
Albania	8	0.97	Libya	3	0.36
Algeria	1	0.12	Moldavia	1	0.12
Argentina	1	0.12	Mongolia	3	0.36
Armenia	1	0.12	Nepal	1	0.12
Azerbaijan	1	0.12	Netherlands	5	0.61

(continued)

Table B. Descriptive statistics for each immigrant background (continued).

Immigrant Background (A - K)	*N*	%	Immigrant Background (L - Z)	*N*	%
Bangladesh	2	0.24	Nigeria	11	1.33
Belarus	3	0.36	North Macedonia	13	1.58
Belgium	2	0.24	Pakistan	7	0.85
Bosnia	48	5.83	Palestine	2	0.24
Brasilia	1	0.12	Philippines	12	1.46
Bulgaria	7	0.85	Poland	17	2.06
Chechen Republic	15	1.82	Romania	51	6.19
Chile	2	0.24	Russia	17	2.06
China	2	0.24	Saudi Arabia	2	0.24
Colombia	2	0.24	Senegal	1	0.12
Congo	1	0.12	Serbia	78	9.47
Croatia	14	1.70	Slovakia	57	6.92
Cuba	1	0.12	Slovenia	1	0.12
Czech Republic	10	1.21	Somalia	5	0.61
Dominican Republic	2	0.24	South Africa	1	0.12
Dubai	1	0.12	Spain	9	1.09
Egypt	26	3.16	Sudan	2	0.24
Ethiopia	1	0.12	Sweden	1	0.12
France	1	0.12	Switzerland	1	0.12
Georgia	1	0.12	Syria	30	3.64
Germany	32	3.88	Tajikistan	1	0.12
Ghana	5	0.61	Thailand	2	0.24
Greece	1	0.12	Tunisia	8	0.97

(continued)

Table B. Descriptive statistics for each immigrant background (continued).

Immigrant Background (A - K)	*N*	%	Immigrant Background (L - Z)	*N*	%
Hungary	32	3.88	Turkey	120	14.56
India	14	1.70	Ukraine	2	0.24
Iran	8	0.97	United Kingdom	7	0.85
Iraq	15	1.82	Venezuela	1	0.12
Italy	3	0.36	Vietnam	4	0.49
Japan	1	0.12	Yemen	1	0.12
Kazakhstan	1	0.12	Other	9	1.09
Kosovo	37	4.49			
Kuwait	2	0.24			
Kyrgyzstan	1	0.12			

Note. Other = student answers that cannot be allocated to a specific country (e.g. "Africa"); *N* of students of immigrant background = 698, N_{total} = 1,649; students can represent up to three immigrant backgrounds (their own and their parents'countries of birth).

Table C. Comparison of students with and without immigrant background.

	Native	Immigrant background
ASC	3.63 (0.76)	3.56 (0.74)
Individual achievement	3.41 (0.90)	3.03 (0.81)
Age	12.25 (0.99)	12.62 (1.13)
% female	44.9	53.8
% NMS	42.4	80.5
% AHS	57.6	19.5

Note. ASC = Academic self-concept; values are means, values in parentheses are standard deviations.

Table D. Descriptive statistics of missing data.

	N	%
ASC	7	0.4
Individual achievement	27	1.6
Immigrant background	14	0.8
Age	3	0.2
Sex	27	1.6
Class-average achievement	0	0.0
Immigrant proportion	0	0.0

Note. ASC = Academic self-concept; class-average achievement and immigrant proportion are aggregated level-2 variables and cannot have missing values.

