# Which class matters? Juxtaposing multiple class environments as frames-of-reference for academic self-concept formation

**Fleischmann, Moritz, Hübner, Nicolas, Marsh, Herbert W., Guo, Jiesi, Trautwein, Ulrich and Nagengast, Benjamin**

Fleischmann, M., Hübner, N., Marsh, H. W., Guo, J., Trautwein, U., & Nagengast, B. (2022). Which Class Matters? Juxtaposing Multiple Class Environments as Frames-of-Reference for Academic Self-Concept Formation. Journal of Educational Psychology, 114(1), 127–143. https://doi.org/10.1037/edu0000491

**Which Class Matters? Juxtaposing Multiple Class Environments as Frames-of-Reference for Academic Self-Concept Formation**

Equally able students have lower academic self-concept in high achieving schools or classes, a phenomenon known as the big fish little pond effect (BFLPE). The class (more so than the school) has been shown to be the pivotal frame-of-reference for academic self-concept formation—a local dominance effect. However, many school systems worldwide employ forms of course-by-course tracking, thus exposing students to multiple class environments. Due to the high correlation between multiple student environments, the frame-of-reference used for academic self-concept formation in course-by-course tracked systems is unclear to date. We addressed this unresolved issue by using data from a comprehensive survey that measured the entire population of Austrian eighth-grade students in the domain of mathematics in 2012. General secondary school students were in the core subjects (i.e., mathematics, German, and English) grouped according to ability, whereas regular class composition was the same in all other subjects. Using cross-classified multilevel models, we regressed math self-concept on average math achievement of students' school, math class, and regular class. Consistent with the local dominance effect, we found the BFLPE on the school level to be weak after controlling for the class levels. We found a stronger BFLPE on the regular class level and the strongest BFLPE on the math class level. Our study demonstrates the importance of multiple class environments as frames-of-reference for academic self-concept formation.

## Educational Impact and Implications Statement

A positive perception of one's academic abilities—termed academic self-concept—is a desirable educational goal. As a consequence of social comparison processes, academic self-concept is assumed to be negatively affected when students are placed in classrooms composed of high achieving classmates. We investigated these comparison processes in course-by-course tracked systems in which students are members of several class environments. Our findings suggest that students build their academic self-concept in a certain domain, for instance in math, to a major extent in comparisons with domain-specific class environments (e.g., the math class) and to a minor extent with domain-unrelated class environments. Our study contributes to the psychological understanding of academic self-concept formation and refines BFLPE implications in course-by-course tracked systems: Students' academic self-concept in a certain domain may be hurt when placing them in high achieving domain-specific classes but it will to a lesser extent also be hurt when placing them in high achieving domain-unrelated classrooms.

**Which Class Matters? Juxtaposing Multiple Class Environments as Frames-of-Reference for Academic Self-Concept Formation**

Academic self-concept—that is, students' perceptions of their academic abilities (Marsh et al., 2016)—is predicted by the average academic achievement of educational environments (i.e., school or class) when controlling for individual achievement differences (Marsh, 1987; Marsh & Parker, 1984). In particular, equally able students have lower academic self-concept in high achieving schools or classes. This frame-of-reference effect—which has been labeled big fish little pond effect (BFLPE; for an overview, see Marsh & Seaton, 2015)—is assumed to be induced by social comparison processes in which students compare their academic achievement with that of their schoolmates or classmates (Huguet et al., 2009; Marsh, Kuyper, et al., 2014).

Typically, research on the BFLPE regresses academic self-concept on aggregated achievement of either the school or the class level. However, average achievement of educational environments is highly correlated, making it difficult to identify the relative strength of both frames-of-reference. To overcome this, Marsh, Kuyper, et al. (2014) employed a three-level approach and found the class to be the pivotal frame-of-reference for academic self-concept formation. In line with the local dominance effect (see Zell & Alicke, 2010), they concluded that local comparison information matters the most for ability self-evaluations.

However, many school systems around the world—for instance, those of many Anglo-Saxon countries—group students according to ability in one or more subjects (often in core subjects like math) while allowing them to remain in the same regular class for the other (untracked) ones. Such an educational practice is referred to as course-by-course tracking (Chmielewski, 2014). Students in course-by-course tracked systems are members of at least two class environments. In such a situation, the question arises to what extent academic self-concept is impacted by the average achievement of multiple class environments. Juxtaposing multiple class environments as frames-of-reference for academic self-concept formation is especially interesting as those educational environments are equal regarding their local proximity but might differ concerning their domain-specific proximity.

Previous research was not able to juxtapose multiple class environments as frames-of-reference for academic self-concept formation because educational large-scale datasets (e.g., PISA, TIMSS) typically do not include information on multiple class environments. And even if information on multiple class environments were included, the fact that typically only a subsample of students from one school is tested would not allow for calculating reliable achievement aggregates on all levels of student nesting. In the present study, we were able to overcome this limitation by making use of data coming from the Austrian national educational standard assessment from 2012 (BIFIE, 2016; Schreiner & Breit, 2012), a comprehensive survey that tested the entire population of Austrian eighth-grade students without special educational needs in the domain of mathematics. Austrian general secondary school students were grouped according to ability in math, German, and English classes and attended all other subjects in the same (untracked) regular class. As the complete student population was tested and information on students' math and regular classes was available, this dataset provided us with an unprecedented opportunity for juxtaposing multiple class environments as frames-of-reference for academic self-concept formation.

**The Big Fish Little Pond Effect and Its Proposed Mechanisms**

The BFLPE, namely the finding that academic self-concept is negatively affected by school- or class-average achievement is supposed to be the result of social comparison processes. Based on classical social comparison theory (Festinger, 1957), there is a human drive for self-evaluation that results in students comparing with school- and classmates, consequently building their academic self-concept based on these comparisons. Thus, an average-ability student would develop a positive self-concept in low achieving educational environments, whereas the opposite would occur in high achieving environments. Several studies support the idea that the BFLPE is driven by social comparison processes (Huguet et al., 2009; Marsh, Kuyper, et al., 2014). Note that other self-belief constructs, such as academic self-efficacy, differ from academic self-concept in that they are prospective rather than retrospective (beliefs about what I can do in the future vs. what I have done in the past) and descriptive rather than evaluative (beliefs about how well behavior matches external standards vs. beliefs about how well behavior matches personal standards). Thus, they are not assumed to be affected by such frame-of-reference effects (e.g., Hulleman, 2016; Marsh et al., 2016). Overall, the BFLPE has received strong empirical support. First, the effect is generalizable across cultures (e.g., Marsh, Abduljabbar et al., 2014; Marsh & Hau, 2003; Nagengast & Marsh, 2012). Additionally, the BFLPE generalizes well over individual characteristics as well as characteristics of educational environments (e.g., Lüdtke et al., 2005; Seaton et al., 2010; Seaton et al., 2011). Finally, frame-of-reference effects affect other desirable outcomes, even though the effects are smaller in size, such as academic effort, interest, participation in physical education, and even long-term income (Göllner et al., 2018; Marsh, 1991; Trautwein et al., 2006; Trautwein et al., 2008)

Building on work from social psychology (e.g., Cialdini & Richardson, 1980; Snyder et al., 1986), there has been speculation that membership in a high achieving educational environment might also come with benefits in terms of academic self-concept because students "bask in the reflected glory" of successful others. To put this idea to an empirical test, the BFLPE model was extended by including a measure capturing the prestige of students' learning environments (Marsh et al., 2000; Trautwein et al., 2009). Students' perception of their educational environment's status positively affected academic self-concept, and controlling for the prestige of students' educational environment led to an even more negative BFLPE.

Membership in a high achieving learning group seems to lead to positive assimilation effects which are counterbalanced by contrastive frame-of-reference effects. Similarly, high within-school track membership (in contrast to low within-school track membership) positively predicts academic self-concept, after controlling for individual and aggregated achievement (Chmielewski et al., 2013; Trautwein et al., 2006), suggesting assimilation effects. Again, accounting for within-school track led to a more negative BFLPE. In other words, equally able students experience a much stronger academic self-concept decline in high achieving learning environments when track level is kept constant. Generally, there are two different interpretations of such assimilation effects. First, these effects might result from track-level assignment, thus being assimilative track-branding effects (Chmielewski et al., 2013). Second, track level might be an indicator of students' prior academic achievement that is not captured by the standardized achievement measure, thus positively predicting academic self-concept (Marsh et al., 2018). As track level and prior achievement are always correlated, correlational analyses cannot clarify the interpretation of track-level effects on academic self-concept.

Early on, researchers speculated that the BFLPE is driven by grading on a curve or class-referenced grading, which is the tendency of teachers to give the best grades to the best students, the worst grades to the worst students, and place the others somewhere in-between (Neumann et al., 2011). For instance, Marsh (1987) theorized that the BFLPE might be the consequence of equally able students getting worse grades in high achieving classes, subsequently leading to lower academic self-concept. This idea has been tested by considering teacher-assigned grades as an additional predictor variable in the BFLPE model. In this model, controlling for grades typically leads to a substantial decline in the size of the BFLPE (e.g., Marsh, 1987; Trautwein et al., 2006). Statistically speaking, equally able students with equal grades have only a slightly lower academic self-concept in high achieving learning environments. It is important to note that these results do not ensure a causal relationship between grades and the BFLPE. There is still the possibility that students compare to each other independent of grades.

**Juxtaposing the School and the Class as Frames-of-Reference for Academic Self-Concept Formation**

Early studies on the BFLPE (e.g., Marsh, 1987, 1991; Marsh & Parker, 1984) most typically used some measure of school-average achievement to predict academic self-concept. By contrast, recent studies more often investigated the effects of class-average achievement on academic self-concept (e.g., Marsh et al., 2001; Marsh, Abduljabbar, et al., 2014; Trautwein et al., 2006). The decision of choosing the school or the class as students' learning environment was typically guided by the properties of the data to be analyzed. The school was chosen when data with a student sample from schools were available. The class was chosen when whole classes were drawn. Because previous studies modeled either the school or the class (but not both) and because achievement aggregates from multiple student environments are typically highly correlated, these examinations were not able to investigate what the pivotal frame-of-reference for academic self-concept formation is. Specifically, a student who attends a school composed of high-achieving students typically also attends a classroom in which average achievement is high. Due to the high correlations between potential frames of reference for academic self-concept formation, contextual effects in an ordinary two-level model (e.g., students within schools) might result from a *noisy reflection* of some other frame-of-reference. For instance, a negative effect of school-average achievement on academic self-concept (when controlling for individual achievement) does not necessarily mean that students form their academic self-concept in comparison with their schoolmates.

From a social comparison literature perspective, clear expectations exist regarding the relative importance of the school and the class as frames-of-reference for academic self-concept formation. In several experiments, it was shown that local comparison information supersedes the influence of distal comparison information on ability self-evaluations (Alicke et al., 2010; Buckingham & Alicke, 2002; Zell & Alicke, 2009). Based on their experimental work, Zell and Alicke (2010) hypothesized the local dominance effect in self-evaluation stating that "when multiple comparison standards are available for self-evaluation, people rely on the most local comparison information while deemphasizing more general, and typically more diagnostic, forms of comparison feedback" (Zell & Alicke, 2010, p. 369).

Marsh, Kuyper, et al. (2014) conducted a study with 15,356 Dutch students nested in 651 classes and 95 schools and juxtaposed the school and the class as frames-of-reference for academic self-concept formation. When modeled separately, school-average achievement, as well as class-average achievement, negatively predicted academic self-concept, controlling for individual achievement. However, when juxtaposed in a joint model, class achievement negatively predicted academic self-concept, whereas school achievement had no effect. These results led Marsh, Kuyper, et al. (2014) to conclude, "This might even suggest that school context really has no effect and its apparent effect is merely a reflection that schools with high school average achievement are made up of classes with high class-average achievement" (p. 58). Similarly, Liem et al. (2013), using a sample of 4,461 Singaporean students from 136 classes and 9 schools, found significant class effects but no school effects in a joint model.

**Juxtaposing Multiple Class Environments as Frames-of-Reference for Academic Self-Concept Formation**

In school systems around the world, including those in the United States, the United Kingdom, Australia, Canada, and New Zealand, many schools track students on a course-by-course basis (Chmielewski, 2014). In contrast to between-school tracking (students from different ability levels attend different schools) and within-school streaming (students from different ability levels attend the same school but are then assigned to different streams for all subjects), course-by-course tracking is defined as "offering courses at varying levels of difficulty in one or more subjects within a school" (Chmielewski, 2014, p. 293). In the following, we will sometimes —for reasons of simplicity—refer to course-by-course tracking as "tracking".

