Harald Clahsen | Loay Balkhair | John-Sebastian Schutter
Ian Cunnings

The time course of morphological processing in a second language

Tracking Effects Depend on Tracking Type: An International Comparison of Students' Mathematics Self-Concept

Anna K. Chmielewski*
Stanford University
Hanna Dumont*
Ulrich Trautwein
University of Tübingen, Germany

The aim of the present study was to examine how different types of tracking—between-school streaming, within-school streaming, and course-by-course tracking—shape students' mathematics self-concept. This was done in an internationally comparative framework using data from the Programme for International Student Assessment (PISA). After controlling for individual and track mean achievement, results indicated that generally for students in course-by-course tracking, high-track students had higher mathematics self-concepts and low-track students had lower mathematics self-concepts. For students in between-school and within-school streaming, the reverse pattern was found. These findings suggest a solution to the ongoing debate about the effects of tracking on students' academic self-concept and suggest that the reference groups to which students compare themselves differ according to the type of tracking.

ANNA K. CHMIELEWSKI is now a postdoctoral fellow in the Pathways to Adulthood program, based in the College of Education at Michigan State University, Erickson Hall, 620 Farm Lane, East Lansing, MI, USA; e-mail: *chmiel34@msu.edu*. Her research interests include socioeconomic gaps in academic achievement, tracking and ability grouping, and large-scale cross-national assessments.

HANNA DUMONT is now a postdoctoral research associate in the Department of Quantitative Methods in Education at the University of Potsdam, Germany. Her research interests include the effects of schools' social and ethnic composition, tracking and ability grouping, parental involvement, and individualized instruction.

ULRICH TRAUTWEIN is professor of educational science and the director of the LEAD Graduate School at the University of Tübingen, Germany; e-mail: *ulrich.trautwein @uni-tuebingen.de*. His research interests include the development of self-concept and motivation in the school context, school effectiveness, and the influence of homework on school achievement.

Most school systems in developed countries around the world practice some form of tracking. *Tracking*—and similar practices known as ability grouping and streaming—groups students into different courses, study programs, or schools according to their achievement. The practice of tracking can be seen as an attempt to respond to heterogeneity in students' abilities and academic achievement by sorting them into more homogeneous groups, allowing teachers to cater instruction to students' needs. Despite this promising objective, there has been a heated debate about the effectiveness of tracking, both in policy and research (Hallinan, 1994; Ireson & Hallam, 2001; Lucas, 1999). This is mostly due to the argument that tracking increases educational inequality. Indeed, previous research has found that students in lower tracks often have less opportunity to learn because of less demanding curricula, lower instructional quality, lower expectations, and lower levels of other educational resources (Brunello & Checchi, 2007; Kelly, 2004; Lucas, 1999; Oakes, 1985) and that these differences in opportunities can affect achievement (Gamoran & Mare, 1989; Hanushek & Wößmann, 2006). Even more alarming, research has also found that assignment to tracks is often biased by race or social class (Lucas & Berends, 2002; Maaz, Trautwein, Lüdtke, & Baumert, 2008).

Whereas most research that has looked at the consequences of tracking for academic achievement has found detrimental effects of being placed in a low track, controversial findings have emerged for the effects of tracking on students' non-cognitive outcomes, specifically, students' academic self-concept. Some researchers have argued that being placed in a low track will lead to lower self-perceptions (Oakes, 1985). Other researchers, however, have reported positive effects of being placed in a low track (and negative effects of being placed in a high track) on students' academic self-concept (Marsh & Hau, 2003; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007). One possible explanation for these conflicting findings on the relationship between tracking and academic self-concept may be that previous research has taken place in a number of different countries that use a wide range of different tracking policies. These differences are usually not taken into account when studying the effects of tracking. In the present study we therefore systematically compare three different types of tracking with respect to their effects on students' self-concept in the domain of mathematics. We do so in an internationally comparative framework using the Programme for International Student Assessment (PISA) 2003 data set. In this sense, the present study examines *social mechanisms* for the effects of tracking—in contrast with *instructional* and *institutional* mechanisms of tracking (see Lucas, 1999; Pallas, Entwisle, Alexander, & Stluka,

1994)—assuming that tracks serve as reference groups for social comparison in the development of academic self-concept.

Students' Academic Self-Concept

A person's self-concept, which can be broadly defined as the person's perception of himself or herself, consists of several hierarchically ordered dimensions (Byrne & Shavelson, 1986; Marsh & Shavelson, 1985). One of these (sub)dimensions is the *academic* self-concept: a person's perception of his or her abilities and competences. Numerous studies have shown that rather than a global academic self-concept, a person has a distinct self-concept for each academic domain (e.g., a mathematical and a verbal self-concept, among others; Marsh, 1990a, 1990b; Marsh & Shavelson, 1985). The multidimensional, domain-specific nature of academic self-concept has been shown to hold across different cultures (Marsh & Hau, 2004) and accordingly should be taken into account when conducting empirical studies on academic self-concept.

It is important to distinguish academic self-concept from related concepts, such as self-esteem and self-efficacy. Self-esteem, it has been suggested, should be used for the global component of a person's self-concept (Marsh, Craven, & Martin, 2006; Marsh & O'Mara, 2008), as empirical studies show that self-esteem more strongly predicts global outcomes, whereas domain-specific academic self-concepts predict academic outcomes (Marsh & Craven, 2006; Swann, Chang-Schneider, & McClarty, 2007). The other related term, self-efficacy, is more context-specific than academic self-concept: It is a person's self-perceived confidence to successfully perform a particular academic task (Bong & Skaalvik, 2003). Studies show that self-efficacy is indeed a separate, empirically distinguishable construct from academic self-concept, even when studied within the same domain (Ferla, Valcke, & Cai, 2009; Lee, 2009).

Academic self-concept is important for students not only because it is a desirable outcome in its own right but also because numerous studies have shown that it positively affects subsequent academic achievement and other educational outcomes, such as academic motivation, effort, and grades (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Marsh & Yeung, 1997; Valentine, DuBois, & Cooper, 2004). Motivation and effort in turn serve as mediators of the association between academic self-concept and academic achievement, in the sense that students with positive views about their abilities engage more in achievement-related behaviors, such as studying for tests and completing homework (Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009), and show higher levels of concentration on tests (Eckert, Schilling, & Stiensmeier-Pelster, 2006). Moreover, prior research has found that academic self-concept has indirect effects on longer-term outcomes, such as educational aspirations and course selection

(Marsh & Yeung, 1998; Mulkey, Catsambis, Steelman, & Crain, 2005). These findings hold for a variety of academic domains, including mathematics self-concept (Marsh et al., 2005; Martin & Debus, 1998; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006), which is the focus of the present study.

The Importance of Social Comparisons for the Development of Academic Self-Concept

Theories on the development of academic self-concept postulate that "the formation of academic self-concepts requires students to compare their self-perceived academic accomplishments to some standard or frame of reference" (Marsh, Chessor, Craven, & Roche, 1995, p. 290). In other words, a student's academic self-concept is shaped not only by his or her actual performance but also through social comparisons. Two important social comparison mechanisms through which academic self-concept is known to be affected, albeit in different directions, are *contrast* and *assimilation effects* (Marsh et al., 1995; Marsh, Kong, & Hau, 2000).

A contrast effect, sometimes also called a *reference group effect* (Bassis, 1977; Marsh, 1987; Richer, 1976), refers to the finding that being surrounded by students with high abilities leads a student to feel more negatively about his or her own abilities. In other words, the same student will have a lower academic self-concept in a group with high mean achievement and a higher academic self-concept in a group with low mean achievement because he or she will compare and contrast his or her own achievement with that of his classmates or schoolmates. On the other hand, being a member of a group of high-achieving students may also make a student feel positively about his or her own abilities and thus result in an *assimilation effect* (Marsh et al., 1995). In this sense, being placed in a high-achieving group will positively affect a student's academic self-concept because he or she will *bask in the reflected glory* (Cialdini et al., 1976) of his or her high-achieving group members.

The existence of contrast effects has been shown in a large number of studies that found negative associations between a group's mean achievement and a student's academic self-concept, after controlling for the student's own achievement. This result, known as the *big-fish-little-pond effect* (BFLPE) or *frog pond effect*, has been the subject of educational, sociological, and psychological research for more than 40 years (Bassis, 1977; Davis, 1966; Espenshade, Hale, & Chung, 2005; Goldsmith, 2011; Marsh, 1987; Marsh et al., 1995, 2007, 2008; Preckel & Brüll, 2010; Thijs, Verkuyten, & Helmond, 2010). It has been replicated numerous times and found across many culturally and economically diverse countries with only small variation in its size (Marsh & Hau, 2003; Seaton, Marsh, & Craven, 2009). Although there has also been evidence of assimilation effects on students' academic self-concept (Marsh et al., 2000; Preckel &

Brüll, 2010; Trautwein et al., 2005; Trautwein, Lüdtke, Marsh, & Nagy, 2009), the evidence is less consistent especially regarding the size of the effects.