B. Mplus Syntaxes for the models of the present investigation

```
TITLE: Research Question 1: Difference in ASC between
      students with and without immigrant background
      manifest two-level regression;
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv age sex im_stat;
      within = ind_achv age sex im_stat;
cluster = class;
missing = all (-99);
DEFINE: center ind_achv age sex im_stat(GROUPMEAN);
ANALYSIS:   type = twolevel;
MODEL:
%within%
      asc on ind_achv age sex im_stat;
OUTPUT; sampstat stand tech1;
```

```
TITLE: Model 0: Replication of the BFLPE
      doubly-manifest multilevel model
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv av_achv age sex;
within = ind_achv age sex;
between = av_achv;
cluster = class;
missing = all (-99);
DEFINE: center ind_achv av_achv age sex(GRANDMEAN);
ANALYSIS:   type = twolevel;
MODEL:
%within%
asc on ind_achv age sex;

%between%
asc on av_achv;

OUTPUT: sampstat stand tech1;
```

```
TITLE: Research Question 2:
      Immigrant background as moderator of the BFLPE
      multiple group multilevel model
      constrained Model 1: parameters are fixed
      doubly-manifest modelling
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv av_achv im_stat age sex;
classes = c(2);
knownclass = c (im_stat = 0 1);
within = ind_achv age sex;
between = av_achv;
cluster = class;
missing = all (-99);
DEFINE: center ind_achv age av_achv sex(GRANDMEAN);
ANALYSIS: type = twolevel mixture;
MODEL:
%within%
%overall%
asc on ind_achv age sex;

%between%
%overall%
asc on av_achv;

OUTPUT: sampstat stand tech1;
```

```
TITLE: Research Question 2:
      Immigrant background as moderator of the BFLPE
      multiple group multilevel model
      unconstrained Model 1: parameters allowed to vary
      doubly-manifest modelling
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv av_achv im_stat age sex;
classes = c(2);
knownclass = c (im_stat = 0 1);
within = ind_achv age sex;
between = av_achv;
cluster = class;
missing = all (-99);
DEFINE: center ind_achv age av_achv sex(GRANDMEAN);
ANALYSIS: type = twolevel mixture;
MODEL:
%within%
%overall%
asc on ind_achv age sex;

%within%
%c#1%
asc on ind_achv age sex;

%within%
%c#2%
asc on ind_achv age sex;

%between%
%overall%
asc on av_achv;

%between%
%c#1%
asc on av_achv;

%between%
%c#2%
asc on av_achv;

OUTPUT: sampstat stand tech1;
```

```
TITLE: Research Question 3:
      Immigrant background and immigrant propotion
      as moderators of the BFLPE
      multiple group multilevel model
      constrained Model 2: parameters are fixed
      doubly-manifest modelling
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv av_achv im_stat age sex im_prop avxprop;
classes = c(2);
knownclass = c (im_stat = 0 1);
within = ind_achv age sex;
between = av_achv im_prop avxprop;
cluster = class;
missing = all (-99);
DEFINE:
      center ind_achv age av_achv sex im_prop(GRANDMEAN);
      avxprop = av_achv * im_prop;
ANALYSIS: type = twolevel mixture;
MODEL:
%within%
%overall%
asc on ind_achv age sex;

%between%
%overall%
asc on av_achv im_prop avxprop;

OUTPUT: sampstat stand tech1;
```

```
TITLE: Research Question 3:
      Immigrant background and immigrant propotion
      as moderators of the BFLPE
      multiple group multilevel model
      unconstrained Model 2: parameters allowed to vary
      doubly-manifest modelling
DATA: file = Daten_Mplus.dat;
VARIABLE:
names =
      track class id age sex m_achv g_achv e_achv
      sessko_1 sessko_2 sessko_3 sessko_4 sessko_5
      asc ind_achv av_achv im_prop im_stat;
usevariables =
      asc ind_achv av_achv im_stat age sex im_prop avxprop;
classes = c(2);
knownclass = c (im_stat = 0 1);
within = ind_achv age sex;
between = av_achv im_prop avxprop;
cluster = class;
missing = all (-99);
DEFINE:
      center ind_achv age av_achv sex im_prop(GRANDMEAN);
      avxprop = av_achv * im_prop;
ANALYSIS: type = twolevel mixture;
MODEL:
%within%
%overall%
asc on ind_achv age sex;

%within%
%c#1%
asc on ind_achv age sex;

%within%
%c#2%
asc on ind_achv age sex;

%between%
%overall%
asc on av_achv im_prop avxprop;

%between%
%c#1%
asc on av_achv im_prop avxprop;

%between%
%c#2%
asc on av_achv im_prop avxprop;

OUTPUT: sampstat stand tech1;
```

C. Complementary Information about the Results

Table E. Values for chi-square difference testing.

	Model 1		Model 2	
	H0	H1	H0	H1
Loglikelihood	-2,636.705	-2,633.392	-2,633.951	-2,629.948
Number of free parameters	9	13	11	17
Scaling correction factor	1.664	1.454	1.519	1.308
cd	0.983		0.921	
χ^2	6.740		8.691	
DF	4		6	

Note. H0 = constrained model; H1 = unconstrained model; cd = difference scaling correction; DF = degrees of freedom.

Table F. Unstandardized parameter estimates for models of the BFLPE.

Parameter	Model 0	Model 1 (constrained)		Model 2 (unconstrained)	
		Native	I.B.	Native	I. B.
Intercept	3.604 (0.018)***	3.568 (0.028)***	3.656 (0.027)***	3.540 (0.33)**	3.660 (0.034)**
Level-1 (individual)					
Individual achievement	.443 (0.023)***	.447 (0.023)***	.447 (0.023)***	.464 (0.026)***	.412 (0.039)***
Age	.041 (0.019)*	.042 (0.020)*	.042 (0.020)*	.074 (0.023)***	.012 (0.028)
Sex	.193 (0.032)***	.194 (0.032)***	.194 (0.032)***	.231 (0.043)***	.142 (0.046) **
Level-2 (classroom)					
Class-average achievement	-.014 (0.06)	.031 (0.062)	.031 (0.062)	-.015 (0.090)	-.112 (0.108)
Immigrant proportion				-.329 (0.137)*	-.140 (0.124)
Interaction				-.052 (0.328)	.094 (0.315)

Note. Standard errors in parentheses; I.B.= immigrant background; Interaction = interaction of class-average achievement and immigrant proportion; a chi-square difference test yielded no significant differences between the parameter estimates of native and students of immigrant background for Model 1; *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.