In course-by-course tracked systems students are, depending on the subject, assigned to different ability tracks that in turn are taught in separate classrooms. These systems usually do not assign students to ability tracks in all of the subjects. For instance, Loveless (2013) showed for the United States that course-by-course tracking in math occurs much more frequently compared with language, science, or history. The fact that students in course-by-course tracked systems are not ability tracked in all subjects typically leads to students attending non-tracked subjects in the same regular class.

In course-by-course tracked school systems, in which students belong to several class environments, students can form their academic self-concept in a certain domain, for example, in math, by comparisons with classmates from their domain-specific class (e.g., math class) and their regular class. In relation to the local dominance effect, both classes are local as students are directly exposed to classmates from both classes in daily teaching lessons. However, they differ concerning their domain-specific proximity. Thus—according to the local dominance theory—one would expect the domain-specific class to be the pivotal frame-of-reference for academic self-concept formation in that domain.

In course-by-course tracked systems, the question is not only to which class environments students compare but also how respective comparison processes might differ regarding assimilation effects. Based on previous research, one would expect that controlling for domain-specific track level should increase the BFLPE on the domain-specific class level because it controls for assimilation (Chmielewski et al., 2013). For instance, students in high achieving math classes are more likely to be high math track members, resulting in a confounding of contrast and assimilation. However, one would not expect that students in high achieving regular classes are more likely to be high math track members. In other words, there might not be such a confounding of contrast and assimilation processes. Thus, controlling domain-specific track level is not expected to increase a BFLPE on the regular class level.

In course-by-course tracked systems, the question is not only to which class environments students compare themselves but also how respective comparison processes might differ regarding grading on a curve. Whereas previous research found frame-of-reference effects of domain-specific class environments to be mediated by grades and interpreted this result as a consequence of class-referenced grading stimulating the BFLPE (Marsh, 1987; Trautwein et al., 2006), no study exists that has investigated if this is the case for domain-unrelated class environments. Investigating this question is especially important as it contributes to a better understanding of the mechanisms of frame-of-reference effects on different class levels. Based on previous research, one would expect that controlling for domain-specific grades will decrease the BFLPE on the domain-specific class level because it controls for the teachers' tendency to conduct class-referenced grading. For instance, students in high achieving math classes are provided with worse math grades resulting in confounding of the BFLPE and grading on a curve. However, one would not expect that controlling domain-specific grades will increase the BFLPE on the regular class level as students in high achieving regular classes are not expected to be provided with worse grades.

To date, research focused on only one class environment—in most cases the domain-specific class environment—as the frame-of-reference for academic self-concept formation. To our knowledge, no study has juxtaposed several class environments as frames-of-reference for academic self-concept formation. One reason for that is that educational large-scale datasets usually do not contain information about multiple class memberships. Another reason for the scarce research on this issue is that it relies on survey designs that test all students within sampled schools. Not testing complete schools will lead to differential sampling rates for different classes that will in turn result in differences in the reliability of aggregates and biased estimates.

The juxtaposition of multiple class environments as frames-of-reference for academic self-concept formation has high theoretical and practical relevance. Regarding the former, it captures the full complexity of academic self-concept formation in course-by-course tracked systems—an issue that previous research neglected. Regarding the latter, disentangling contextual effects of multiple class environments comes with implications for the composition of learning environments.

## The Present Study

The present study is based on data from the Austrian national educational standard assessment in 2012 (BIFIE, 2016; Schreiner & Breit, 2012), which measured all Austrian eighth-grade students in the domain of math. Austrian general secondary school students were assigned to one of three tracks (low, medium, high) in the core subjects of mathematics, German, and English, based on teachers' subjective impression of students' achievement. Students from the different tracks were usually taught in separate classrooms according to curricula that differed in performance requirements. As there might have been students who were good in all three core subjects, the class composition of core subjects might have been more or less similar. Secondary school students attended all other subjects (e.g., history, geography, biology, chemistry, physics, music, domestic education) in the same regular class that was not grouped according to ability. Thus, in our multilevel data, students (level 1) were nested in the cross-classification between math classes (level 2a) and regular classes (level 2b) that were nested within schools (level 3; see

Figure 1 for a graphical description of the data structure; a more detailed explanation of the complex data structure can be found in the Data section).

As the Austrian national educational standard assessment in 2012 (BIFIE, 2016; Schreiner & Breit, 2012) identified students' school, math class, and regular class and additionally measured all students—what enabled us to build reliable math achievement aggregates on all levels of the data hierarchy—these data were perfectly suited for juxtaposing multiple class environments as frames-of-reference for academic self-concept formation, thus filling the research gap concerning the pivotal frames-of-reference for academic self-concept formation in course-by-course tracked systems. In order to take a closer look at the different mechanisms of the level-specific frame-of-reference effects, we were also interested in how additionally modeling math track and math grades affected the different contextual effects. In detail, we hypothesized the following:

Hypothesis 1 (H1; see also Figure 2a): When considered separately, each of the three math achievement aggregates (school, math class, and regular class math achievement) is expected to have a negative effect on math self-concept, controlling for individual math achievement. Thus, we expected to find a school, a math class, and a regular class BFLPE.

Hypothesis 2 (H2; see also Figure 2b): When all three math achievement aggregates are considered together, controlling for individual math achievement, we expected the math class BFLPE to be more negative than the regular class BFLPE, which we in turn expected to be more negative than the school BFLPE.

Hypothesis 3 (H3; see also Figure 2c): When additionally modeling track level, we expected it to contribute positively to math self-concept and result in a more negative math class BFLPE. However, we did not expect it to substantially change the regular class BFLPE. In an exploratory endeavor, we were also interested in whether effects differed between pure and mixed math classes (i.e., math classes with students from the same math track [pure] vs. math classes with students from different math tracks [mixed]).

Hypothesis 4 (H4; see also Figure 2d): In a preliminary analysis, we were interested in whether grades are impacted by frame-of-reference effects. When additionally modeling grades in the BFLPE model, we expected it to contribute positively to math self-concept and result in a less negative math class BFLPE. However, we did not expect it to substantially change the regular class BFLPE.

# Method

## The Austrian Educational System

In Austria, children attend primary school from Grades 1 to 4, then attend secondary school from Grade 5 onward (for a detailed description of the Austrian school system, see Bruneforth et al., 2016). Depending on their primary school achievement, students attend either (a) academic or (b) general secondary school. Academic secondary school provides students with deepened general knowledge and requirements for a transition to university. In the school year 2011-2012—in which eighth-grade student data for the present study were collected—about 33% of all Austrian eighth-grade students attended this school type. In contrast, general secondary school prepares students for vocational training or the transition to higher education. In the school year 2011-2012, about 67% of all Austrian eighth-grade students attended this school type (Schreiner & Breit, 2012). Because our study's main goal was to juxtapose multiple class environments as frames-of-reference for academic self-concept formation—and this could only be done with the subsample of course-by-course tracked students from general secondary schools—we focused on this group of students in the present paper. For reasons of transparency, we report descriptive statistics and analyses on the pivotal frames-of-reference for the total student sample as well as the academic secondary school subsample in the supplemental material (see Tables S12–S15).

Students attending general secondary school were assigned to one of three tracks (low, medium, high) in three core subjects including mathematics. Generally, students from different tracks were provided with different curricula that differed concerning the topics to be addressed as well as the depth with which the topics were treated. However, in the end, it was left to the teacher to decide how to design the curriculum. In math classes with students from different math tracks, it was also left to the teacher to decide how to deal with math class heterogeneity in terms of track level. Some teachers differentiated their teaching by instructing students from one track while students from the other tracks worked on their own. Some teachers provided very much the same classroom instruction to students from different tracks, however, provided tests with varying degrees of difficulty according to students' track level. Other teachers provided the same tests to students from varying tracks but applied different grading schemes. Note that beginning in the school year 2012-2013, course-by-course tracking was successively abolished and does not exist anymore today (Eder et al., 2015).

**Data**

In Austria, the Federal Institute for Educational Research, Innovation, and Development of the Austrian School System (BIFIE) conducts national educational monitoring. The examinations of Austria's educational standards are conducted as comprehensive surveys, aiming at measuring all Austrian students attending the fourth or the eighth grade without special educational needs. The national educational standard assessment from 2012 (BIFIE, 2016; Schreiner & Breit, 2012), which is the database for the present study, was conducted in May 2012. The assessment was aimed at testing all Austrian eighth-grade students without special educational needs in the domain of mathematics. About 4% of the students could not be tested, mostly due to absence at the main and alternative testing dates. The Austrian national educational standard assessment is prescribed by law and does not require the consent of students or parents. Data access was approved by the BIFIE and required consent to data protection regulations.

Our sample included 50,208 students (48% female, $M_{age}$= 14.44 years) from 1,078 general secondary schools, 3,449 math classes, and 2,729 regular classes. On average, there were *M* = 3.20 (*SD* = 1.11) math classes and *M* = 2.53 (*SD* = 0.96) regular classes per school with *M* = 14.56 (*SD* = 5.32) students per math class and *M* = 18.40 (*SD* = 4.10) students per regular class. The math classes were on average smaller than regular classes because schools that contained only two regular classes split the student body into three math classes according to the three track levels. As noted above, math classes typically contained students from one and the same math track. However, in small schools, math classes might also have contained students from different math tracks. Generally, 73% of all math classes were composed of students from one and the same math track. In the subsample of students from mixed math classes, we found that every student attends his math class with *M* = 12.54 (*SD* = 6.54) other students from his regular class, indicating a moderate overlap between both class environments for students in mixed math classes. Generally, students spent about 15% of the weekly lesson time in each of the three core subjects (in total 45%) and the other 55% in their regular class.

## Instruments

**Math self-concept**. Math self-concept (MSC) was assessed using four items (i.e., *Usually I am good in mathematics; Mathematics is harder for me than for many of my classmates; I am just not good in mathematics; I learn quickly in mathematics*), which were answered on a 4-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*; BIFIE, 2012). The academic self-concept scale used in the Austrian Educational Standard Assessment basically corresponds to the academic self-concept scale employed by the TIMSS study (the TIMSS study comprises one additional item *I am good at working out difficult mathematics problems;* Mullis et al., 2016). Multiple studies successfully used this scale (e.g., Guo et al., 2018) and confirmed its construct validity by demonstrating measurement invariance across cultural contexts (e.g., Marsh et al., 2015; Marsh, Abduljabbar, et al., 2014) or gender (e.g., Abu-Hilal et al., 2013). We interpret this body of research as evidence for our scale's psychometric quality. For subsequent analyses, a mean score comprising the items was constructed (at least two items had to be completed for mean score calculation; $\alpha$ = .85). As the procedure of aggregating Likert scale items by calculating mean scores depends on the equidistance between adjacent scale points (Carifio & Perla, 2008), we also created IRT-based academic self-concept WLE scores based on the rating scale model (de Ayala, 2009). These WLE scores correlated with the mean scores by $r$ = .96. Additionally, we ran our main analyses with IRT-based academic self-concept WLE scores as the outcome. The results did not substantially differ from those using the mean scores (Tables S16 and S17). Table 1 contains descriptives for all model variables (for the level-specific correlations, see Table S10). Average MSC in our sample was $M$ = 2.97 ($SD$ = 0.76). Most of the MSC variation was located on the individual level ($\hat{p}_{ind}$ = .77). Math class variation in MSC was lower ($\hat{p}_{mcl}$ = .18), and variability on the regular class level ($\hat{p}_{rcl}$ = .00) and the school level ($\hat{p}_{sch}$ = .05) was even lower. The low variance proportions on the higher level student hierarchies are a common finding in research on the BFLPE (e.g., see Nagengast & Marsh, 2012; Trautwein et al., 2006). The relatively low variation in academic self-concept between educational environments is consistent with BFLPE theory proposing that students build their academic self-concept as a consequence of social comparison processes within educational environments. Table S1 in the online supplemental material presents the descriptives for each math track separately. Students in the high track ($M$ = 3.16, $SD$ = 0.70) had higher MSC than those in the medium track ($M$ = 2.91, $SD$ = 0.73) and those in the low track ($M$ = 2.55, $SD$ = 0.79). To assess the quality of our academic self-concept scale, we conducted tests of differential item functioning with ordinal logistic regression techniques (Crane et al., 2006) for gender, migration background and track level using the R-Package *lordif* (Choi et al., 2011). We did not find evidence for meaningful DIF based on the pseudo $R^2$ measure (Jodoin & Gierl, 2001; Kim et al., 2007; Zumbo, 1999) as well as $\beta$ differences (Crane et al., 2004; Crane et al., 2007; Jodoin & Gierl, 2001).