The study of contrast and assimilation effects is complicated by the fact that they occur at the same time, such that observed levels of academic self-concept can be understood as the net effects of these two counterbalancing processes (Marsh et al., 2000). This is particularly relevant for the study of tracking and academic self-concept: Because high track level is typically associated with high track mean achievement, an observed association between track level and academic self-concept is likely to conflate assimilation and contrast effects. Thus, the net association between track level and academic self-concept after controlling for individual achievement depends on which effect, contrast or assimilation, outweighs the other. There is reason to believe that the mixed findings on the effects of tracking on students' self-concept reported in the literature (Oakes, 1985; Marsh et al., 2007) are driven by differences in the underlying contrast and assimilation effects. Whereas contrast effects appear not to differ systematically between different national or cultural contexts (Marsh & Hau, 2003; Seaton et al., 2009), in the present study, we suggest that the presence and size of assimilation effects may depend on the *type of tracking* students experience, as this determines whom students use to evaluate their own abilities, that is, their reference groups. Before discussing this idea in more detail and presenting previous empirical findings on the association between tracking and academic self-concept, we first introduce the three different types of tracking considered in the present study.

Different Types of Tracking

Although tracking is practiced in almost all countries throughout the developed world, the nature and extent of tracking differs greatly between—and even within—countries. Given that the nature of tracking may have differential effects on students' academic self-concept, it is important to distinguish between different types of tracking. One important feature by which types of tracking can be described is their institutional level (Trautwein et al., 2006; Trautwein, Köller, Lüdtke, & Baumert, 2005). That is, tracking can take place *between* schools or *within* schools. Additionally, tracking can sort either *students* or *courses*, meaning that a student's track could be an overarching program that determines all course work throughout school, or track could designate only the level of a particular course, and students could be allowed to take different levels of courses in different subjects (Lucas, 1999). Whereas tracking students into overarching programs can occur either between or within schools (both of which we will call *streaming* in the present paper), tracking courses subject by subject can occur only within schools. Thus, tracking can be categorized into the following three types: (a) *Between-school streaming* we define as a system where students of different abilities go to completely different schools. This can

be considered the most pronounced type of tracking. Perhaps the most well-known example of this type of tracking is the three-tiered tracking system practiced in several German states. (b) *Within-school streaming* refers to systems where students are tracked within schools and assigned to different tracks, programs, or streams for all subjects. This is practiced in many schools in Belgium, Portugal, or Luxembourg, for instance. (c) *Course-by-course tracking* we define as a system where students are tracked within schools only with respect to certain subjects. This type of tracking is very common in the comprehensive school systems of the United States, the United Kingdom, and Australia. It should be noted that multiple types of tracking can occur in one country. For instance, several German states have not only single-track schools (between-school streaming) but also some schools practicing within-school streaming and others practicing course-by-course tracking. Thus, the different types of tracking are mutually exclusive not with respect to countries but only with respect to schools.

Empirical Studies on the Association Between Tracking and Academic Self-Concept

Although previous research has not systematically investigated the effects of tracking type on academic self-concept, aggregating several studies of single tracking types provides some limited evidence on this question. As we review the following studies, note that several did not measure individual and track mean achievement, so in these cases we can make only inferences about the existence and size of contrast and assimilation effects underlying the association between track level and academic self-concept across the different types of tracking.

A study by Marsh et al. (2000), conducted in the highly achievement-segregated high school system in Hong Kong (which can be thought of as similar to between-school streaming), found strong contrast and small assimilation effects, resulting in students having higher academic self-concepts in low-achieving schools than in high-achieving schools. Similarly, Schwarzer, Lange, and Jerusalem (1982) found track level to be negatively associated with students' academic self-concept a few months after students were grouped into different between-school streaming tracks. A study of within-school streaming found that immediately after track placement, high-track students had higher academic self-concepts than low-track students, but over the course of 3 years, low-track students' academic self-concept became higher than that of high-track students (Liu, Wang, & Parkins, 2005). However, the last two studies did not control for individual and mean academic achievement, meaning that we can only infer that contrast effects outweighed assimilation effects. In studies of course-by-course tracking, the evidence is mixed. Some studies in the United States—though not controlling for individual and track mean achievement—found track level

to be positively associated with students' academic self-concept, which could be an indication of strong assimilation effects (Chiu et al., 2008; Oakes, 1985); others reported no association between students' mathematics track and their mathematics self-concept (Fuligni, Eccles, & Barber, 1995), which may indicate that contrast and assimilation effects cancel each other out. Studies of course-by-course tracking within single-stream schools (i.e., when students are tracked within an already narrowed achievement distribution) either found assimilation and contrast effects to be of similar size (Preckel & Brüll, 2010) or contrast effects to be stronger than assimilation effects (Trautwein, Lüdtke, Marsh, & Nagy, 2009). With one exception (Ireson & Hallam, 2009), studies comparing course-by-course tracking to untracked settings mostly observed that low-achieving tracked students had higher (and high-achieving students had lower) academic self-concepts than the corresponding untracked students (Catsambis, Mulkey, & Crain, 2001; Marsh et al., 1995; Mulkey et al., 2005; Trautwein, Köller, & Kämmerer, 2002).

While all of the studies reported above were conducted in single countries, only one study, by Dupriez, Dumay, and Vause (2008), has examined students' academic self-concept internationally across a large set of countries with different tracking systems. The authors categorized tracking systems across 26 developed countries from the most to the least rigid types of tracking and compared the academic self-concept of high- and low-achieving students. Although this study compared low- versus high-achieving students rather than students in different tracks, meaning differences between tracks were not directly observed, it does fit the pattern emerging from the studies conducted in single countries: The authors found that low-achieving students' academic self-concept was highest in the earliest and most rigidly tracked countries (e.g., Germany, the Netherlands, and Hungary), while it was lowest in those countries with very little tracking at all (e.g., Sweden and Finland). Results for countries that track students at an older age (e.g., France, Italy, and Spain) or that use course-by-course tracking (e.g., the United States, Canada, Australia, and the United Kingdom) were in between these two extremes. Taken together, previous empirical studies on the association between tracking and academic self-concept seem to suggest that the association between track level and students' academic self-concept does differ according to the type of tracking. Furthermore, studies that have compared different types of tracking with respect to outcomes other than academic self-concept, such as self-esteem (Van Houtte, Demanet, & Stevens, 2012), study involvement (Van Houtte & Stevens, 2009), and achievement expectancies in mathematics (Reuman, 1989), lend further support for the hypothesis. However, the three different types of tracking have not yet been compared systematically with respect to their effects on students' self-concept—the aim of the present study. In the following, we lay out how different types of tracking may have different consequences for the development of students' academic self-concept.

Tracks as Reference Groups for Social Comparison

The pattern emerging above, that the effects of tracking on academic self-concept differ according to the type of tracking, may be the result of different social comparison processes (Reuman, 1989; Van Houtte et al., 2012). In other words, different types of tracking may create different reference groups for social comparison. Given that between-school and—to a lesser degree—within-school streaming are explicit and rigid types of tracking, one might suppose that students are more aware of their track membership in these types of tracking than in course-by-course tracking. However, we argue that whom students are exposed to on an everyday basis is more important than the rigidity of their track. Students in each of the three types of tracking do indeed interact with different groups of students in their daily lives: Students in between-school streaming are never confronted with students in other tracks, as they attend completely different school buildings. Students in within-school streaming have a moderate amount of exposure to students in other tracks, as they attend the same school but spend all of their time at school with students in their own track. Finally, students in course-by-course tracking have the highest level of exposure to students in other tracks, as they may be tracked differently in different academic subjects and are also exposed to students of all achievement levels when attending non-tracked subjects, such as physical education. Theories on intergroup behavior do in fact state that a person's choice of reference groups depends on both *proximity* (Tajfel & Turner, 1986) and *salience* (Richer, 1976). We therefore argue that the most salient reference group for students in course-by-course tracking is all students of their age in the entire school. Students in course-by-course tracking observe the grouping process on an everyday basis and are thus constantly reminded of the relative status of their track within the entire age cohort. In contrast, students in within- and between-school streaming most likely compare themselves only to students from their own track and "forget" about students in other tracks after they have been grouped into tracks at the beginning of secondary school or the school year, meaning that the broader age cohort becomes less salient as a reference group. In between-school streaming, this should be even more pronounced than in within-school streaming.