**Math achievement.** Math achievement (MACH) was measured using a math competencies test that was based on the competency model of the Austrian educational standards (Schreiner & Breit, 2012). The competency model of the Austrian educational standards—similar to the PISA concept of mathematical literacy (OECD, 2017a)—focuses on the mastery of processes, the understanding of concepts, and the ability to deal with different everyday situations and problems within a competence area on the basis of sustainably networked knowledge. The test lasted about 90 minutes and was delivered by means of a multi-matrix design that contained several test booklets. Students completed approximately 48 items, mostly being presented in a multiple-choice format. There also existed a limited amount of half-open and open item formats.

The BIFIE provides ten plausible values (PVs) that represent the likely distribution of a person's ability (von Davier et al., 2009; Wu, 2005). Large-scale assessment studies typically use PVs because such a procedure allows taking into account the uncertainty of person parameter estimation, thus allowing for correctly estimating associations with other variables. However, due to the multi-matrix design, it was not possible to calculate marginal reliabilities for the PVs. Thus, we calculated an alternative reliability coefficient as it is used in PISA, deducting the within-person PV variance proportion from one. A reliability coefficient close to one indicates that PVs vary within individuals only to a small extent, thus pointing to high measurement accuracy (Adams, 2005; OECD, 2017b). This reliability coefficient was 0.91. MACH showed high variation on the individual level ($\hat{\rho}_{ind}$ = .38) and the math class level ($\hat{\rho}_{mcl}$ = .43), whereas variability on the regular class level ($\hat{\rho}_{rcl}$ = .06) and the school level ($\hat{\rho}_{sch}$ = .14) was lower. This finding empirically underlines the group assignment mechanism. Average MACH was *M* = 504.24 (*SD* = 86.85). MACH correlated with MSC by *r* = .39. Students from the high math track (*M* = 564.77, *SD* = 69.50) had higher MACH as compared to students from the medium track (*M* = 477.38, *SD* = 59.10) and the low track (*M* = 414.02, *SD* = 56.06).

**Math grade**. Students self-reported their math grade (MGRA) in the last half-year report card. As students were measured in May 2012 and report cards were provided in February 2012, this kind of performance feedback was still relatively current. Generally, it can be assumed that self-reported grades provide a reliable measure of actual grades (Sticca et al., 2017). In Austria, grades are given on a scale from 1 to 5 with 1 representing the best grade. For subsequent analyses, we inverted the grade variable so that higher grades reflect higher achievement. MGRA mainly varied on the individual level ($\hat{\rho}_{ind}$ = .63) with variance proportions on the math class level ($\hat{\rho}_{mcl}$ = .19), the regular class level ($\hat{\rho}_{rcl}$ = .07), and the school level ($\hat{\rho}_{sch}$ = .10) being substantially smaller. Average MGRA was *M* = 3.19 (*SD* = 0.94). The correlation between MGRA and MACH was *r* = .39.

**Math track.** Students self-reported the math track they were associated with. Math track can be regarded as a level 1 variable as there where some math classes that contained students from several math tracks (also see above). Generally, track assignment was mainly based on teachers' subjective impression of their students' achievement. We created two dummy variables for the low and the high math tracks with medium math track representing the reference category.

**Statistical Analyses**

We applied multilevel linear regression analyses (Hox et al., 2017; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) using the statistical computing software R (R Core Team, 2008) and the package lme4 (Bates et al., 2015). Generally, all analyses were run with 10 datasets that differed concerning the achievement variables (10 plausible values are provided by the BIFIE). Results then were pooled by Rubin's (1987) rules using the lmer_pool function drawn from the package miceadds (Robitzsch et al., 2018).

We addressed our research question by calculating multilevel models in which we regressed math self-concept on math achievement aggregates on all levels of student nesting. In these models, level 1 variables were standardized and all three achievement aggregates (math achievement for level 3 school, level 2a math class, and level 2b regular class) were calculated based on the standardized measure, but not re-standardized. As a result, all math achievement variables are in the same metric, namely standard deviations of individual math achievement, making coefficients comparable across levels and models. By grand mean centering of level 1 variables, respective higher level effects can be interpreted as effects of the higher level aggregates, controlling for individual variables, also referred to as contextual effects (Enders & Tofighi, 2007). We also ran our analyses using a centering within cluster centering (CWC) approach. The respective results can be found in the supplemental material (Table S11).

For juxtaposing multiple class environments as frames-of-reference for academic self-concept formation, we regressed math self-concept on individual math achievement and math achievement aggregates at the school, the math class, and the regular class levels:

$$Self-concept_{i(j_1,j_2)k}=\gamma_{0000}+\gamma_{1000}\cdot achievement_{i(j_1,j_2)k}+\gamma_{0001}\cdot school\ achievement_l+\gamma_{0100}\cdot math\ class\ achievement_{j_1k}+$$

$\gamma_{0001}$, $\gamma_{0100}$, and $\gamma_{0010}$ can be interpreted as the BFLPEs on the school, the math class, and the regular class levels. Respectively $v_{000k}$, $u_{0j_10k}$, and $u_{00j_2k}$ are the random school, math class, and regular class effects. $e_{i(j_1,j_2)k}$ is the residual term. The residual terms are assumed to be normally distributed and independent of each other (Beretvas, 2011).

In our modeling sequence, we started by first regressing academic self-concept on the average achievement measure on the different levels separately. We calculate these models because they illustrate the widespread procedure of considering only one out of several educational environments. However, we note that modeling only average achievement scores from all levels of student nesting simultaneously reveals the pivotal frames-of-reference for academic self-concept formation as it controls for the confounding of the aggregate achievement measures.

We ran all our statistical models using a complete case analysis approach (also known as "listwise deletion"). Thus, cases that had missing values on at least one model variable were excluded. The procedure resulted in exclusion rates between 1 and 8%, depending on the statistical model. Research on missing data suggests that when the loss of cases is small—like it was in our study—a complete case analysis will result in negligible parameter bias (Graham, 2009). However, we also conducted a robustness check, in which we calculated all models with multiply imputed data (Table S7). The results were the same as for the complete case analysis approach.

# Results

## Separate BFLPEs (H1)

Before starting the actual modeling procedure, we estimated an intercept-only model in order to estimate the variance proportions of academic self-concept on the different levels of the modeling hierarchy. As already mentioned in the method section, the variance proportions were 77% (individual), 18% (math class), 0% (regular class), and 5% (school). In H1, we assumed that when considered separately, each of the three math achievement aggregates (school, math class, and regular class math achievement) should negatively predict math self-concept, controlling for individual math achievement. As shown in Table 2 and consistent with our hypothesis, we found

BFLPEs on the school (Model 1; $b = -.42$, $p < .001$, 95% CI = [-.45, -.39]), the math class (Model 2: $b = -.43$, $p < .001$, CI = [-.45, -.41]), and the regular class levels (Model 3; $b = -37$, $p < .001$, CI = [-.39, -.35]) when modeled separately. Some of the variance components (e.g., on the math class as well as the regular class level) were higher than in the empty model. Note that in multilevel regression models—in contrast to ordinary single-level regression models—the inclusion of new predictors can increase higher level variance in the outcomes (Gelman, 2008).

**Pivotal Frames-of-Reference (H2)**

In order to reveal the pivotal frames-of-reference for academic self-concept formation, we examined math achievement aggregates on all levels of student nesting simultaneously in one model (Model 4; see Table 2). In H2, we expected the math class BFLPE to be more negative than the regular class BFLPE and the regular class BFLPE to be more negative than the school level BFLPE. We found the math class BFLPE to amount to $b = -.37$, $p < .001$, 95% CI = [-.39, -.35], whereas the regular class BFLPE was $b = -.11$, $p < .001$, CI = [-.15, -.08] and the school BFLPE was $b = -.06$, $p = .004$, CI = [-.10, -.02]. Whereas the math class BFLPE was smaller than the regular class BFLPE on an inferential level ($p < .001$), this was not the case for the regular class BFLPE and the school BFLPE ($p = .157$). The corresponding interpretation is that equally able students in equally able schools and regular classes had a substantially lower self-concept in high achieving math classes. Equally able students in equally able schools and math classes had lower self-concept in high achieving regular classes, but equally able students in equally able math and regular classes had only a little lower self-concept in high achieving schools. These results are in line with local dominance theory as the size of the level-specific BFLPEs differs as a function of the proximity of respective learning environments. The variance proportions were 72% (individual), 13% (math class), 6% (regular class), and 9% (school).

**Track Level and the BFLPE (H3)**

Next, we modeled math track as an additional predictor variable (Model 5; Table 3). We found high math track students to have a more positive self-concept as opposed to medium track students ($b = .19$, $p < .001$, 95% CI = [.17, .22]). Conversely low track students had lower self-concept ($b = -.32$, $p < .001$, CI = [-.35, -.30]) as opposed to medium track students. In line with H3, the math class BFLPE changed (from $b = -.37$, $p < .001$, 95% CI = [-.39, -.35] in Model 4) to $b = -.53$, $p < .001$, CI = [-.56, -.51]. This change was also supported on an inferential level (95% CIs did not overlap). Thus, equally able students in equal math tracks experienced a more severe math class BFLPE. Also in line with H3, the regular class BFLPE did not change substantially from $b = -.11$, $p < .001$, CI = [-.15, -.08] in Model 4 to $b = -.09$, $p < .001$, CI = [-.13, -.06] in Model 5 (95% CIs did overlap). The school BFLPE did change substantially, from $b = -.06$, $p = .004$, CI = [-.10, -.02] in Model 4 to $b = .05$, $p = .024$, CI = [-.01, .09] in Model 5 (95% CIs did not overlap). The variance proportions were 72% (individual), 13% (math class), 7% (regular class), and 8% (school).

To check if associations differed between students from pure and mixed math classes, we calculated an additional set of analyses in which we included all interactions between a “mixed” dummy variable and the model variables (Table 4). Mixed math classes consisted of students from different math tracks, whereas pure math classes included classes with students from the same math track. Generally, we did not find differences in the frame-of-reference effects between students from mixed and pure math classes. However, we indeed found differences in the track-level effects. Students in mixed math classes experienced more negative track-level effects from the low track as indicated by the negative interaction between the low track and the mixed dummy ($b = -.19$, $p < .001$). Additionally, students in mixed math classes experienced a more positive track-level effect from the high track as indicated by the positive interaction between the high track and the mixed dummy (b = .14, p < .001).

**School Grades and the BFLPE (H4)**

To investigate the frames-of-reference for grade provision, we regressed grades on achievement aggregates on all levels of student nesting (see Table 5). As expected in H4, we found the math class average achievement effect to be most pronounced ($b$ = -.30, $p$ < .001). We found the school average achievement effect to be less negative ($b$ = -.11, $p$ < .001) and the regular class effect to be positive ($b$ =.07, $p$ < .001).

Following this, we modeled math grades (in addition to standardized achievement on all levels of student nesting) as an additional predictor of academic self-concept (Model 6; see Table 3). We found math grades to have a strong positive effect on math self-concept ($b$ = .39, $p$ < .001, 95% CI = [.38, .40]). Students with better grades had higher self-concept. In line with H4, we found the math class BFLPE to be substantially reduced (from $b$ = -.37, $p$ < .001, 95% CI = [-.39, -.35] in Model 4) to $b$ = -.26, $p$ < .001, CI = [-.28, -.24]. This change was also supported on an inferential level (95% CIs did not overlap). Additionally we found that the school ($b$ = -.02, $p$ = .297, CI = [-.06, .02]; $b$ = -.06 in Model 4) and the regular class BFLPEs ($b$ = -.14, p < .001, CI = [-.18, -.11]; b = -.11 in Model 4) were not substantially changed by the inclusion of school grades (95% CIs did overlap). This suggests that equally able students with equal grades did not experience a more or less severe school or regular class BFLPE. The variance proportions were 73% (individual), 12% (math class), 7% (regular class), and 9% (school).

**Robustness Checks and Additional Analyses**

To check the robustness of our results, we ran additional sets of analyses. First, as the academic self-concept item "*Mathematics is harder for me than for many of my classmates*" directly referred to social comparisons, we reran all models with a self-concept score in which this item was excluded. This did not change the results (Table S2). Second, we reran all models with the inclusion of covariates, namely sex, age, SES, and migration background. Also, this did not substantially change the results (Table S3). Third, we ran additional analyses with math class average track level instead of track level. We did this because prestige measures in the assimilation effects literature often represent class-level variables. In our main analyses, this was not true for students from math classes that contain students from several math tracks. Modeling average track level instead of track level did not substantially change the results (Table S4). In additional exploratory analyses, we also calculated the interactions between the track-level dummies as well as the BFLPEs on the different levels. We found that low-track students experienced a more positive school-level BFLPE as opposed to medium-track students, whereas the opposite was true for high-track students (Table S5). Furthermore, high-track students experienced a more negative math class level BFLPE as opposed to medium-track students but a more positive regular class level BFLPE. We also calculated all models with only the social comparison item as the dependent variable. The results were the same as for the complete scale (Table S6). Moreover, we calculated all models with the help of multiply imputed data (Table S7). The results were the same as for the complete case analysis approach. Additionally, we ran all analyses for the subsample of students from math classes that contain only students from the same track (Table S8). This did not substantially change the results. Finally, in an exploratory endeavor, we also conducted moderation analyses in which we specified the interactions between the achievement aggregates and sex, age, migration, and SES (Table S9). None of these interactions were statistically significantly different from zero.

## Discussion

When regressing math self-concept on math achievement on all levels of student nesting, math class achievement had the strongest negative effect (math class BFLPE), regular class achievement had a less negative effect (regular class BFLPE), and school achievement had the least negative effect (school BFLPE). Additionally, controlling for track level increased the math class BFLPE but did not substantially change the regular class BFLPE. Additionally, controlling for grades decreased the math class BFLPE but did not substantially change the regular class BFLPE. In sum, our study suggests that in course-by-course tracked systems, multiple class environments may act as frames-of-reference for academic self-concept formation and that mechanisms of respective social comparison processes might differ from each other.

Our paper offers several unique contributions to the BFLPE literature and more broadly to the literature on academic self-concept formation. Our study's overall contribution is the investigation of the BFLPE and its potential mechanisms in course-by-course tracked systems in which students are members of not one but multiple class environments. More specifically, our study is the first to (a) juxtapose multiple class environments as frames-of-reference for academic self-concept formation and (b) investigate assimilation and grading on a curve as a potential mechanism for frame-of-reference effects of these multiple class environments. Regarding (a) we found the (domain-specific) math class achievement and to a weaker extent the (domain-unrelated) regular class achievement to negatively predict domain-specific academic self-concept. This finding suggests that students in course-by-course tracked systems evaluate their abilities against students from not one but multiple class environments. Regarding (b) we found BFLPEs of multiple class environments to differentially react to controlling for track level and grades. The math class BFLPE increased when controlling for track level and decreased when controlling for grades; in contrast, there was no substantial change to the regular class BFLPE. One interpretation of our results is that math class BFLPE is counterbalanced assimilation and associated with grading on a curve, whereas this is not the case for the regular class BFLPE. Additionally, our study contributes to the educational psychological literature by investigating differential track-level effects for students from math classes that contain students from the same math track (pure math classes) and students from math classes that contain students from different math tracks (mixed math classes). We found more pronounced track-level effects in mixed math classes, suggesting track-level saliency to amplify this prestige effect. Also in additional analyses, we investigated frame-of-reference effects on grades in course-by-course tracked school systems. We found grades to be negatively predicted by math class and to a lower extent also by school achievement, suggesting that teachers conduct class- and school-referenced grading.

**Pivotal Frames-of-Reference for Academic Self-Concept Formation**

Generally, we found all math achievement aggregates on all levels of student nesting (math class, regular class, and school) to negatively predict math self-concept when modeled separately. However, when conjointly modeling all these predictors, the math class BFLPE was dominant. These results provide renewed evidence that the use of traditional large-scale datasets that do not allow for modeling all levels of student nesting is most likely to result in biased estimates of level-specific BFLPEs.

When conjointly modeling math achievement aggregates on all levels of student nesting, we found a small school BFLPE. This result is somewhat in contrast to that of Marsh, Kuyper, et al. (2014) who did not find a school BFLPE when class achievement was taken into account. However, Marsh, Kuyper, et al. (2014) conducted their study with a Dutch student sample in which students were tracked in relation to all classes. We also note that the very large sample size in our study meant that even a small BFLPE at the school level was highly significant.

When conjointly modeling math achievement aggregates on all levels of student nesting, we also found a regular class BFLPE that was smaller than the math class BFLPE. As both educational environments might be considered to be similar concerning their local proximity but differ concerning their domain-specific proximity, these results are in line with—but also clarify and extend—local dominance theory. But how can average math achievement of regular classes affect students' math self-concept? Our explanation is that it is likely that students had a relatively accurate perception of the math achievement of regular classes because track membership was highly salient. Thus, students might have had lower math self-concept in regular classes with high math achievement as a consequence of being surrounded by lots of students from the high math track. Conversely, students might have had higher math self-concept in regular classes with low math achievement as a consequence of being surrounded by lots of students from the low math track.

**Track Level and the BFLPE**

When additionally modeling track level, we found students from higher math tracks to have higher math self-concept. One interpretation of this finding is that students experienced assimilative track branding effects (e.g., "I am in a high math track, thus I am good at math"). However, as already noted in the introduction, information on track level is confounded with students' prior achievement, as track designation is based on prior achievement (Marsh et al., 2018). Unfortunately, we cannot resolve these opposing interpretations with data available in the present investigation. Thus, the disentanglement of positive track branding effects and effects of prior achievement is a fruitful direction for future research.

When additionally modeling track level, the math class BFLPE increased, whereas this did not change the regular class BFLPE. We interpret this finding as a consequence of the math class BFLPE being counterbalanced by assimilation, whereas this was not the case for the regular class BFLPE. This result suggests that frame-of-reference effects of multiple class environments might differ in their mechanisms.

In addition, we found students from mixed math classes to experience more pronounced track-level effects on academic self-concept. We interpret this as a consequence of increased salience of track level in mixed math classes.

**School Grades and the BFLPE**

We found school grades to be negatively predicted by math class achievement and school achievement, whereas the effect of regular class achievement was slightly positive. This result suggests that teachers—next to providing grades on a class-referenced basis—additionally grade on a school-referenced basis. The rather unexpected frame-of-reference effect on the school level may result from two aspects. First, several math classes from one school might be taught by the same teacher. These teachers might evaluate students in their classes on the same scale, for instance with the same tests, which induced the frame-of-reference effect on the school level. Unfortunately, no teacher ID is provided in the data so we are not able to empirically test our assumption. Another explanation for our finding might be that math teachers use common testing standards. For instance, they might use identical test materials and standardized result protocols, which are comparable across classes within schools. We also found that regular-class achievement positively affected grades when controlling for achievement on all other levels of student nesting. This finding is somewhat surprising as we would not have expected any associations between regular class achievement and grades. In other words, why should teachers provide better grades for students that come from a high achieving regular class? We can only speculate about possible mechanisms. For instance, teachers might perceive students from high achieving regular classes to be more competent, thus providing them with better school grades.

When math grades were included in the model, there was a substantial positive effect of math grades on math self-concept. Additionally, the math class BFLPE decreased substantially whereas this was not the case for the regular class BFLPE. We interpret this finding in that the math class BFLPE was associated with grading on a curve. This result suggests that in frame-of-reference effects of multiple class environments might differ in their mechanisms. As already noted in the introduction, previous research interpreted this to mean that the BFLPE was caused, at least in part, by grading-on-a-curve driving the BFLPE (e.g., Marsh, 1987; Trautwein et al., 2006). However, hypothesized causal effects are difficult to test with correlational data. Indeed, recent discussion suggests that there is a strong evolutionary basis for social comparison processes (Frank, 2011). Marsh et al. (2018) argued that this explains why the BFLPE is so cross-culturally robust, and this supports claims that social comparison processes underpinning the BFLPE are pan-human and universal (Marsh & Seaton, 2015). From this perspective, it might be possible that the social comparison processes underlying the BFLPE are so strong that they are independent of the provision of class-referenced grades because students socially compare themselves in relation to other students in a similar fashion whether or not they are assigned with school grades. Thus, for example, would the size of the BFLPE decrease if students were not assigned grades at all or were assigned grades in relation to a common metric rather than grading on a curve? Although beyond the scope of the present investigation, we note that more research is needed to determine whether grading-on-a-curve is a causal contributor to the BFLPE or merely an effect that is correlated with the BFLPE.

**Limitations and Directions for Future Research**

Although our study is based on strong data, some potential limitations should be addressed in the future. First, students were tracked not only in relation to math but also in German and English. However, we had no information about German and English class membership. Thus, every student was associated with two more class environments that were not included in our analysis. Future research should aim at juxtaposing all class environments as frames-of-reference for academic self-concept formation. However, such an endeavor requires a comprehensive dataset with complete information on students' multiple course memberships.

A second potential limitation of the present investigation is that it is based on cross-sectional population data, thus we cannot provide firm causal inference. For two reasons, we argue that our correlational approach, which is of course not perfect, still provides a rather strong design to investigate the desired research questions. First, an internally valid juxtaposition of multiple class environments as frames-of-reference for academic self-concept formation would require the random assignment to multiple class environments that differ in their average achievement. More specifically, it would require randomizing students to schools with different achievement levels, while simultaneously keeping class achievement constant. Likewise, it would require randomizing students to classes with different achievement levels, while simultaneously keeping school achievement constant. For ethical, organizational, and political reasons there is no chance to conduct such a study. Second, our study shows that achievement aggregates from different student environments are highly correlated and that controlling for all student environments results in a completely different picture. For instance, the school BFLPE shrinks by about 85% (from -.42 to -.06). Thus, we argue that our study which controls for achievement aggregates of different student environments has substantially improved in internal validity in contrast to previous studies that considered only one student environment (e.g., the school).

Third—although data from the Austrian national educational standard assessment represents a comprehensive survey, thus providing nearly perfect external validity for the Austrian context—it remains unclear to what extent our results are generalizable to other countries and educational systems. Due to differences in teacher communication or grading policies, it might be the case that pivotal frames-of-reference for academic self-concept formation in other student populations deviate from those we found. Thus, future studies should replicate our findings in different cultural contexts. Additionally, prior research has produced evidence that the BFLPE is stronger in math as opposed to verbal domains (e.g., Guo et al., 2018). In this paper, we focused on mathematics, as this was the central domain of the national educational standard assessments in 2012. Future research is needed to test the generalizability of our results to other domains (e.g., language). Limitations of external validity also concern the transferability of the results to other age groups. For instance, the local dominance effect might be stronger in younger age groups that evaluate their abilities primarily concerning very proximal environments, whereas older age groups might take into account also less proximal comparison information. Limitations of external validity also concern the transferability of the results to other educational systems.

Finally, track level, as well as grades, were self-reported by students. Thus self-report bias, such as social desirability, might have impacted the reliability and validity of our measures. As it is rather unlikely that self-report bias differentially occurred for different groups of students, we think that it did not affect the relationship between the variables. If this would have been the case, however, the grade- and track-level estimates that we found would have been conservative estimates. Concerning grades, there is also empirical evidence that self-reported measures provide reliable indicators of actual grades (Sticca et al., 2017). Additionally, neither grades nor track level were the central constructs in Hypothesis 3 and Hypothesis 4.

**Practical Implications**

Generally, the very basic BFLPE finding—equally able students have lower self-concept in high achieving educational environments—has a variety of practical implications. These implications can be divided into (a) implications for individual educational careers and (b) implications for educational systems. Regarding (a), the BFLPE predicts that individual educational careers that will result in changes in the average achievement of a student's educational environment will be accompanied by changes in the student's academic self-concept. In this context, the BFLPE predicts that school transfers, educational transitions, course choices, track changes, or grade retention of a student may be accompanied by changes in his academic self-concept (e.g., Wouters et al., 2012). Regarding (b), the BFLPE predicts that changing educational systems concerning the composition of educational environments will result in changes in students' academic self-concept. Specifically, this means that every form of ability segregation (e.g., different forms of tracking) should increase the academic self-concept of low achievers because it decreases the average achievement of these students' educational environments (Hübner et al., 2017; Hübner et al., 2020). Vice versa, the BFLPE predicts that ability desegregation (e.g., detracking) will decrease the academic self-concept of low achievers because it increases the average achievement of these students' educational environments.

Given these predictions of the BFLPE, the question arises on how educational policymakers should shape their school systems to reduce the negative consequences of the frame-of-reference effect. First of all, it has to be noted that the BFLPE is a "zero-sum game" (Trautwein & Möller, 2016). This means that a low achieving student that encounters a high achieving classroom will have lower academic self-concept but he will also lower the class average achievement of that class, increasing the academic self-concept of other students. Similarly, detracking will result in an academic self-concept decline of low achievers but an academic self-concept increase of high achievers. Additionally, the BFLPE applies to student motivation in terms of academic self-concept, but there is still an ongoing discussion of how selective student environments affect students' academic achievement (Dicke et al., 2018; Stäbler et al., 2017). Some studies suggest that selective learning environments positively impact students' academic achievement via peer spillover effects (Ammermueller & Pischke, 2009; Burke & Sass, 2013). However, also opposite effects have been found (e.g., Dicke et al., 2018; Televantou et al., 2015). Nevertheless, suggestions have been made for counteracting the negative consequences of the BFLPE. For instance, Marsh and Seaton (2015) suggest avoiding a competitive environment, enhancing students' feeling of connection, or valuing students' unique accomplishments as potential measures to reduce the negative consequences of BFLPE.