Hypotheses

Based on the considerations in the previous section, we hypothesize that students in types of tracking that expose them more frequently to peers in other tracks are more likely to include those other tracks in their reference group and will thus experience stronger assimilation effects than students whose tracks are more isolated. Accordingly, we expect assimilation effects to be strongest in course-by-course tracking and weakest—if they exist at all—in between-school streaming, with within-school streaming taking

a middle position. Based on previous research (Marsh & Hau, 2003; Seaton et al., 2009), we assume that contrast effects exist in all types of tracking and are similar in size. So when considering students with the same achievement level, this should result in different net associations between track level and academic self-concept for each type of tracking. More specifically, based on previous findings, we expect assimilation effects to outweigh contrast effects in course-by-course tracking, resulting in high-track students having higher academic self-concepts than low-track students, and contrast effects to outweigh assimilation effects in within- and between-school streaming, resulting in low-track students having higher academic self-concepts than high-track students.

We tested our hypotheses about the effects of tracking type on academic self-concept in an internationally comparative framework using the PISA 2003 data set. Using international data allowed us to examine a large number of students in each of the three types of tracking, across a wide range of country contexts. The 2003 data set provides a unique opportunity to study this research question, as it is the only cycle of PISA that collected information on course-by-course tracking as well as between- and within-school streaming. PISA also includes achievement measures, making it possible to disentangle contrast and assimilation effects. We focus on mathematics self-concept because this was the focal domain of achievement in the PISA 2003 data set.[1] PISA also included a number of variables that have been shown to be associated with students' academic self-concept and/or track placement and which we were able to add as controls. We included students' mathematics course grades, as they have been found to mediate much of the impact of track level and track mean achievement on academic self-concept (Preckel & Brüll, 2010; Skaalvik & Skaalvik, 2002; Trautwein et al., 2006). We also included gender, as boys usually have a higher mathematics self-concept than girls (Marsh & Yeung, 1998; Nagy et al., 2010). Last, we included student socioeconomic status (SES) and ethnicity, because low-SES students and minority students often disproportionately attend low tracks (Buchmann & Park, 2009; Lucas, 2001; Maaz et al., 2008; Oakes, 1985).

Method

Sample

We used as our data set PISA 2003, conducted by the Organization for Economic Cooperation and Development (OECD). PISA is a large-scale international assessment that tests nationally representative samples of 15-year-olds in each participating country, regardless of grade. It is conducted every 3 years as a cross-sectional assessment of mathematics, reading, and science. PISA 2003 was selected as the data set for this analysis because it includes the most extensive international data available on course-by-course tracking in mathematics.

Research in the United States and other course-by-course tracking countries has shown that mathematics is the most important domain of course-by-course tracking. It is the subject most commonly tracked (Ireson & Hallam, 2009), mathematics track placement may determine placement in other subjects (Lucas & Berends, 2002), and advanced mathematics course work is a strong predictor of college attendance and completion (Adelman, 1999).

Country PISA samples are nationally representative of all 15-year-olds who are enrolled in school. The samples represent both public and private school students but do not represent difficult-to-access students, such as those in geographically remote or special education schools. These exclusions were not to exceed 5% of any country's 15-year-old population (OECD, 2005b). In 2003, about 276,000 students in 41 countries participated, including all 30 OECD countries at the time (wealthy democratic countries with market-based economies) and 11 non-OECD countries. Following other comparative research (e.g., Dupriez et al., 2008), we limited our sample to OECD countries, excluding OECD members Mexico and Turkey, in order to obtain a more homogeneous sample of wealthy countries. We counted the Flemish and French communities of Belgium as two "countries" because they represent two distinct cultural contexts and also have two separate school systems. We also excluded the 9 countries that either do not practice any tracking for 15-year-olds or did not collect student tracking data.[2] Similarly, we excluded students in untracked comprehensive schools.[3] We excluded untracked schools because the goal of our study was not to explain variance in mathematics self-concept across countries (for which there is already a large literature) but to explain gaps in mathematics self-concept between high- and low-track students in different types of tracking. Untracked schools cannot be included in such a study because they have no track data. Thus, the data used in our analyses were nationally representative of all tracked students. After exclusions, the sample consisted of about 99,000 students in about 3,400 schools in 20 countries. PISA uses two-stage sampling of first schools and then students (OECD, 2005b), resulting in a complex sample design. All statistics reported below use the appropriate student and school sampling weights, as recommended in the PISA 2003 data analysis manual (OECD, 2005a).

Instruments

The means, standard deviations, and intercorrelations of all variables are presented in Appendix A, Table A1, in the online journal.

Mathematics Self-Concept

The dependent variable of interest, mathematics self-concept, was a PISA-created index based on student responses to a set of items based on the commonly used Self-Description Questionnaire (SDQ) II (Marsh,

1990c; OECD, 2005b). The wording of the items was as follows: "How much do you disagree or agree with the following statements about how you feel when studying mathematics?" (1) I am just not good at mathematics (inverted), (2) I get good grades in mathematics, (3) I learn mathematics quickly, (4) I have always believed that mathematics is one of my best subjects, and (5) In my mathematics class, I understand even the most difficult work. Responses were on a 4-point Likert-type scale with the categories *strongly agree, agree, disagree*, and *strongly disagree*. The items were combined into a single scale using an item response theory model. The reliability of the PISA mathematics self-concept index ranged from 0.81 in Hungary to 0.93 in Iceland with a median of 0.89 across the 20 countries in the sample (OECD, 2005b). To ease interpretation of our results, we restandardized the scale to have a mean of 0 and a standard deviation of 1 within our sample of 20 countries.

Student Track

In countries that track students in between- or within-school streaming, students responded to the questionnaire item, "Which of the following programs are you in?" The country-specific responses varied in number and we recoded them into *high, middle*, and *low* categories where possible, except in cases where there were only two tracks. The categories generally corresponded to *academic, high vocational*, and *low vocational*, respectively. We used as resources the International Encyclopedia of National Systems of Education (Postlethwaite, 1995), Eurydice.org (Eurydice, 2011), and any descriptions of national education systems in the empirical literature reviewed above. Of the four countries in our data set that practice within-school course-by-course tracking for some or all students, students in three countries (Australia, Germany, and Iceland), were asked, "What type of mathematics class are you taking?" and responded with high-level, medium-level, or basic-level mathematics classes. In the fourth country, the United States, students were also asked, "What type of mathematics class are you taking?" and responded in six categories: Pre-Algebra or General Mathematics, Algebra I, Geometry, Algebra II, Pre-Calculus or Calculus, or Other. Since students in American high schools typically progress through these hierarchical classes year by year in a particular order, mathematics track can be recoded from a combination of this variable and a variable for student grade (Schiller, Schmidt, Muller, & Houang, 2010). Mathematics class was recoded as *middle mathematics track* if the student was at grade level (i.e., Algebra I in 9th grade, Geometry in 10th grade, etc.), *high mathematics track* if the student was above grade level, and *low mathematics track* if the student was below grade level.

Type of Tracking in the School

As outlined above, we theoretically distinguished between three types of tracking: between-school streaming, within-school streaming, and course-by-course tracking. We identified between-school streaming schools as those where all students reported the same stream and within-school streaming schools as those where there was variation in student-reported streams. We identified course-by-course tracking schools as those where all students were in a comprehensive study program and where there was information available for student mathematics classes. This included nearly all schools in Australia, Iceland, and the United States and *Gesamtschulen* (comprehensive schools) in Germany. As explained in the Sample section above, we excluded untracked comprehensive schools. Table 1 displays all systems in the sample, sorted by tracking type, along with the percentage of students in each track. Although there are 20 countries, there are a total of 28 entries in the table, as 6 countries have schools represented in each of two different types of tracking, and 1 country, Germany, has schools of all three types. We refer to each type of tracking within each country as a "tracking system."

Mathematics Achievement

The OECD-developed PISA mathematics assessment is designed to capture not mastery of a specific school curriculum but rather students' mathematical "literacy," that is, their ability to apply their knowledge and skills in real-world situations (OECD, 2003). PISA conceptualized the achievement of country populations and subpopulations as latent variables measured with uncertainty. Student mathematics scores consisted of five plausible values for each student rather than a single score. All statistics reported below take into account the uncertainty associated with plausible values, as recommended in the PISA 2003 data analysis manual (OECD 2005a). The mean score of mathematics achievement in OECD countries is 500 and the standard deviation is 100. To ease interpretation of our results, we restandardized student achievement to have a mean of 0 and a standard deviation of 1 across our sample of 20 countries. We then calculated school-level and track-level mean achievement.