Unfortunately, there is almost no evidence for the effectiveness of such endeavors. For instance, several studies show that the size of the BFLPE seems not to be affected by feedback practices (Lüdtke et al., 2005) or motivational climate (Wouters et al., 2013). Thus, the BFLPE has repeatedly been described as an unavoidable aspect of human nature (Marsh et al., 2020). Against this background, we were able to identify only one study that found differentiated instruction to weaken the BFLPE (Roy et al., 2015). However, in sum, studies consider the BFLPE an unavoidable aspect of human nature (Marsh et al., 2020). Our study investigated BFLPE and its proposed mechanisms within course-by-course tracked school systems in which students are members of multiple class environments. Accordingly, our findings allow for a refinement of BFLPE predictions presented above. Regarding individual educational careers, our study results suggest that a student's academic self-concept in a certain domain may be strongly hurt when placing him in high achieving domain-specific classes and may also be hurt, though to a much lesser extent, when placing him in high achieving domain-unrelated classrooms or schools.

Our study also comes with further implications for educational practice. For example, we found track-level effects to be more pronounced in math classes that contain students from more than one math track. As track-level effects on academic self-concept negatively affect low achievers and have the opposite effect for high achievers, these results remind practitioners to carefully think about the arrangement of learning environments. We also found frame-of-reference effects on grades on the math class level and to a weaker extent on the school level. This finding can be interpreted to mean that grades might not only be class-referenced but also be school-referenced. Thus, making grades a more valid instrument for student assessment requires better coordination not only between teachers but also between respective schools.

Although there is compelling evidence that, in general, a high self-concept is beneficial for students, whereas a low self-concept is rather harmful (e.g., Marsh et al., 2016), it is important to mention that a low but realistic self-concept may have advantages over a high but unrealistic self-concept. For example, compared to an unrealistically high self-concept, a low but realistic self-concept may positively impact academic decision making, such as choosing a more suitable program of study, and decrease dropout. More research is needed on the importance of academic self-concept for students' academic careers.

**Conclusion**

The present study was aimed at testing predictions from local dominance theory by taking a closer look at the pivotal frames-of-reference for academic self-concept formation in course-by-course tracked school systems. More specifically, we were interested in juxtaposing multiple class environments as frames-of-reference for academic self-concept formation. Data from a comprehensive survey that measured the entire population of Austrian eighth-grade students without special educational needs were well-suited for addressing our research question as general secondary school students were tracked in the core subjects (i.e., mathematics, German, and English) according to ability, whereas regular class composition was the same in all other (non-tracked) subjects. We found math class achievement and to a weaker extent regular class achievement to negatively affect math self-concept, when controlling for achievement on all levels of student nesting. Our finding is in line with local dominance theory and suggests the more proximal domain-specific and to a lower extent the domain-unrelated environments to be frames-of-reference for academic self-concept formation.

**References**


Abu-Hilal, M. M., Abdelfattah, F. A., Alshumrani, S. A., Abduljabbar, A. S., & Marsh, H. W. (2013). Construct validity of self-concept in TIMSS's student background questionnaire: A test of separation and conflation of cognitive and affective dimensions of self-concept among Saudi eighth graders. *European Journal of Psychology of Education*, *28*(4), 1201–1220. https://doi.org/10.1007/s10212-012-0162-1

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, *31*(2-3), 162–172. https://doi.org/10.1016/j.stueduc.2005.05.008

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, *21*(2), 174–177. https://doi.org/10.1177/0956797609357718

Ammermueller, A., & Pischke, J.-S. (2009). Peer effects in european primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, *27*(3), 315–348. https://doi.org/10.1086/603650

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Beretvas, S. N. (2011). Cross-classified and multiplemembership models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 313–334). Routledge.

BIFIE. (2012). *Schülerfragebogen Standardüberprüfung 8. Schulstufe 2012. Gedruckte Version.* BIFIE.

BIFIE. (2016). *Datensatz zur Standardüberprüfung 2012 Mathematik, 8. Schulstufe. Schülerebene, M812I. Forschungsdatenbibliothek (FDB). Nicht-imputierter Datensatz, v2.0.* BIFIE.

Bruneforth, M., Chabera, B., Vogtenhuber, S., & Lassnigg, L. (2016). *ECD review of policies to improve the effectiveness of resource use in schools. Country background report for Austria*. Bundesministerium für Bildung und Frauen. https://doi.org/10.1787/9789264256729-en

Buckingham, J. T., & Alicke, M. D. (2002). The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. *Journal of Personality and Social Psychology*, *83*(5), 1117–1130. https://doi.org/10.1037/0022-3514.83.5.1117

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51–82. https://doi.org/10.1086/666653

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Chmielewski, A. K. (2014). An international comparison of achievement inequality in within- and between-school tracking systems. *American Journal of Education*, *120*(3), 293–324. https://doi.org/10.1086/675529

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type. *American Educational Research Journal*, *50*(5), 925–957. https://doi.org/10.3102/0002831213489843

Choi, Seung W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations.

Cialdini, R. B., & Richardson, K. D. (1980). Two indirect tactics of image management: Basking and blasting. *Journal of Personality and Social Psychology*, *39*(3), 406–415. https://doi.org/10.1037/0022-3514.39.3.406

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques. Difdetect and difwithpar. *Medical Care*, *44*, 115-123. https://doi.org/10.1097/01.mlr.0000245183.28384.ed

Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., Hays, R. D., & Teresi, J. A. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *16 Suppl 1*, 69–84. https://doi.org/10.1007/s11136-007-9185-5

Crane, P. K., van Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine*, *23*(2), 241–256. https://doi.org/10.1002/sim.1713

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*, *110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Eder, F., Altrichter, H., Hofmann, F., & Weber, C. (Eds.). (2015). *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten.* Leykam.

Enders, C. K., & 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

Festinger, L. (1957). A theory of social comparison processes. *Human Relations*, *7*, 117–140. https://doi.org/10.1177/001872675400700202

Frank, R. H. (2011). *The Darwin economy: Liberty, competition, and the common good*. Princeton University Press.

Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W., & Trautwein, U. (2018). It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychological Science*, 1-12. https://doi.org/10.1177/0956797618794454

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data. *Learning and Instruction*, *58*, 210–219.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge Taylor & Francis Group.

Hübner, N., Wagner, W., Hochweber, J., Neumann, M., & Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *Journal of Educational Psychology*, *112*(1), 204–220. https://doi.org/10.1037/edu0000351

Hübner, N., Wille, E., Cambria, J., Oschatz, K., Nagengast, B., & Trautwein, U. (2017). Maximizing gender equality by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *Journal of Educational Psychology*, *109*(7), 993–1009. https://doi.org/10.1037/edu0000183

Huguet, P., Dumas, F., Marsh, H. W., Wheeler, L., Seaton, M., Nezlek, J., Suls, J., & Régner, I. (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

Hulleman, C. S. (2016). Student motivation: Current theories, constructs, and interventions within an expectancy-value framework. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial Skills and School Systems in the 21st Century.* Springer International Publishing.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, *14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

Kim, S.-H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF Detection and Effect Size Measures for Polytomously Scored Items. *Journal of Educational Measurement*, *44*(2), 93–116. https://doi.org/10.1111/j.1745-3984.2007.00029.x

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung, A. S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming. *American Educational Research Journal*, *50*(2), 326–370. https://doi.org/10.3102/0002831212464511

Loveless, T. (2013). *The 2013 Brown Center report on American education: How well are American students learning?* Brookings Institution.

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology*, *30*(3), 263–285. https://doi.org/10.1016/j.cedpsych.2004.10.002

Marsh, H. W. (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. W. (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. https://doi.org/10.2307/1162948

Marsh, H. W., Abduljabbar, A. S., Morin, A. J., Parker, P. D., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*(1), 258–271. https://doi.org/10.1037/a0037485

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics. *Journal of Cross-Cultural Psychology*, *45*(5), 777–804. https://doi.org/10.1177/0022022113519858

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of east and west german school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, *38*(2), 321–350. https://doi.org/10.3102/00028312038002321

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037//0022-3514.78.2.337

Marsh, H. W., Kuyper, H., Morin, A. J., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Martin, A. J., Yeung, A. S., & Craven, R. (2016). Competence self-perceptions. In C. Dweck & D. Yaeger (Eds.), *Handbook of competence and motivation.* Guilford Press.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human cross-cultural generalizability over 68 countries. *European Journal of Personality*, *3*(2), 180–202. https://doi.org/10.1002/per.2232

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., 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. W., & Seaton, M. (2015). The big-fish–little-pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. *Advances in Motivation Science*, *2*, 127–184.

Mullis, I.V.S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015. International results in mathematics*. IEA.

Nagengast, B., & Marsh, H. W. (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

Neumann, M., Trautwein, U., & Nagy, G. (2011). Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the university entrance qualification in Germany. *Studies in Educational Evaluation*, *37*(4), 206–217. https://doi.org/10.1016/j.stueduc.2012.02.002

OECD. (2017a). *Pisa 2015 assessment and analytical framework. Science, reading, mathematic, financial literacy and collaborative problem solving*. OECD.

OECD. (2017b). *PISA 2015 Technical Report*. OECD.

R Core Team. (2008). *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing.

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

Robitzsch, A., Grund, S., & Henke, T. (2018). *Miceadds: Some additional multiple imputation functions, especially for mice.*

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110–116. https://doi.org/10.1016/j.lindif.2015.07.009

Schreiner, C., & Breit, S. (2012). *Standardüberprüfung 2012 Mathematik, 8. Schulstufe. Bundesergebnisbericht*. BIFIE.

Seaton, M., Marsh, H. W., & Craven, R. G. (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

Seaton, M., Marsh, H. W., Yeung, A. S., & Craven, R. (2011). The big fish down under: Examining moderators of the ‘big-fish little-pond’ effect for Australia's high achievers. *Australian Journal of Education*, *55*(2), 93–114. https://doi.org/10.1177/000494411105500202

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Snyder, C. R., Lassegard, M., & Ford, C. E. (1986). Distancing after group success and failure: Basking in reflected glory and cutting off reflected failure. *Journal of Personality and Social Psychology*, *51*(2), 382–388. https://doi.org/10.1037/0022-3514.51.2.382

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish’s achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology*, *109*(2), 191–207. https://doi.org/10.1037/edu0000135

Sticca, F., Goetz, T., Bieg, M., Hall, N. C., Eberle, F., & Haag, L. (2017). Examining the accuracy of students' self-reported academic grades from a correlational and a discrepancy perspective: Evidence from a longitudinal study: Evidence from a longitudinal study. *PloS One*, *12*, 1-13. https://doi.org/10.1371/journal.pone.0187367

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: Consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, *26*(1), 75–101. https://doi.org/10.1080/09243453.2013.871302

Trautwein, U., Gerlach, E., & Lüdtke, O. (2008). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of reference effects. *Journal of Educational Psychology*, *100*(4), 988–1001. https://doi.org/10.1037/0022-0663.100.4.988

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

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*(4), 853–866. https://doi.org/10.1037/a0016306

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial Skills and School Systems in the 21st Century* (pp. 187–214). Springer International Publishing.

von Davier, M. von, Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Vol. 2. Issues and methodologies in large scale assessments* (pp. 9–36). IEA-ETS Research Institute.

Wouters, S., Colpin, H., van Damme, J., & Verschueren, K. (2013). Endorsing achievement goals exacerbates the big-fish-little-pond effect on academic self-concept. *Educational Psychology*, *35*(2), 252–270. https://doi.org/10.1080/01443410.2013.822963

Wouters, S., Fraine, B. de, Colpin, H., van Damme, J., & Verschueren, K. (2012). The effect of track changes on the development of academic self-concept in high school: A dynamic test of the big-fish–little-pond effect. *Journal of Educational Psychology*, *104*(3), 793–805. https://doi.org/10.1037/a0027732

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2-3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005

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

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review*, *14*(4), 368–384. https://doi.org/10.1177/1088868310366144

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Dierential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Directorate of Human Resources Research and Evaluation.

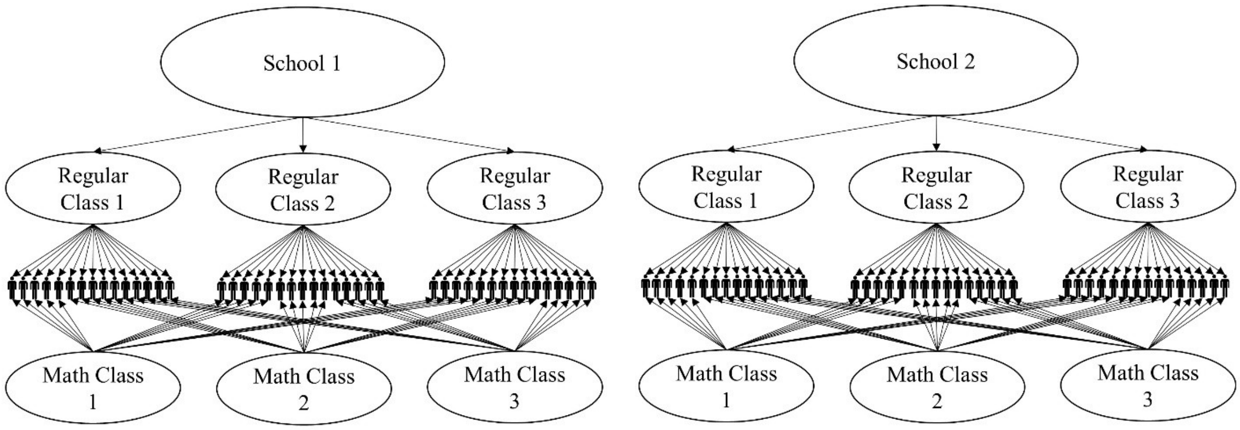
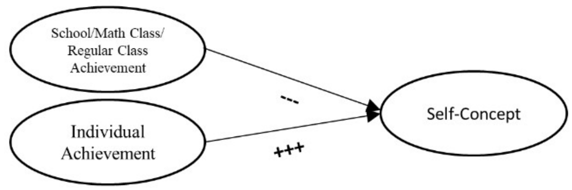
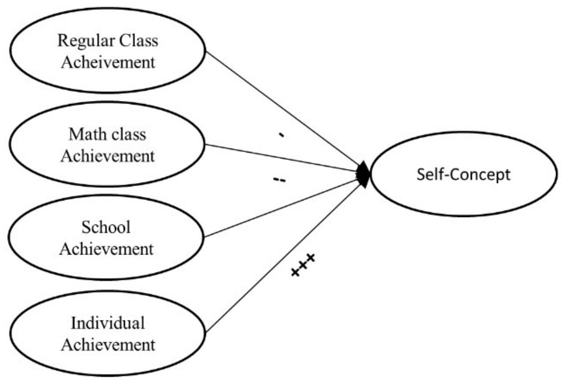



*Figure 1*. Graphical illustration of the cross-classified data structure (exemplary for two schools).

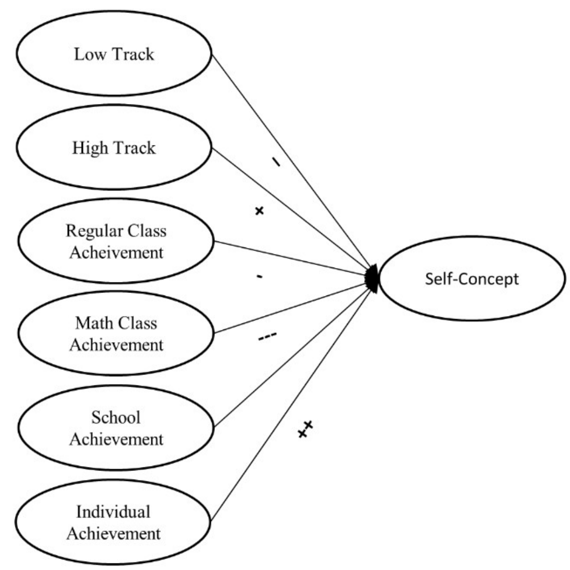
a) H1: Separate BFLPEs




b) H2: Pivotal Frames-of-Reference




c) H3: Track Level and the BFLPE




d) H4: School Grades and the BFLPE

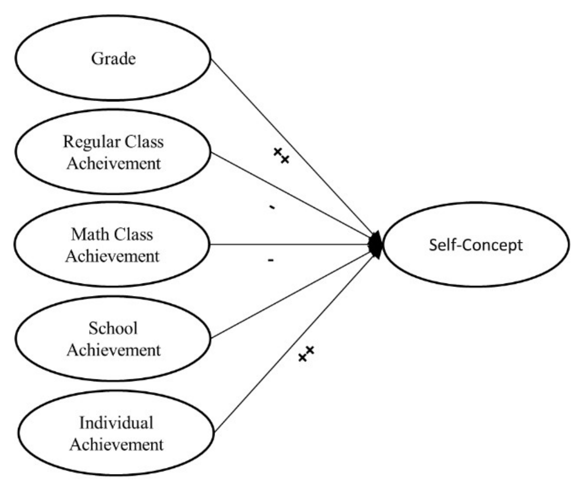



*Figure 2.* Graphical illustration of statistical models testing the study hypotheses. All variables refer to the domain of mathematics. +/++/+++ represent expected positive effects in different strengths, -/--/--- represent expected negative effects in different strengths.

Table 1

*Descriptive Statistics of Model Variables*

| | Mis | *M* | *SD* | $VP_{ind}$ | $VP_{mcl}$ | $VP_{rcl}$ | $VP_{sch}$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Self-concept | .01 | 2.97 | 0.76 | .77 | .18 | .00 | .05 | | | | | |
| 2. Individual achievement | .00 | 504.24 | 86.85 | .38 | .43 | .06 | .14 | .39 | | | | |
| 3. Math class achievement | .00 | 504.24 | 67.58 | | | | | .16 | .78 | | | |
| 4. Regular class achievement | .00 | 504.24 | 49.74 | | | | | .09 | .57 | .70 | | |
| 5. School achievement | .00 | 504.24 | 44.09 | | | | | .06 | .51 | .65 | .89 | |
| 6. Grade | .02 | 3.19 | 0.94 | .63 | .19 | .07 | .10 | .51 | .39 | .20 | .15 | .12 |

*Note.* All variables refer to the domain of mathematics. Variables 1 to 5 are in their original metric. Grade is reverse coded in that higher values indicate better grades. Descriptive statistics were calculated using a complete case analysis approach. Variance proportions were estimated using random intercept models that modeled all levels of student nesting: students (*VPind*), math class (*VPmcl*), regular class (*VPrcl*), and school (*VPsch*). Mis = percent missing.

Table 2

*Pivotal Frames-of-Reference for Academic Self-Concept Formation*

| | Model 0 | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | | | | | | | | | | |
| Individual achievement | | .55 | .01 | [.54,.56] | <.001 | .67 | .01 | [.66,.68] | <.001 | .56 | .01 | [.55,.57] | <.001 | .67 | .01 | [.66,.69] | <.001 |
| School achievement | | -.42 | .01 | [-.45,-.39] | <.001 | | | | | | | | | -.06 | .02 | [-.10,-.02] | .004 |
| Math class achievement | | | | | | -.43 | .01 | [-.45,-.41] | <.001 | | | | | -.37 | .01 | [-.39,-.35] | <.001 |
| Regular class achievement | | | | | | | | | | -.37 | .01 | [-.39,-.35] | <.001 | -.11 | .02 | [-.15,-.08] | <.001 |
| *Random Effect Variances* | | | | | | | | | | | | | | | | | |
| Individual | .97 | .87 | | | | .87 | | | | .87 | | | | .87 | | | |
| Math class | .22 | .24 | | | | .15 | | | | .23 | | | | .15 | | | |
| Regular class | .00 | .08 | | | | .08 | | | | .06 | | | | .08 | | | |
| School | .07 | .03 | | | | .12 | | | | .06 | | | | .10 | | | |
| *Variance Proportions* | | | | | | | | | | | | | | | | | |
| Individual | .77 | .71 | | | | .71 | | | | .71 | | | | .72 | | | |
| Math class | .18 | .20 | | | | .12 | | | | .19 | | | | .13 | | | |
| Regular class | .00 | .06 | | | | .07 | | | | .05 | | | | .06 | | | |
| School | .05 | .02 | | | | .10 | | | | .05 | | | | .09 | | | |

*Note.* $N = 49{,}625$. All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

Table 3

*Pivotal Frames-of-Reference for Academic Self-Concept Formation With Math Grade and Track Level*

| | Model 5 ($N$ = 46,078) | | | | Model 6 ($N$ = 48,978) | | | |
|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | |
| Individual achievement | .61 | .01 | [.60, .63] | <.001 | .45 | .01 | [.44, .46] | <.001 |
| School achievement | .05 | .02 | [.01, .09] | .024 | -.02 | .02 | [-.06, .02] | .297 |
| Math class achievement | -.53 | .01 | [-.56, -.51] | <.001 | -.26 | .01 | [-.28, -.24] | <.001 |
| Regular class achievement | -.09 | .02 | [-.13, -.06] | <.001 | -.14 | .02 | [-.18, -.11] | <.001 |
| Low track | -.32 | .01 | [-.35, -.30] | <.001 | | | | |
| High track | .19 | .01 | [.17, .22] | <.001 | | | | |
| Grade | | | | | .39 | .00 | [.38, .40 | <.001 |
| *Random Effect Variances* | | | | | | | | |
| Individual | .86 | | | | .80 | | | |
| Math class | .15 | | | | .13 | | | |
| Regular class | .08 | | | | .07 | | | |
| School | .10 | | | | .10 | | | |
| *Variance Proportions* | | | | | | | | |
| Individual | .72 | | | | .73 | | | |
| Math class | .13 | | | | .12 | | | |
| Regular class | .07 | | | | .07 | | | |
| School | .08 | | | | .09 | | | |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track variables are dummy variables with reference category medium track. Because of the complete case analysis approach, *N*s differed slightly for the statistical models. CI = 95% confidence interval.

Table 4

*Pivotal Frames-of-Reference for Academic Self-Concept Formation: Differences Between Students from Pure and Mixed Math Classes*

| | *B* | *SE* | CI | *p* |
|---|---|---|---|---|
| *Fixed Effects* | | | | |
| Mixed | -.04 | .02 | [-.08, .00] | .082 |
| Achievement | .69 | .01 | [.67, .71] | <.001 |
| School achievement | .00 | .04 | [-.07, .08] | .941 |
| Math class achievement | -.52 | .03 | [-.58, -.45] | <.001 |
| Regular class achievement | -.08 | .03 | [-.13, -.02] | .006 |
| Low track | -.19 | .03 | [-.26, -.13] | <.001 |
| High track | .09 | .04 | [.02, .16] | .009 |
| Mixed x achievement | -.14 | .01 | [-.17, -.11] | <.001 |
| Mixed x school achievement | .06 | .05 | [-.03, .16] | .199 |
| Mixed x math class achievement | .03 | .04 | [-.05, .10] | .499 |
| Mixed x regular class achievement | -.05 | .04 | [-.12, .03] | .220 |
| Mixed x low track | -.19 | .04 | [-.26, -.11] | <.001 |
| Mixed x high track | .14 | .04 | [.06, .22] | <.001 |
| *Random Effects Variances* | | | | |
| Individual | .86 | | | |
| Math class | .15 | | | |
| Regular class | .08 | | | |
| School | .10 | | | |
| *Variance Proportions* | | | | |
| Individual | .72 | | | |
| Math class | .13 | | | |
| Regular class | .07 | | | |

| | |
|---|---|
| School | .08 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. “Mixed” is a dummy variable indicating if students belong to pure (students from one math track; value 0) or mixed (students from several math tracks; value 1) math classes. The track-level variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table 5

*Pivotal Frames-of-Reference for Grade Provision*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | | | | | | | | | |
| Individual achievement | .49 | .01 | [.48,.51] | <.001 | .57 | .01 | [.56,.59] | <.001 | .49 | .01 | [.48,.50] | <.001 | .57 | .01 | [.56,.59] | <.001 |
| School achievement | -.25 | .02 | [-.28,-.22] | <.001 | | | | | | | | | -.10 | .03 | [-.15,-.06] | <.001 |
| Math class achievement | | | | | -.30 | .01 | [-.32,-.28] | <.001 | | | | | -.30 | .01 | [-.32,-.28] | <.001 |
| Regular class achievement | | | | | | | | | -.15 | .01 | [-.18,-.13] | <.001 | .07 | .02 | [.03,.11] | .001 |
| *Random Effects Variances* | | | | | | | | | | | | | | | | |
| Individual | .86 | | | | .86 | | | | .86 | | | | .86 | | | |
| Math class | .27 | | | | .23 | | | | .27 | | | | .23 | | | |
| Regular class | .08 | | | | .10 | | | | .10 | | | | .10 | | | |
| School | .15 | | | | .17 | | | | .15 | | | | .17 | | | |
| *Variance Proportions* | | | | | | | | | | | | | | | | |
| Individual | .63 | | | | .64 | | | | .63 | | | | .64 | | | |
| Math class | .20 | | | | .17 | | | | .19 | | | | .17 | | | |
| Regular class | .06 | | | | .08 | | | | .07 | | | | .07 | | | |
| School | .11 | | | | .12 | | | | .11 | | | | .12 | | | |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

**Supplemental material**

Table S1

*Descriptive Statistics for Different Math Tracks*

| | Low track | | | | Medium track | | | | High track | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | 1 | 2 | *M* | *SD* | 1 | 2 | *M* | *SD* | 1 | 2 |
| 1. Self-concept | 2.55 | 0.79 | | | 2.91 | 0.73 | | | 3.16 | 0.70 | | |
| 2. Achievement | 414.02 | 56.06 | .25 | | 477.38 | 59.10 | .24 | | 564.77 | 69.50 | .32 | |
| 3. Grade | 2.69 | 0.88 | .41 | .19 | 3.06 | 0.81 | .41 | .17 | 3.46 | 0.96 | .51 | .32 |

*Note*. All variables refer to the domain of math. Variables 1 to 2 are in their original metric. Grade is reverse coded in that higher values indicate better grades. Descriptive statistics were calculated using a complete case analysis approach..