Mathematics Course Grades

Teacher-assigned mathematics course grade was based on student responses to one of the following two questions: either "In your last school report, what was your mark in Mathematics?" or "In your last school report, how did your mark in mathematics compare with the pass mark?" Most countries in the sample collected one of these two variables, but the resulting grade variable differed across countries. In six systems, students reported

Table 1
Descriptions of Tracking Systems

Tracking System	Analytic Sample Size (Unweighted)	Percentage			
		High Track	Medium Track	Low Track	Track Missing
Course-by-course tracking					
Australia	12,551	36.14	47.65	16.21	12.69
Germany	426	46.54	33.56	19.90	17.52
Iceland	3,167	41.60	42.20	16.20	35.79
United States	5,456	28.98	49.22	21.80	11.99
Within-school streaming					
Belgium-Flanders	3,234	25.52	43.04	31.44	0
Belgium-French	2,552	40.31	32.23	27.46	0.04
France	578	84.58	0	15.42	0
Germany	460	0	73.54	26.46	1.69
Ireland	731	64.71	29.40	5.89	0
Luxembourg	2,793	73.44	14.27	12.29	0
Netherlands	1,616	43.81	46.34	9.85	0.20
Portugal	2,181	82.39	0	17.61	0
Switzerland	329	26.73	27.33	45.94	0.70
Between-school streaming					
Austria	4,597	21.29	53.67	25.04	0
Belgium-Flanders	1,825	84.48	4.68	10.84	2.01
Belgium-French	1,185	64.82	16.87	18.32	0.02
Czech Republic	6,263	14.32	69.61	16.07	0
France	2,183	84.62	0	15.38	0
Germany	3,627	37.81	32.07	30.11	0.04
Greece	4,152	78.12	0	21.88	0
Hungary	4,339	37.97	41.29	20.74	0
Ireland	677	94.90	5.10	0	0
Italy	11,057	40.17	33.99	25.84	0
Japan	4,707	75.47	0	24.53	0
Korea	5,137	73.44	0	26.56	0
Netherlands	2,376	4.69	1.03	94.28	0.09
Slovak Republic	6,835	58.12	21.51	20.37	0
Switzerland	750	41.39	4.66	53.95	0.17

Note. Countries are sorted alphabetically within categories. Countries with more than one type of tracking appear in multiple categories.

grades on a 100-point scale (Belgium-Flanders, Belgium-French, Iceland, Ireland, Portugal, and the Slovak Republic). In six countries, students reported grades in 5 to 10 categories (Austria, the Czech Republic, Germany, Hungary, Italy, and Korea). In four countries, students reported

only whether they received a passing or failing mathematics grade (Australia, Greece, the Netherlands, and the United States). And in four countries, students did not report any mathematics grades (France, Japan, Luxembourg, and Switzerland). Additionally, the same grade categories were assigned to students at different rates in different countries (e.g., 12% of students received a failing mathematics grade in the United States, while 28% did in the Netherlands and only 4% did in Greece). Therefore, to create greater comparability across countries, we recoded mathematics grade into a percentile for pooled analyses of all countries but kept the variable in its original format for individual country analyses. For pooled analyses of all countries, students in countries that did not collect grade data were recoded to the mean, that is, the 50th percentile mathematics grade. Percentiles were then transformed onto a scale that ranged from 0 to 1 rather than 0 to 100; thus, they can be interpreted as the proportion of students earning grades lower than the student in question.

Gender

We used student-reported gender and recoded female to 1 and male to 0, making males the reference group.

SES

We used as our measure of SES the OECD-created "Index of Social and Cultural Status," which combines highest level of parental educational attainment, highest level of parental occupational status, and number of home possessions, including books, computers, and educational resources, and other possessions (OECD, 2005b). To ease interpretation of our results, we restandardized SES to have a mean of 0 and a standard deviation of 1 across our sample of 20 countries.

Ethnic/Language Minority Status

Although it is difficult to create a comparable measure of ethnicity across countries with widely disparate ethnic compositions, we did so using all pertinent data from PISA student reports. We created a dummy variable that equaled 1 for any student born in another country or who had at least one parent born in another country, for any student whose family spoke a language at home other than the national language(s), and/or for any non-White student in the United States (student race was only reported in the United States).

Statistical Analyses

The goal of our analyses was to test our hypothesis that the strength of assimilation effects varies across the three types of tracking and drives the different net associations between track level and mathematics self-concept

for each type of tracking. We tested the hypothesis in two different ways: In the first set of analyses, we pooled all students across the 20 countries into a single model that allowed a broad comparison of the three different types of tracking. Second, we conducted a finer-grained analysis of individual countries in order to examine whether the general results held for all countries. Each set of analyses began with a descriptive model (Model 1), in which we looked at the association between track level and mathematics self-concept for each type of tracking. In order to account for the fact that high-track students generally score higher than low-track students, Model 2 then controlled for individual achievement. This allowed us to determine which effect, contrast or assimilation, outweighed the other in each type of tracking. Next, Model 3 disentangled contrast and assimilation effects by adding a control for track mean achievement (i.e., contrast effects). A positive regression coefficient of track level then reflects assimilation effects (Trautwein et al., 2006). Finally, Model 4 tested the robustness of these findings by adding additional controls (mathematics course grades, gender, SES, and ethnicity).

A hierarchical linear model (HLM), also known as a multilevel model, was the appropriate model choice for our data because it accounts for clustered standard errors at the school and country levels (see Raudenbush & Bryk, 2002). Our first (pooled) model was a three-level HLM (students nested within schools, nested within countries) that had random effects at the school level but only fixed effects and no random effects at the country level (because of insufficient degrees of freedom with a sample of only 20 countries). We included country fixed effects because each country represents a different cultural context, and fixed effects allow us to account for different mean levels of mathematics self-concept in each country (see Appendix C in the online journal for more detail on this issue). Each individual country model was a two-level HLM (students within schools). Model equations are presented in Appendix B in the online journal. The pooled model and individual country models were conceptually similar but differed in three key ways. First, the pooled model used interaction terms to examine differences in mathematics self-concept between types of tracking. These consisted of dummies for course-by-course tracking and between-school streaming at the school level and the cross-level interactions Course-by-Course Tracking × High Track and Between-School Streaming × High Track.[4] On the other hand, the individual country models compared mathematics self-concept between high- and low-track students separately for each type of tracking within each country. Second, in the pooled analysis, we had to combine middle and low tracks into a single reference group in order to harmonize track coding across countries with two and three tracks. In the individual country models, we were able to include all three tracks in each analysis, if applicable. Third, in the pooled models, we used school mean achievement rather than track mean achievement. We made this decision in order to maintain consistency across between-school streaming countries

(where track mean achievement was a school-level variable) and within-school streaming and course-by-course tracking countries (where track mean achievement varied within schools). In the individual country models, analyzing each country separately allowed us to enter track mean achievement differently for each type of tracking.

Throughout the pooled models, country fixed effects were grand-mean centered. In both the pooled and the individual country models, achievement and SES variables were centered within the entire international sample through the standardization procedure described in the Instruments section above; mathematics course grades were country mean centered through the conversion to percentiles described above. Gender and ethnicity were grand-mean centered, and track variables were uncentered, so that the intercept was the estimated mean mathematics self-concept for middle-track students across the entire sample. All models were estimated with HLM 7.0 software using the provided school weights at Level 2 and individual weights at Level 1, transformed to conditional within-school weights, following Rabe-Hesketh and Skrondal (2006). In the pooled analyses, the school-level weights also incorporated the OECD's PISA country weights. Country weights are *house* weights as opposed to *senate* weights, meaning each country contributed to the analysis in proportion to the size of its 15-year-old population, rather than each country contributing equally regardless of size. We chose to use house weights for the pooled analyses both for consistency with most prior literature containing pooled analyses of PISA data and because house weights were consistent with the purpose of the pooled model to ignore country location and broadly compare all OECD students in the three types of tracking. On the other hand, the individual country models effectively weight all countries equally because all results are presented side by side for large and small countries. Thus, each of our two analyses tests the robustness of our findings under a different country weighting approach.

All variables except achievement were derived from student questionnaires and had low levels of missingness, ranging from 0.10% for gender to 5.46% for track. As missing data can bias results, we imputed all missing data using multiple imputation by iterative chained equations (STATA's -mi impute- command), creating five imputed data sets (Schafer & Graham, 2002). We then used HLM's multiple imputation function, which runs each model five times and computes the mean of the estimates and standard errors that reflect uncertainty due to imputation. Similarly, the plausible values of achievement in PISA can also be understood in a multiple imputation framework. The proper technique for dealing with plausible values uses the same standard error formula to account for imputation variance (OECD, 2005a, 2005b). Therefore, we also used HLM's multiple imputation function to correctly deal with plausible values of achievement.