Table S2

*Pivotal Frames-of-Reference for Academic Self-Concept Formation Without Self-Concept Item That Refers to Social Comparison*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| Individual achievement | .53 | .01 | [.52, .54] | <.001 | .64 | .01 | [.63, .66] | <.001 | .54 | .01 | [.53, .55] | <.001 | .65 | .01 | [.63, .66] | <.001 | .59 | .01 | [.57 .60] | <.001 | .43 | .01 | [.41, .44] | <.001 |
| School achievement | -.41 | .01 | [-.43, -.38] | <.001 | | | | | | | | | -.08 | .02 | [-.13, -.04] | <.001 | .03 | .02 | [-.01 .08] | .162 | -.04 | .02 | [-.08, .00] | .042 |
| Math class achievement | | | | | -.41 | .01 | [-.43, -.39] | <.001 | | | | | -.35 | .01 | [-.37, -.33] | <.001 | -.52 | .01 | [-.54 -.49] | <.001 | -.24 | .01 | [-.26, -.22] | <.001 |
| Regular class achievement | | | | | | | | | -.35 | .01 | [-.38, -.33] | <.001 | -.10 | .02 | [-.13, -.06] | <.001 | -.07 | .02 | [-.11 -.04] | <.001 | -.13 | .02 | [-.16, -.09] | <.001 |
| Low track | | | | | | | | | | | | | | | | | -.31 | .01 | [-.33 -.28] | <.001 | | | | |
| High track | | | | | | | | | | | | | | | | | .21 | .01 | [.18 .24] | <.001 | | | | |
| Grade | | | | | | | | | | | | | | | | | | | | | .39 | .00 | [.38, .40] | <.001 |

*Note.* The self-concept item that refers to social comparison is *Mathematics is harder for me than for many of my classmates*. All variables refer to the domain of math. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S3

*Pivotal Frames-of-Reference for Academic Self-Concept Formation Controlling for Covariates*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| Sex | -.35 | .01 | [-.36, -.33] | <.001 | -.33 | .01 | [-.35, -.32] | <.001 | -.35 | .01 | [-.36, -.33] | <.001 | -.33 | .01 | [-.35, -.32] | <.001 | -.34 | .01 | [-.36, -.32] | <.001 | -.41 | .01 | [-.42, -.39] | <.001 |
| Age | .00 | .00 | [-.01, .01] | .552 | -.01 | .00 | [-.02, .00] | .008 | .00 | .00 | [-.01, .00] | .353 | -.01 | .00 | [-.02, .00] | .006 | .00 | .00 | [-.01, .01] | .617 | .01 | .00 | [.00, .01] | .133 |
| SES | -.01 | .00 | [-.02, .00] | .045 | .00 | .00 | [-.01, .01] | .605 | -.01 | .00 | [-.01, .00] | .188 | .00 | .00 | [-.01, .01] | .837 | -.01 | .00 | [-.02, .00] | .045 | -.01 | .00 | [-.02, -.01] | <.001 |
| Migration | .20 | .01 | [.17, .22] | <.001 | .21 | .01 | [.18, .23] | <.001 | .20 | .01 | [.18, .23] | <.001 | .19 | .01 | [.16, .21] | <.001 | .20 | .01 | [.18, .23] | <.001 | .15 | .01 | [.13, .17] | <.001 |
| Individual achievement | .54 | .01 | [.53, .55] | <.001 | .65 | .01 | [.64, .67] | <.001 | .55 | .01 | [.54, .56] | <.001 | .65 | .01 | [.64, .67] | <.001 | .59 | .01 | [.58, .61] | <.001 | .41 | .01 | [.40, .43] | <.001 |
| School achievement | -.34 | .01 | [-.37, -.31] | <.001 | | | | | | | | | -.04 | .02 | [-.08, .01] | .101 | .09 | .02 | [.05, .14] | <.001 | .00 | .02 | [-.04, .04] | .990 |
| Math class achievement | | | | | -.39 | .01 | [-.41, -.37] | <.001 | | | | | -.35 | .01 | [-.37, -.33] | <.001 | -.52 | .01 | [-.55, -.49] | <.001 | -.21 | .01 | [-.23, -.19] | <.001 |
| Regular class achievement | | | | | | | | | -.30 | .01 | [-.33, -.28] | <.001 | -.09 | .02 | [-.13, -.05] | <.001 | -.06 | .02 | [-.10, -.02] | .002 | -.12 | .02 | [-.15, -.08] | <.001 |
| Low track | | | | | | | | | | | | | | | | | -.34 | .01 | [-.37, -.31] | <.001 | | | | |
| High track | | | | | | | | | | | | | | | | | .22 | .01 | [.19, .24] | <.001 | | | | |
| Grade | | | | | | | | | | | | | | | | | | | | | .41 | .00 | [.41, .42] | <.001 |

*Note.* All variables refer to the domain of math. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. Sex is a dummy variable with 0 for male and 1 for female. Migration is a dummy variable with 0 for no migration background and 1 for migration background. The track variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S4

*Pivotal Frames-of-Reference for Academic Self-Concept Formation With Math Class Average Track Level*

| | *B* | *SE* | CI | *p* |
|---|---|---|---|---|
| Individual achievement | .67 | .01 | [.66, .68] | <.001 |
| School achievement | .02 | .02 | [-.02, .07] | .375 |
| Math class achievement | -.53 | .02 | [-.57, -.49] | <.001 |
| Regular class achievement | -.09 | .02 | [-.13, -.06] | <.001 |
| Math class average track level | .17 | .02 | [.13, .21] | <.001 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

Table S5

*Track Interaction Model*

| | *B* | *SE* | CI | *p* |
|---|---|---|---|---|
| Achievement | .55 | .01 | [.53, .58] | <.001 |
| School achievement | .08 | .03 | [.01, .15] | .022 |
| Math class achievement | -.44 | .02 | [-.49, -.39] | <.001 |
| Regular class achievement | -.15 | .03 | [-.21, -.09] | <.001 |
| Low track | -.34 | .03 | [-.40, -.29] | <.001 |
| High track | .15 | .02 | [.12, .18] | <.001 |
| Achievement x Low track | .00 | .02 | [-.05, .04] | .848 |
| Achievement x High track | .11 | .02 | [.08, .15] | <.001 |
| School achievement x Low track | .17 | .05 | [.06, .27] | .002 |
| School achievement x High track | -.20 | .05 | [-.29, -.11] | <.001 |
| Math class achievement x Low track | -.06 | .04 | [-.13, .02] | .126 |
| Math class achievement x High track | -.09 | .03 | [-.15, -.02] | .010 |
| Regular class achievement x Low track | .01 | .05 | [-.09, .11] | .885 |
| Regular class achievement x High track | .11 | .04 | [.02, .19] | .010 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S6

*Pivotal Frames-of-Reference for Academic Self-Concept Formation with Only the Social Comparison Item as the Outcome*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p |
| Individual achievement | .39 | .01 | [.38, | <.001 | .51 | .01 | [.49, .52] | <.001 | .41 | .01 | [.40, .42] | <.001 | .51 | .01 | [.50, .53] | <.001 | .47 | .01 | [.46, .49] | <.001 | .36 | .01 | [.34, .37] | <.001 |
| School achievement | -.30 | .01 | [-.32, -.28] | <.001 | | | | | | | | | .00 | .02 | [-.04, .04] | .983 | .07 | .02 | [.03, .11] | .001 | .03 | .02 | -[.01, .07] | .105 |
| Math class achievement | | | | | -.35 | .01 | [-.37, -.33] | <.001 | | | | | -.30 | .01 | [-.32, -.28] | <.001 | -.40 | .01 | [-.42, -.37] | <.001 | -.22 | .01 | [.24, -.20] | <.001 |
| Regular class achievement | | | | | | | | | -.29 | .01 | [-.31, -.27] | <.001 | -.12 | .02 | [-.15, -.08] | <.001 | -.11 | .02 | [-.14, -.07] | <.001 | -.14 | .02 | [-.17, -.11] | <.001 |
| Low track | | | | | | | | | | | | | | | | | -.25 | .01 | [-.28, -.22] | <.001 | | | | |
| High track | | | | | | | | | | | | | | | | | .09 | .01 | [.06, .12] | <.001 | | | | |
| Grade | | | | | | | | | | | | | | | | | | | | | .27 | .00 | [.26, .28] | <.001 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track-level variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S7

*Pivotal Frames-of-Reference for Academic Self-Concept Formation with Multiple Imputation*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p |
| Individual achievement | .55 | .01 | [.54, | <.001 | .67 | .01 | [.65, .68] | <.001 | .56 | .01 | [.55, .57] | <.001 | .67 | .01 | [.66, .69] | <.001 | .62 | .01 | [.61, .64] | <.001 | .48 | .01 | [.46, .49] | <.001 |
| School achievement | -.42 | .01 | [-.45, -.40] | <.001 | | | | | | | | | -.06 | .02 | [-.11, -.02] | .004 | .01 | .02 | [-.03, .06] | .564 | -.03 | .02 | [-.07, .01] | .204 |
| Math class achievement | | | | | -.43 | .01 | [-.45, -.41] | <.001 | | | | | -.37 | .01 | [-.40, -.35] | <.001 | -.49 | .01 | [-.52, -.47] | <.001 | -.27 | .01 | [-.29, -.25] | <.001 |
| Regular class achievement | | | | | | | | | -.37 | .01 | [-.40, -.35] | <.001 | -.12 | .02 | [-.15, -.08] | <.001 | -.09 | .02 | [-.13, -.05] | <.001 | -.14 | .02 | [-.17, -.11] | <.001 |
| Low track | | | | | | | | | | | | | | | | | -.27 | .01 | [-.29, -.24] | <.001 | | | | |
| High track | | | | | | | | | | | | | | | | | .14 | .01 | [.11, .17] | <.001 | | | | |
| Grade | | | | | | | | | | | | | | | | | | | | | .36 | .00 | [.36, .37] | <.001 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track-level variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S8

*Pivotal Frames-of-Reference for Academic Self-Concept Formation with the Pure Student Sample*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p | B | SE | CI | p |
| Individual achievement | .53 | .01 | [.51, | <.001 | .68 | .01 | [.66, .70] | <.001 | .53 | .01 | [.52, .55] | <.001 | .68 | .01 | [.66, .70] | <.001 | .69 | .01 | [.67, .71] | <.001 | .42 | .01 | [.40, .44] | <.001 |
| School achievement | -.40 | .03 | [-.45, -.35] | <.001 | | | | | | | | | -.07 | .03 | [-.14, -.01] | .033 | .00 | .04 | [-.08, .08] | .968 | -.03 | .03 | [-.09, .03] | .362 |
| Math class achievement | | | | | -.43 | .01 | [-.46, -.40] | <.001 | | | | | -.39 | .02 | [-.42, -.36] | <.001 | -.51 | .03 | [-.58, -.44] | .000 | -.25 | .01 | [-.28, -.22] | <.001 |
| Regular class achievement | | | | | | | | | -.30 | .02 | [-.34, -.27] | <.001 | -.08 | .03 | [-.13, -.03] | .002 | -.08 | .03 | [-.13, -.03] | .004 | -.11 | .02 | [-.16, -.06] | <.001 |
| Low track | | | | | | | | | | | | | | | | | -.19 | .03 | [-.25, -.12] | <.001 | | | | |
| High track | | | | | | | | | | | | | | | | | .09 | .04 | [.01, .16] | .019 | | | | |
| Grade | | | | | | | | | | | | | | | | | | | | | .41 | .01 | [.39, .42] | <.001 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track-level variables are dummy variables with reference category medium track. CI = 95% confidence interval.

Table S9

*Pivotal Frames-of-Reference for Academic Self-Concept Formation with Moderators*

| | Sex | | | | Age | | | | Migration | | | | SES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| Moderator | -.33 | .01 | [-.35, -.32] | <.001 | .01 | .00 | [.00, .02] | .008 | .19 | .01 | [.16, .21] | <.001 | -.01 | .00 | [-.02, .00] | .094 |
| Achievement | .65 | .01 | [.63, .66] | <.001 | .67 | .01 | [.66, .69] | <.001 | .68 | .01 | [.67, .69] | <.001 | .68 | .01 | [.66, .69] | <.001 |
| School achievement | -.06 | .03 | [-.11, -.01] | .031 | -.07 | .02 | [-.11, -.02] | .002 | -.02 | .02 | [-.07, .03] | .426 | -.07 | .02 | [-.11, -.02] | .003 |
| Math class achievement | -.34 | .01 | [-.37, -.32] | <.001 | -.37 | .01 | [-.39, -.35] | <.001 | -.37 | .01 | [-.39, -.34] | <.001 | -.38 | .01 | [-.40, -.35] | <.001 |
| Regular class achievement | -.12 | .02 | [-.17, -.08] | <.001 | -.11 | .02 | [-.15, -.08] | <.001 | -.11 | .02 | [-.15, -.07] | <.001 | -.11 | .02 | [-.14, -.07] | <.001 |
| Moderator x school achievement | -.02 | .04 | [-.09, .05] | .529 | .03 | .02 | [-.01, .06] | .142 | -.02 | .05 | [-.12, .07] | .612 | -.01 | .02 | [-.05, .02] | .507 |
| Moderator x math class achievement | -.01 | .01 | [-.04, .02] | .475 | -.01 | .01 | [-.02, .01] | .272 | .00 | .02 | [-.05, .04] | .905 | .01 | .01 | [.00, .03] | .070 |
| Moderator x regular class achievement | .04 | .03 | [-.03, .11] | .249 | .00 | .02 | [-.03, .03] | .900 | .01 | .05 | [-.08, .10] | .871 | .02 | .02 | [-.02, .05] | .312 |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. Sex is a dummy variable with 0 for male and 1 for female. Migration is a dummy variable with 0 for no migration background and 1 for migration background. The column names indicate the moderator under investigation. CI = 95% confidence interval.

Table S10

*Correlations Using a Within-Cluster Centering (CWC) Approach*

| | 1 | 2 |
|---|---|---|
| *Individual level* | | |
| 1. Self-concept | | |
| 2. Achievement | .38 | |
| 3. Grade | .50 | .37 |
| *Math class level* | | |
| 1. Self-concept | | |
| 2. Achievement | .58 | |
| 3. Grade | .60 | .53 |
| *Regular class level* | | |
| 1. Self-concept | | |
| 2. Achievement | .35 | |
| 3. Grade | .49 | .48 |
| *School level* | | |
| 1. Self-concept | | |
| 2. Achievement | .30 | |
| 3. Grade | .49 | .41 |

*Note.* Individual-level variables are centered at the two classenvironments, and class-level variables are centered at the school environment.

Table S11

*Pivotal Frames-of-Reference for Academic Self-Concept Formation using a Within Custer Centering (CWC) Approach*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | | | | | | | | | |
| Individual achievement | .63 | .01 | [.62, .64] | <.001 | .59 | .01 | [.58, .61] | <.001 | .62 | .01 | [.61, .63] | <.001 | .67 | .01 | [.66, .69] | <.001 |
| School achievement | .74 | .02 | [.71, .77] | <.001 | | | | | | | | | .79 | .01 | [.77, .82] | <.001 |
| Math class achievement | | | | | .35 | .01 | [.33, .37] | <.001 | | | | | .30 | .01 | [.28, .32] | <.001 |
| Regular class achievement | | | | | | | | | .66 | .02 | [.62, .70] | <.001 | .56 | .02 | [.52, .59] | <.001 |
| *Variance Proportions* | | | | | | | | | | | | | | | | |
| Individual | .87 | | | | .87 | | | | .87 | | | | .87 | | | |
| Math class | .28 | | | | .17 | | | | .26 | | | | .15 | | | |
| Regular class | .18 | | | | .17 | | | | .05 | | | | .08 | | | |
| School | .00 | | | | .36 | | | | .37 | | | | .10 | | | |
| *Random Effect Variances* | | | | | | | | | | | | | | | | |
| Individual | .65 | | | | .55 | | | | .56 | | | | .72 | | | |
| Math class | .21 | | | | .11 | | | | .17 | | | | .13 | | | |
| Regular class | .13 | | | | .11 | | | | .03 | | | | .06 | | | |
| School | .00 | | | | .23 | | | | .24 | | | | .09 | | | |
| *Contextual Effects Estimates* | | | | | | | | | | | | | | | | |
| School BFLPE | | | | | | | | | | | | | -.06 | | | |
| Math class BFLPE | | | | | | | | | | | | | -.37 | | | |
| Regular class BFLPE | | | | | | | | | | | | | -.11 | | | |

*Note.* Individual-level variables are centered at the two class environments, and class-level variables are centered at the school environment.

Table S12

*Descriptive Statistics in the Total Sample*

| | Mis | *M* | *SD* | $VP_{ind}$ | $VP_{cl}$ | $VP_{sch}$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Self-concept | .01 | 2.97 | 0.76 | .78 | .16 | .06 | | | | |
| 2. Individual achievement | .00 | 535.36 | 94.31 | .37 | .33 | .30 | .38 | | | |
| 3. Class achievement | .00 | 535.36 | 74.33 | | | | .12 | .79 | | |
| 4. School achievement | .00 | 535.36 | 60.52 | | | | .05 | .64 | .81 | |
| 5. Grade | .02 | 3.16 | 1.01 | .69 | .19 | .12 | .54 | .36 | .11 | .04 |

*Note.* All variables refer to the domain of mathematics. Variables 1 to 5 are in their original metric. Grade is reverse coded in that higher values indicate better grades. Descriptive statistics were calculated using a complete case analysis approach. Variance proportions were estimated using random intercept models that modeled all levels of student nesting: students (*VPind*), class (*VPcl*), and school (*VPsch*). Mis = percent missing.

Table S13

*Pivotal Frames-of-Reference for Academic Self-Concept Formation in the Total Sample*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| Individual achievement | .67 | .00 | [.66,.68] | <.001 | .76 | .01 | [.75,.77] | <.001 | .76 | .00 | [.75,.77] | <.001 |
| School achievement | -.58 | .01 | [-.60,-.57] | <.001 | | | | | -.22 | .01 | [-.24,-.19] | <.001 |
| Class achievement | | | | | -.56 | .01 | [-.57,-.54] | <.001 | -.45 | .01 | [-.47,-.43] | <.001 |
| *Random Effect Variances* | | | | | | | | | | | | |
| Individual | .85 | | | | .85 | | | | .16 | | | |
| Class | .25 | | | | .17 | | | | .11 | | | |
| School | .05 | | | | .13 | | | | .85 | | | |
| *Variance Proportions* | | | | | | | | | | | | |
| Individual | .74 | | | | .74 | | | | .15 | | | |
| Class | .22 | | | | .15 | | | | .09 | | | |
| School | .04 | | | | .11 | | | | .76 | | | |

*Note*. All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

Table S14

*Descriptive Statistics in the Academic Secondary School Subsample*

| | Mis | *M* | *SD* | $VP_{ind}$ | $VP_{cl}$ | $VP_{sch}$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Self-concept | .01 | 2.99 | 0.77 | .83 | .13 | .04 | | | | |
| 2. Individual achievement | .00 | 599.86 | 74.66 | .55 | .19 | .26 | .51 | | | |
| 3. Class achievement | .00 | 599.86 | 39.17 | | | | .08 | .52 | | |
| 4. School achievement | .00 | 599.86 | 31.38 | | | | .04 | .42 | .80 | |
| 5. Grade | .02 | 3.12 | 1.14 | .73 | .17 | .10 | .61 | .51 | .07 | .04 |

*Note.* All variables refer to the domain of mathematics. Variables 1 to 5 are in their original metric. Grade is reverse coded in that higher values indicate better grades. Descriptive statistics were calculated using a complete case analysis approach. Variance proportions were estimated using random intercept models that modeled all levels of student nesting: students (*VPind*), class (*VPcl*), and school (*VPsch*). Mis = percent missing.

Table S15

*Pivotal Frames-of-Reference for the Academic Secondary School Subsample*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | | | | | |
| Individual achievement | .62 | .01 | [.61,.63] | <.001 | .64 | .01 | [.63,.65] | <.001 | .64 | .01 | [.63,.65] | <.001 |
| School achievement | -.53 | .02 | [-.57-.49] | <.001 | | | | | -.18 | .03 | [-.24,-.13] | <.001 |
| Class achievement | | | | | -.48 | .02 | [-.51,-.45] | <.001 | -.37 | .02 | [-.42,-.32] | <.001 |
| *Random Effect Variances* | | | | | | | | | | | | |
| Individual | .81 | | | | .81 | | | | .81 | | | |
| Class | .21 | | | | .18 | | | | .17 | | | |
| School | .00 | | | | .05 | | | | .05 | | | |
| *Variance Proportions* | | | | | | | | | | | | |
| Individual | .21 | | | | .17 | | | | .17 | | | |
| Class | .21 | | | | .17 | | | | .17 | | | |
| School | .00 | | | | .05 | | | | .05 | | | |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

Table S16

*Pivotal Frames-of-Reference for Academic Self-Concept Formation With IRT-Scaled WLE Scores as Outcome*

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | | | | | | | | | |
| Individual achievement | .54 | .01 | [.53 .55] | <.001 | .66 | .01 | [.64 .67] | <.001 | .55 | .01 | [.54 .56] | <.001 | .66 | .01 | [.65 .67] | <.001 |
| School achievement | -.42 | .01 | [-.45 -.40] | <.001 | | | | | | | | | -.10 | .02 | [-.14 -.05] | <.001 |
| Math class achievement | | | | | -.42 | .01 | [-.44 -.40] | <.001 | | | | | -.36 | .01 | [-.38 -.34] | <.001 |
| Regular class achievement | | | | | | | | | -.36 | .01 | [-.39 -.34] | <.001 | -.09 | .02 | [-.13 -.06] | <.001 |
| *Random Effect Variances* | | | | | | | | | | | | | | | | |
| Individual | .88 | | | | .87 | | | | .88 | | | | .87 | | | |
| Math class | .24 | | | | .15 | | | | .22 | | | | .15 | | | |
| Regular class | .08 | | | | .08 | | | | .07 | | | | .08 | | | |
| School | .05 | | | | .13 | | | | .08 | | | | .11 | | | |
| *Variance Proportions* | | | | | | | | | | | | | | | | |
| Individual | .71 | | | | .71 | | | | .70 | | | | .72 | | | |
| Math class | .19 | | | | .12 | | | | .18 | | | | .12 | | | |
| Regular class | .06 | | | | .07 | | | | .06 | | | | .07 | | | |
| School | .04 | | | | .10 | | | | .06 | | | | .09 | | | |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. CI = 95% confidence interval.

Table S17

*Pivotal Frames-of-Reference for Academic Self-Concept Formation With Math Grade and Track Level With IRT-Scaled WLE Scores as Outcome*

| | Model 5 | | | | Model 6 | | | |
|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | CI | *p* | *B* | *SE* | CI | *p* |
| *Fixed Effects* | | | | | | | | |
| Individual achievement | .60 | .01 | [.59, .61] | <.001 | .44 | .01 | [.42, .45] | <.001 |
| School achievement | .02 | .02 | [-.03, .06] | .411 | -.05 | .02 | [-.09, -.01] | .007 |
| Math class achievement | -.52 | .01 | [-.55, -.50] | <.001 | -.24 | .01 | [-.26, -.22] | <.001 |
| Regular class achievement | -.07 | .02 | [-.11, -.03] | <.001 | -.12 | .02 | [-.16, -.09] | <.001 |
| Low track | -.28 | .01 | [-.31, -.26] | <.001 | | | | |
| High track | .22 | .01 | [.19, .25] | <.001 | | | | |
| Grade | | | | | .39 | .00 | [.38, .40] | <.001 |
| *Random Effect Variances* | | | | | | | | |
| Individual | .86 | | | | .80 | | | |
| Math class | .15 | | | | .13 | | | |
| Regular class | .08 | | | | .07 | | | |
| School | .11 | | | | .11 | | | |
| *Variance Proportions* | | | | | | | | |
| Individual | .72 | | | | .72 | | | |
| Math class | .13 | | | | .11 | | | |
| Regular class | .07 | | | | .07 | | | |
| School | .09 | | | | .09 | | | |

*Note.* All variables refer to the domain of mathematics. Level 1 variables are standardized, and manifest level 2 aggregates are composed of standardized level 1 variables. The track variables are dummy variables with reference category medium track. CI = 95% confidence interval.