Results

Pooled Three-Level Models

Table 2 displays the results for the three-level models pooling all 20 countries. Model 1 shows the simple descriptive differences in mathematics self-concept between students in the high track and students in the low and middle tracks (hereafter "low track") across the three types of tracking. Model 1 revealed that the relationship between track and mathematics self-concept did indeed depend on the type of tracking: The statistically significant interaction term Course-by-Course Tracking × High Track indicated that the mathematics self-concept gap between high- and low-track students in course-by-course tracking was significantly greater than the gap between tracks in the reference group, within-school streaming. For between-school streaming, the gap between tracks was about the same as that in within-school streaming, as the non-significant interaction term Between-School Streaming × High Track indicates. In terms of point estimates obtained by summing the main and interaction effects, this can be nicely illustrated: The difference was much more pronounced in course-by-course tracking, where the gap was 0.42 standard deviations, compared to only 0.12 and 0.05 standard deviations for within- and between-school streaming, respectively. General linear hypothesis tests showed that this gap was significantly different from 0 for course-by-course tracking ($p < .001$) and for within-school streaming ($p = .001$) but not for between-school streaming ($p = .064$). Despite these differences, the point estimates indicated that high-track students had higher mathematics self-concepts than low-track students in all three types of tracking.

Model 2 added individual student mathematics achievement to the model in order to see whether contrast or assimilation effects outweighed the other. The most striking finding from Model 2 is that the high-track coefficient became negative. That is, for the reference group of within-school streaming, after controlling for individual achievement, high-track students actually had significantly lower mathematics self-concepts than low-track students, a gap of about -0.14 standard deviations. Similarly in between-school streaming, the interaction was negative and significant, indicating that in this type as well, high-track students had lower self-concepts, and in fact the gap between tracks was -0.33 standard deviations, even larger than in within-school streaming. But for course-by-course tracking, the interaction remained large and positive and the estimated gap was 0.19 standard deviations, so high-track students still had higher mathematics self-concepts than low-track students, as in Model 1. All three gaps were significantly different from 0 ($p < .001$). The finding that high-track students had lower mathematics self-concepts than low-track students in within- and between-school tracking supports our prediction that in these types of tracking,

Table 2
Coefficients From Pooled Three-Level Models Predicting
Mathematics Self-Concept

Variable	Model 1 β (*SE*)	Model 2 β (*SE*)	Model 3 β (*SE*)	Model 4 β (*SE*)
Student level				
High track	0.12**	−0.14***	−0.10**	−0.06*
	(0.04)	(0.03)	(0.03)	(0.03)
Mathematics achievement		0.42***	0.54***	0.35***
		(0.02)	(0.02)	(0.02)
Mathematics grade (percentile, 0-1)				1.77***
				(0.04)
Female				−0.28***
				(0.02)
SES				0.02
				(0.01)
Ethnic/language minority				0.17***
				(0.03)
School level				
Course-by-course tracking	−0.12	−0.01	−0.05	−0.01
	(0.07)	(0.08)	(0.07)	(0.06)
Course-by-Course	0.30***	0.32***	0.27***	0.23***
Tracking × High Track	(0.06)	(0.05)	(0.05)	(0.05)
Between-school streaming	0.09*	0.20***	0.09*	0.09**
	(0.04)	(0.04)	(0.04)	(0.03)
Between-School	−0.06	−0.19 ***	−0.02	−0.04
Streaming × High Track	(0.04)	(0.04)	(0.04)	(0.04)
School mean mathematics			−0.28***	−0.22***
achievement			(0.03)	(0.03)
Intercept	−0.10**	−0.05	−0.04	−0.08**
	(0.03)	(0.03)	(0.03)	(0.03)
Country fixed effects	x	x	x	x
Variance explained (students)	0.02	0.15	0.15	0.28
Variance explained (schools and countries)	0.00	0.00	0.03	0.44
n (students)	99,004	99,004	99,004	99,004
n (schools)	3445	3445	3445	3445
n (countries)	20	20	20	20

Note. x indicates that fixed effects were included in the model. All continuous variables except mathematics grade were standardized (*M* = 0, *SD* = 1) at the student level across the full sample of 20 countries.
*p < .05. **p < .01. ***p < .001.

contrast effects outweigh assimilation effects, while in course-by-course tracking, assimilation effects outweigh contrast effects.

In Model 3, we tested whether this pattern was indeed driven by differences in assimilation effects, specifically that assimilation effects are strongest in course-by-course tracking and weakest in between-school streaming—our main hypothesis. As assimilation effects are operationalized in terms of the coefficient on track after controlling for individual and mean achievement (i.e., contrast effects), we added school mean mathematics achievement to the model. Consistent with a long line of research, we found a negative coefficient for school mean mathematics achievement, thus evidence for contrast effects. If assimilation effects are indeed strongest for students in course-by-course tracking, we would expect the Course-by-Course Tracking × High Track interaction to remain positive and significant and the estimated gap in mathematics self-concept between high- and low-track students to remain positive and larger than the gaps for the other two types of tracking. In fact, the interaction was positive in Model 3 (it was only slightly reduced in magnitude from Model 2 and did not change in significance), and the estimated gap was 0.17 standard deviations ($p < .001$), confirming our hypothesis. With respect to the other two types of tracking, we expected assimilation effects to be weaker for between-school streaming than for within-school streaming. If this is the case, we would expect the Between-School Streaming × High Track coefficient to remain negative, the estimated gap for within-school streaming to be small and positive, and the gap for between-school streaming to be the smallest of all, or even zero. However, contrary to our predictions, the estimated gaps for within- and between-school streaming both remained negative in Model 3, at -0.10 and -0.12 standard deviations, respectively ($p = .003$ and $p < .001$), meaning we observed no assimilation effects at all for these two types of tracking. Furthermore, the Between-School Streaming × High Track interaction was close to 0 and not significant. Thus, there was no significant difference in assimilation effects between within- and between-school streaming. This suggests that the main distinction in the size of assimilation effects across different types of tracking is between course-by-course tracking and both forms of streaming (within- and between-school). Moreover, consistent with prior literature, additional models not reported here revealed that contrast effects did not differ systematically by type of tracking, as interactions between school mean achievement and type of tracking were small in size and not statistically significant. This also means that the difference between within- and between-school streaming observed in Model 2 was not attributable to differences between the two types of tracking in the size of assimilation or contrast effects, so it must be due to differences in the distributions of covariates in the two types of tracking. Closer examination indeed showed that achievement gaps between tracks were larger in between-school streaming than in within-school streaming.

In Model 4, we tested the robustness of these results after adding four control variables: mathematics course grades, gender, SES, and ethnicity. Again, the interaction terms remained in the same direction as previous models, but the Couse-by-Course Tracking × High Track interaction was somewhat reduced in magnitude and significance. The point estimates of the gaps in mathematics self-concept for course-by-course tracking, within-schools streaming, and between-school streaming were 0.16, −0.06, and −0.11 standard deviations, respectively. All three were statistically significant ($p < .001$, $p = .047$, and $p = .004$, respectively). The results from Model 4 indicated that the difference in high-track coefficients between course-by-course tracking and within-school streaming was partially but not fully accounted for by mathematics course grades and demographics. However, this change in coefficients may have been underestimated because mathematics grades were unmeasured in some countries and were measured only as a dichotomous pass/fail in some others.

Individual Country Models

In order to examine whether the results from the pooled models were consistent across all countries, we next estimated individual two-level models for each of the tracking systems in our sample. The key coefficients from these models—the coefficients on high track and low track with middle track as the reference group—are reported in Table 3 (systems with only one track coefficient had only two tracks). As the results for Models 1 and 2 were fully consistent with the pooled models, and Model 3 (and 4) tests our main hypothesis, Table 3 lists the track coefficients only for Models 3 and 4, sorted by type of tracking; countries with more than one type of tracking appear in multiple categories.

As in the pooled models, Model 3 tested our main hypothesis that assimilation effects differ according to the type of tracking by controlling for individual and track mean achievement, while Model 4 tested the robustness of these results by adding controls for mathematics course grades and demographics. In Model 3, for course-by-course tracking, high-track coefficients ranged from 0.16 in the United States to 0.59 in Iceland, with a median of 0.32; low-track coefficients ranged from −0.08 in the United States to −0.36 in Iceland, with a median of −0.33. In other words, students in high tracks had high mathematics self-concepts, and students in low tracks had low mathematics self-concepts, which is consistent with the results from the pooled models. The results for both within- and between-school streaming were the reverse of those for course-by-course tracking: The high-track coefficients were slightly negative (median = −0.04) and the low-track coefficients were positive (median = 0.21), indicating that students in low tracks had higher mathematics self-concepts than students in high tracks. These results can also be seen in Figure 1, in which the differences

Table 3
**Track Coefficients From Individual Country Models Predicting
Mathematics Self-Concept**

Tracking System	Model 3		Model 4	
	High β	Low β	High β	Low β
Course-by-course tracking				
Australia	.42***	−.32***	.40***	−.26***
Germany	.21	−.34	.10	−.27
Iceland	.59***	−.36***	.31***	−.12*
United States	.16**	−.08	.13*	−.07
Within-school streaming				
Belgium-Flanders	.01	.14*	.00	.00
Belgium-French	−.02	.40***	.08	−.01
France	−.16		−.10	
Germany		.40**		.19*
Ireland	−.04	.02	−.07	−.22
Luxembourg	.07	.68***	.04	.55***
Netherlands	−.08	.27**	−.07	.22**
Portugal	.08		.04	
Switzerland	−.52*	.70***	−.43*	.70***
Between-school streaming				
Austria	.01	.21*	.01	.08
Belgium-Flanders	−.11	.21	−.08	−.13
Belgium-French	.01	.65***	−.10	.21
Czech Republic	−.13	.07	−.09	.00
France	−.32*		−.34*	
Germany	−.25*	.17	−.17*	−.01
Greece	.07		.10	
Hungary	−.04	.34***	−.12**	.10
Ireland	.09		.28	
Italy	.00	−.15**	−.03	−.11**
Japan	−.04		−.02	
Korea	−.04		.09*	
Netherlands		.42**		.28*
Slovak Republic	.10	.04	.01	−.06
Switzerland	−.44**	.09	−.53***	−.09

Note. Countries are sorted alphabetically within categories. Countries with more than one type of tracking appear in multiple categories. Italicized estimates for Model 4 indicate that mathematics course grades were unavailable for that country and could not be included among the controls.
*$p < .05$. **$p < .01$. ***$p < .001$.
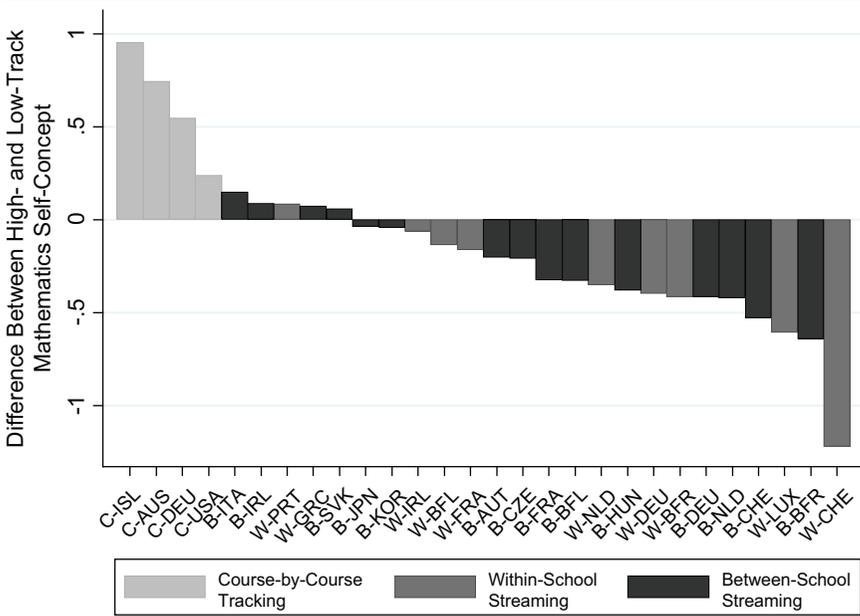
Figure 1. **Estimated difference in mathematics self-concept between high and low track students from individual country models (Model 3).**

Note. Models control for individual and mean achievement.

between the estimated high-track and low-track mathematics self-concepts for each country are plotted: The four largest positive differences between high- and low-track mathematics self-concept were in the four course-by-course tracking systems in Iceland, Australia, Germany, and the United States. Five systems had smaller positive differences: Italy, Ireland, Portugal, Greece, and the Slovak Republic. The rest of the systems all had negative differences between high- and low-track mathematics self-concepts, some of which were quite large.

Furthermore, it is evident from both Table 3 and Figure 1 that differences in mathematics self-concept between high- and low-track students were not systematically larger or smaller for between-school streaming than for within-school streaming. This is also consistent with the results of the pooled Model 3 above, where within- and between-school streaming were not significantly different. In the individual country models, a further test of this finding is possible: Because the set of the countries represented differs somewhat between the two types of tracking, this comparison can also be made only for countries that have both within- and between-school

streaming. In those seven countries (Belgium-Flanders, Belgium-French, France, Germany, Ireland, the Netherlands, and Switzerland), the direction of the track coefficients was the same in every case except Ireland, where all track coefficients were close to 0. Variation in significance levels of coefficients was likely due to differences in sample sizes across types of tracking. Finally, in the one country with all three types of tracking, Germany, all three types followed the general pattern: For course-by-course tracking, high-track coefficients were positive and low-track coefficients were negative, while for within- and between-school streaming, high-track coefficients were negative and low-track coefficients were positive. Thus, the findings from the individual country models fully support the findings from the pooled models that assimilation effects could be detected only in course-by-course tracking and not in within- or between-school streaming. Moreover, consistent with our expectations and the findings from the pooled models, the coefficients for track mean achievement representing contrast effects, which are not reported in Table 3, were negative in nearly all systems (with the exception of two where they were close to 0) and did not differ systematically by type of tracking.

The results for Model 4, when controls for mathematics course grades and demographics were added, were generally consistent with Model 3. However, many coefficients were smaller in size and/or less significant, as these controls accounted for varying amounts of the difference between high- and low-track mathematics self-concept across countries. Still, the overall pattern holds: Assimilation effects were observed only in course-by-course tracking. In addition to the analyses presented above, we conducted several supplemental analyses as robustness checks of our findings, specifically, our coding of track in the pooled models, the level at which mean achievement was computed, and differences in the variance of self-concept across countries. In every case, results were consistent with those presented here. More information on these analyses can be found in Appendix C in the online journal.

Discussion

The aim of the present study was to examine how different types of tracking—between-school streaming, within-school streaming, and course-by-course tracking—shape students' mathematics self-concept. This was done in an internationally comparative framework using the PISA 2003 data set, which allowed us to observe a large number of students in each of the three types of tracking across a wide range of country contexts. Furthermore, by disentangling contrast and assimilation effects, we were able to move beyond the descriptive association between track level and academic self-concept that was the focus of several previous studies. Overall, the results of our analyses indicate that, after controlling for

individual achievement, for students in course-by-course tracking, high-track students have higher mathematics self-concepts and low-track students have lower mathematics self-concepts, thus showing that assimilation effects outweighed contrast effects. Inversely, for students in between-school and within-school streaming, high-track students have lower mathematics self-concepts and low-track students have higher mathematics self-concepts, thus showing that contrast effects outweighed assimilation effects. As hypothesized, this pattern was driven by differences in observed assimilation effects across the three types of tracking. Only for students in course-by-course tracking did we observe a positive association between track level and mathematics self-concept, after controlling for contrast effects. We found no evidence for assimilation effects in either between- or within-school streaming, even though we hypothesized that there would be some weak assimilation effects at least in within-school streaming. In fact, in many countries, there was actually a negative association between track level and mathematics self-concept for students in between- and within-school streaming that was not explained by contrast effects.

These findings support the idea that when students are grouped only for certain courses, they observe the grouping process on an everyday basis and are thus constantly reminded of the relative status of their track. This is in line with research showing that in between- and within-school streaming, there is a positive association between track level and academic self-concept immediately after the grouping process, but it diminishes over time when students are no longer exposed to peers of all achievement levels (Liu et al., 2005; Schwarzer et al., 1982), suggesting that for students in between- and within-school streaming, the salient reference group shifts over time from the entire age cohort to only those in their own track. In contrast, for students in course-by-course tracking, their reference group remains the entire age cohort in their school. Taken together, these findings can explain prior inconsistencies in the literature and suggest a solution to the ongoing debate about the effects of tracking on students' academic self-concept. By systematically taking into account the type of tracking, we found a consistent pattern for the effects of tracks on students' academic self-concept.

Theoretical Significance of the Study

Previous research on tracking has established three mechanisms through which tracking affects student outcomes: instructional, institutional, and social mechanisms (Lucas, 1999; Pallas et al., 1994). Whereas the instructional mechanism deals with differences between tracks in instructional quality and the pace and content of subject matter, the institutional mechanism points to differences in the wider society's expectations and perceptions of students as a consequence of track *charters*—explicit, socially defined labels that transcend individual schools (Meyer, 1977). The social

mechanism, which is the focus of the present study, emphasizes that tracks are different social contexts for students' development.

By addressing social contexts, our study fills an important gap in the comparative literature, which has mainly studied instructional and institutional differences between national tracking systems. Moreover, our findings provide a remarkable contrast to those of instructional and institutional research. Prior studies on the instructional mechanism of tracking have found that countries with the most rigid and explicit forms of tracking (between-school, early selection) produce the largest achievement gaps between tracks (Brunello & Checchi, 2007; Dupriez & Dumay, 2006; Duru-Bellat & Suchaut, 2005; Hanushek & Wößmann, 2006). Likewise, studies on the institutional mechanism have shown that more rigid tracking systems produce the largest gaps between tracks in educational aspirations (Buchmann & Dalton, 2002; Buchmann & Park, 2009). However, for the social mechanism, at least for the outcome of mathematics self-concept, our findings revealed that the gap between high- and low-track students is actually most pronounced in the least rigid type of tracking, course-by-course tracking. In contrast, in between- and within-school streaming, being in a low track appears to have fewer negative consequences for student mathematics self-concept. Therefore, it appears that track effects on academic self-concept are better predicted by the psychological reality of students' daily lives than the meaning attached to tracks by society.

Our study also adds further nuance to our understanding of the relationships between the instructional and social mechanisms of tracking. Perhaps our most surprising finding was the observed negative effect of high track on academic self-concept for between- and within-school streaming in some countries after partialling out individual and mean achievement. One possible explanation for this negative effect may be instructional differences between tracks. Teachers in low tracks may deliberately attempt to boost their students' academic self-concept by being more supportive and celebrating the accomplishments of each individual student, which has been shown to increase students' academic self-concept (Liu et al., 2005; Lüdtke, Köller, Marsh, & Trautwein, 2005), whereas high-track teachers may maintain a more demanding and competitive atmosphere. On the other hand, it is widely known in the tracking literature that teachers in low tracks provide lower-level instruction (Baumert et al., 2010; Oakes, 1985) and have lower expectations of their students (Brophy, 1983), which in turn influences students' own perceptions of their abilities (Good, 1987). Thus, different dimensions of teacher behavior may have opposite effects on students' academic self-concept. Although this was not the focus of our study, future studies should investigate how teacher instruction interacts with the social comparison taking place between students attending different tracks. The apparently contrasting effects of tracking on achievement and academic self-concept lead to a complex picture with potentially conflicting messages

for policy and practice. In the next section we will therefore discuss the practical significance of our study and outline potential avenues for tracking policy.

Practical Significance of the Study

The present study has important implications for educational policies on tracking across many countries. In the United States, the detracking movement of the 1990s did not succeed in dismantling course-by-course tracking in the majority of middle and high schools (Loveless, 2013). In contrast, many European countries have passed detracking policies in recent years. For example, in Germany, there have been several attempts to introduce one school for students of all abilities or combine tracks as a means to reduce achievement gaps between tracks (Trautwein, Baumert, & Maaz, 2007). However, our findings indicate that moving to less rigid and explicit types of tracking may not only have positive effects for lower-achieving students but could at the same time have negative side effects for their academic self-concept. Any form of detracking implies an increased heterogeneity of the student body, meaning that lower-achieving students are suddenly confronted with higher-achieving students, thus potentially leading to a drop in their academic self-concept. Negative effects on students' academic self-concept as a result of detracking policies may have long-term consequences for their educational trajectories, as it has been shown that academic self-concept affects subsequent academic achievement as well as educational aspirations and course selection (Marsh et al., 2005; Marsh & Yeung, 1997, 1998; Mulkey et al., 2005; Valentine et al., 2004).

In order for detracking policies to be beneficial for all students, they should be accompanied by policies for instructional practices that bolster the academic self-concept of low-achieving students. Such instructional practices may involve putting less emphasis on competition and ranking of students in the classroom and instead valuing the accomplishments of each individual student. To conclude, our study implies that if academic self-concept is not taken into account, lower-achieving students may not be able to fully take advantage of the increased educational opportunities that detracking provides them.

Limitations of the Present Study

Several limitations of the present study should be noted. First, our study considered only two possible reference groups for students: the school level and the track level. However, the reference groups that students use are known to be far more complex. In fact, previous research indicates that students use multiple reference groups that "compete" with one another, including not only the entire school or entire class but also selected students within or outside a student's class (Marsh et al., 2008; Skaalvik & Skaalvik,

2002). As students in PISA did not report to whom they primarily compared themselves, we had no information about other potential reference groups. Similarly, since PISA students did not report what they considered to be the status or standing of their tracks, we can only infer the processes and mechanisms that were driving the observed assimilation effects. Furthermore, since the basic sampling unit in PISA 2003 was schools rather than classes or tracks within schools, for students in the same track within a school, we could not know whether they were actually in the same classroom.

Moreover, the nature of our data set did not enable a study design that would allow causal inference. Track assignment is intentionally non-random and often produces groups with widely differing socioeconomic and achievement compositions. Therefore, differences between tracks in mathematics self-concept, achievement, or mathematics course grades cannot be taken as the effects of track placement since they are likely also the result of selection. That our findings were robust after adding a number of control variables to our model speaks for their validity. Finally, at the country level, differences between countries in types of tracking are clearly not randomly assigned but are instead the result of historical and cultural differences between countries. With respect to this potential threat to validity, the seven countries in our data set that practice more than one type of tracking provide crucial evidence. As was revealed in the individual country models, the results for these countries were consistent with our general findings in every case except one. The strictest test of all is provided by Germany, which is the only country in our data set where all three types of tracking are practiced. For Germany, the exact same pattern of results was found as for the entire sample of countries. Therefore, the consistency and robustness of our results both between and even within countries strongly suggests that type of tracking is an important determinant of the relationship between track level and academic self-concept.

Notes

[*]The first two authors contributed equally to this work and are listed in alphabetical order.

[1]Previous research has shown that there are no qualitative differences in the development of mathematics self-concept between different cultural settings (Nagy et al., 2010), indicating that it is valid to analyze the effect of tracking type on students' mathematics self-concept across a large number of countries.

[2]The nine countries without tracking data were Canada, Denmark, Finland, New Zealand, Norway, Poland, Spain, Sweden, and the United Kingdom. All of these countries practice course-by-course tracking to some degree, but levels of tracking in some

countries are much higher (e.g., the United Kingdom and Canada) than in others (e.g., Finland).

[3]In between- and within-school streaming systems, untracked students were those who reported being in a comprehensive track, usually lower secondary school. Their representation in each country cohort was as follows: Austria, 0%; Belgium-Flanders, 0%; Belgium-French, 0%; Czech Republic, 0%; France, 38.32%; Germany, 0.65%; Greece, 8.85%; Hungary, 6.08%; Ireland, 63.68%; Italy, 1.57%; Japan, 0%; Korea, 1.57%; Luxembourg, 0%; Netherlands, 0%; Portugal, 35.09%; Slovak Republic, 0%; Switzerland, 80.25%. In course-by-course tracking systems, untracked schools were those where principals reported that the school did not group students by ability for mathematics classes that differed in content and/or difficulty in any cases, and where all students reported that they were in the same mathematics track or all were missing. Their representation in each country cohort was as follows: Australia, 0%; Germany, 0%; Iceland, 5.16%; United States, 0%.

[4]In order to be able to compare different types of tracking simultaneously in one model, we statistically had to handle type of tracking as a school variable, even though conceptually it is a system variable.

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Bassis, M. S. (1977). The campus as a frog pond: A theoretical and empirical reassessment. *American Journal of Sociology*, *82*, 1318–1326. doi:10.1086/226467

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Yi-Miau, T. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133–180. doi:10.3102/0002831209345157

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

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*, 631–661. doi:10.1037//0022-0663.75.5.631

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*, 781–861. doi:10.1111/j.1468-0327.2007.00189.x

Buchmann, C., & Dalton, B. (2002). Interpersonal influences and educational aspirations in 12 countries: The importance of institutional context. *Sociology of Education*, *75*, 99–122. doi:10.2307/3090287

Buchmann, C., & Park, H. (2009). Stratification and the formation of expectations in highly differentiated educational systems. *Research in Social Stratification and Mobility*, *27*, 245–267. doi:10.1016/j.rssm.2009.10.003

Byrne, B. M., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology*, *78*, 474–481. doi:10.1037//0022-0663.78.6.474

Catsambis, S., Mulkey, L. M., & Crain, R. L. (2001). For better or for worse? A nationawide study of the social psychological effects of gender and ability grouping in mathematics. *Social Psychology of Education*, *5*, 83–115. doi:10.1023/A:1012675523595

Chiu, D., Beru, Y., Watley, E., Wubu, S., Simson, E., Kessinger, R., . . . Wigfield, A. (2008). Influences of math tracking on seventh-grade students' self-beliefs and social comparisons. *Journal of Educational Research*, *102*, 125–135. doi:10.3200/JOER.102.2.125-136

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, *34*, 366–375. doi:10.1037//0022-3514.34 .3.366

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, *82*, 17–31. doi:10.1086/224257

Dupriez, V., & Dumay, X. (2006). Inequalities in school systems: Effect of school structure or of society structure? *Comparative Education*, *42*, 243–260. doi:10.1080/03050060600628074

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review*, *52*, 245–273. doi:10.1086/528764

Duru-Bellat, M., & Suchaut, B. (2005). Organisation and context, efficiency and equity of educational systems: What PISA tells us. *European Educational Research Journal*, *4*, 181–194.

Eckert, C., Schilling, D., & Stiensmeier-Pelster, J. (2006). Einfluss des Fähigkeitsselbstkonzepts auf die Intelligenz und Konzentrationsleistung [The influence of academic self-concept on performance in intelligence and concentration tests]. *Zeitschrift für Pädagogische Psychologie*, *20*, 41–48. doi:10.1024/ 1010-0652.20.1.41

Espenshade, T. J., Hale, L. E., & Chung, C. Y. (2005). The frog pond revisited: High school academic context, class rank, and elite college admission. *Sociology of Education*, *78*, 269–293. doi:10.1177/003804070507800401

Eurydice. (2011). *Eurydice: The information network on education in Europe*. Retrieved from http://ec.europa.eu/education/linkhomepage/eurydice_en.htm

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, *19*, 499–505. doi:10.1016/j.lindif.2009.05.004

Fuligni, A. J., Eccles, J. S., & Barber, B. L. (1995). The long-term effects of seventh-grade ability grouping in mathematics. *Journal of Early Adolescence*, *15*, 58–89. doi:10.1177/0272431695015001005

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology*, *94*, 1146–1183. doi:10.1086/229114

Goldsmith, P. R. (2011). Coleman revisited: School segregation, peers, and frog ponds. *American Educational Research Journal*, *48*, 508–535. doi:10.3102/ 0002831210392019

Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education*, *38*, 32–47. doi:10.1177/ 002248718703800406

Hallinan, M. (1994). Tracking: From theory to practice. *Sociology of Education, 67*, 79–84, 89–91.

Hanushek, E., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal*, *116*, 63–76. doi:10.1111/j.1468-0297.2006.01076.x

Ireson, J., & Hallam, S. (2001). *Ability grouping in education*. London, UK: Sage.

Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction*, *19*, 201–213. doi:10.1037/a0015558

Kelly, S. (2004). Are teachers tracked? On what basis and with what consequences? *Social Psychology of Education*, *7*, 55–72. doi:10.1023/B:SPOE.0000010673.78910.f1

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, *19*, 355–365. doi:10.1016/j.lindif.2008.10.009

Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology*, *75*, 567–586. doi:10.1348/000709905X42239

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

Lucas, S. R. (1999). *Tracking inequality: Stratification and mobility in American high schools*. New York, NY: Teachers College Press.

Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology*, *106*(6), 1642–1690. doi:10.1086/321300

Lucas, S. R., & Berends, M. (2002). Sociodemographic diversity, correlated achievement, and de facto tracking. *Sociology of Education*, *75*, 328–348. doi:10.2307/3090282

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*, 263–285.

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Developmental Perspectives*, *2*, 99–106. doi:10.1111/j.1750-8606.2008.00048.x

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

Marsh, H. W. (1990a). Influences of internal and external frames of reference on the formation of mathe and English self-concepts. *Journal of Educational Psychology*, *82*, 107–116. doi:10.1037//0022-0663.82.1.107

Marsh, H. W. (1990b). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review*, *2*, 77-171. doi:10.1007/BF01322177

Marsh, H. W. (1990c). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept. An interim test manual and a research monograph*. San Antonio, TX: Psychological Corporation.

Marsh, H. W., Chessor, D., Craven, R. G., & Roche, L. (1995). The effect of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal*, *32*(2), 285–319. doi:10.2307/1163433

Marsh, H. W., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science*, *1*, 133–163. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H., Craven, R., & Martin, A. (2006). What is the nature of self-esteem? Unidimensional and multidimensional perspectives. In M. Kernis (Ed.), *Self-esteem: Issues and answers* (*Vol. 1*, pp. 16–25). New York, NY: Psychology Press.

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*, 364. doi:10.1037/0003-066 X.58.5.364

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizibility of the Internal/External Frame of Reference predictions across 26 countries. *Journal of Educational Psychology*, *96*, 56–67. doi:10.1037/0022-0663.96.1.56

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*, 337–349. doi:10.1037//0022-3514.78.2.337

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

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

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*, 107–125. doi:10.1207/s15326985 ep2003_1

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*, 631–669. doi:10.3102/0002831207306728

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

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitidinal data. *Journal of Educational Psychology*, *89*, 41–54. doi:10.1037//0022-0663.89.1.41

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*(4), 705. doi:10.3102/00028312035004705

Martin, A. J., & Debus, R. L. (1998). Self-reports of mathematics self-concept and educational outcomes: The roles of ego-dimensions and self-conciousness. *British Journal of Educational Psychology*, *68*, 517–535. doi:10.1111/j.2044-8279 .1998.tb01309.x

Meyer, J. W. (1977). The effects of education as an institution. *American Journal of Sociology*, *83*, 55–77. doi:10.1086/226506

Mulkey, L. M., Catsambis, S., Steelman, L. C., & Crain, R. L. (2005). The long-term effects of ability grouping in mathematics: A national investigation. *Social Psychology of Education*, *8*, 137–177. doi:10.1007/s11218-005-4014-6

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, L., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence*, *20*, 482–506. doi:10.1111/j.1532-7795.2010.00644.x

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Organization for Economic Cooperation and Development. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science, and problem solving knowledge and skills*. Paris, France: Author.

Organization for Economic Cooperation and Development. (2005a). *PISA 2003 data analysis manual: SAS users*. Paris, France: Author.

Organization for Economic Cooperation and Development. (2005b). *PISA 2003 technical report*. Paris, France: Author.

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-grouping effects: Instructional, social, or institutional. *Sociology of Education, 67*, 27–46. doi:10.2307/2112748

Postlethwaite, T. N. (1995). *International encyclopedia of national systems of education*. Oxford, UK: Pergamon.

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

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, 169*, 805–827. doi:10.1111/j.1467-985X.2006.00426.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

Reuman, D. A. (1989). How social comparison mediates the relation between ability-grouping practices and students' achievement expectancies in mathematics. *Journal of Educational Psychology, 81*, 178. doi:10.1037//0022-0663.81.2.178

Richer, S. (1976). Reference-group theory and ability grouping: A convergence of sociological theory and educational research. *Sociology of Education, 49*, 65–71. doi:10.2307/2112394

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. doi:10.1037//1082-989X.7.2.147

Schiller, K. S., Schmidt, W. H., Muller, C., & Houang, R. T. (2010). Hidden disparities: How courses and curricula shape opportunities in mathematics during high school. *Equity and Excellence in Education, 43*, 414–433. doi:10.1080/10665684.2010.517062

Schwarzer, R., Lange, B., & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel [Self-concept development after a reference-group change]. *Zeitschrift für Entwicklung spsychologie und Pädagogische Psychologie, 14*, 125–140.

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

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist, 37*, 233–244. doi:10.1207/S15326985EP3704_3

Swann, W. B., Chang-Schneider, C., & McClarty, K. L. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist, 62*, 84–94. doi:10.1037/0003-066X.62.2.84

Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A Further examination of the big-fish–little-pond effect. *Sociology of Education, 83*, 333. doi:10.1177/0038040710383521

Trautwein, U., Baumert, J., & Maaz, K. (2007). Hauptschulen = Problemschulen? [Hauptschulen = Problematic schools?]. *Aus Politik und Zeitgeschichte, 28*, 3–9.

Trautwein, U., Köller, O., & Kämmerer, E. (2002). Effekte innerer und äußerer Leistungsdifferenzierung auf selbstbezogene Fähigkeitskognitionen, die wahrgenommene Unterrichtspartizipation und die wahrgenommene soziale Akzeptanz [The effects of within-class and between-class ability grouping on academic self-concepts, reported classroom participation, and perceived social integration]. *Psychologie in Erziehung und Unterricht, 49*, 273–286.

Trautwein, U., Köller, O., Lüdtke, O., & Baumert, J. (2005). Student tracking and the powerful effects of opt-in courses on self-concept: Reflected-glory effects do exist after all. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *New frontiers for self-research* (*Vol. 2*, pp. 307–327). Greenwich, CT: Information Age.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group compostion and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*, 788–806. doi: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*, 853–866. doi:10.1037/a0016306

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, L., & Niggli, A. (2009). Different forces, same consequences: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97*, 1115–1128. doi:10.1037/a0017048

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

Van Houtte, M., Demanet, J., & Stevens, P. A. J. (2012). Self-esteem of academic and vocational students: Does within-school tracking sharpen the difference? *Acta Sociologica, 55*, 73–89. doi:10.1177/0001699311431595

Van Houtte, M., & Stevens, P. A. J. (2009). Study involvement of academic and vocational students: Does between-school tracking sharpen the difference? *American Educational Research Journal, 46*, 943. doi:10.3102/0002831209348789