Journal of Educational Psychology
2006, Vol. 98, No. 4, 788–806

0022-0663/06/$12.00 DOI: 10.1037/0022-0663.98.4.788

Tracking, Grading, and Student Motivation: Using Group Composition and Status to Predict Self-Concept and Interest in Ninth-Grade Mathematics

Ulrich Trautwein and Oliver Lüdtke
Max Planck Institute for Human Development

Herbert W. Marsh
Oxford University and University of Western Sydney

Olaf Köller
Humboldt University

Jürgen Baumert
Max Planck Institute for Human Development

Assigning students to different classes on the basis of their achievement levels (tracking, streaming, or ability grouping) is an extensively used strategy with widely debated consequences. The authors developed a model of the effects of tracking on self-concept and interest that integrates the opposing predictions of "assimilation" and "contrast" effects, which specifies teacher-assigned grades as a major mediating variable, and tested it in 2 settings in which track level is clearly associated with different status—systematic tracking as a function of school type (Study 1, N = 14,341 German 9th-grade students) and separate streams within a comprehensive school system (Study 2, N = 3,243 German 9th-grade students). The results support predictions that students' math self-concept and math interest differ as a function of the achievement of their reference group, their own achievement, and their teacher-assigned grades. No systematic association between track level and math self-concept was found once individual student achievement, school-/stream-average achievement, and teacher-assigned grades were controlled.

Keywords: self-concept, interest, tracking, assimilation, contrast

The educational systems of most industrialized countries around the world use some form of achievement grouping (also known as *tracking* or *streaming*). In many countries, achievement grouping has been the subject of heated political and scientific debate for many years now (see Ireson & Hallam, 2001; LeTendre, Hofer, & Shimizu, 2003; Lucas, 1999). Its critics argue that students in low-achievement tracks are at a relative disadvantage to their counterparts in high-achievement tracks, that they receive lower quality teaching, and develop lower educational aspirations (see Lucas, 1999, for an in-depth analysis). Furthermore, many critics of tracking have argued that being placed in a low-achieving group has negative effects on students' motivation. These critics maintain that sorting students according to their achievement level will lead to assimilation effects that enhance student motivation in higher tracks and undermine it in lower ones (see Oakes, 1985). However, this notion has been questioned by other researchers who have recently documented *negative* effects on self-concept of being placed in a high-achievement group and *positive* effects of being placed in a low-achievement group in several studies (e.g., Marsh & Craven, 2002; Marsh & Hau, 2003). These researchers argue that students tend to use their peers in the class or school as a reference group to form their self-views. Given the same individual achievement, this naturally leads to less favorable comparisons in high-achieving groups and to more favorable comparisons in low-achieving groups (known as the *contrast effect*).

Although there are various forms of differential student placement in the United States (see Lucas, 1999), and although researchers there continue to debate the advantages and disadvantages of different types of tracking (e.g., Plucker et al., 2004), explicit forms of tracking have essentially been abolished. Consequently, researchers wishing to examine the potentially counteracting effects of placement in high- and low-achievement tracks on student motivation have to turn to educational systems that have retained explicit tracking structures. In the present article, we examine the assimilation and contrast hypotheses in a large German sample. Because the German school system is probably the most strictly differentiated in the Western industrialized countries, our sample allows for a conservative test of assimilation effects: If they are not found in this sample, then it is unlikely that they will be found in other less overt forms of ability tracking.

This article adds to the debate on tracking in three ways. First, we argue that the scope for assimilation effects may vary, depending on the form of tracking used (Trautwein, Köller, Lüdtke, & Baumert, 2005), and differentiate between two alternative forms—between-school tracking, in which students are assigned to different school types on the basis of their prior achievement such that students in each school type form a relatively homogeneous group in terms of achievement levels and test scores, and within-school

Ulrich Trautwein, Oliver Lüdtke, and Jürgen Baumert, Center for Educational Research, Max Planck Institute for Human Development, Berlin, Germany; Herbert W. Marsh, Department of Psychology, Oxford University, Oxford, England and SELF Research Centre, University of Western Sydney, Sydney, Australia; Olaf Köller, Institute for Educational Progress, Humboldt University, Berlin, Germany.

Correspondence concerning this article should be sent to Ulrich Trautwein, Center for Educational Research, Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany. E-mail: trautwein@mpib-berlin.mpg.de

tracking, in which relatively heterogeneous groups of students within a school are assigned to different streams within that school on the basis of their prior achievement. Hence, we investigate contrast and assimilation effect from two perspectives—systematic tracking as a function of school type (see Study 1) and separate streams within a comprehensive school system (see Study 2). Second, in both studies, we scrutinize the extent to which assimilation and contrast effects are associated with grading practices. On the basis of prior theorizing, we argue that contrast effects will be mediated heavily by teacher-assigned grades (cf. Marsh, 1987, 1993), and we test this mediation model empirically. Third, we extend prior research on reference group effects on student motivation, which has tended to use academic self-concept as the outcome variable by including students' academic interest as an additional outcome variable.

Aspects of Tracking

Around the world, students are grouped into "clusters" (e.g., streams, schools, or school types) in a variety of ways. Many researchers have argued that placing an individual student in one of these clusters—commonly referred to as *streaming*, *tracking*, or *ability grouping*—will affect her or his achievement, future educational career, morale, and happiness (see Ireson & Hallam, 2001; Lucas, 1999; Oakes, 1985). The nature and extent of tracking differs across countries, states, and/or school districts, making the term somewhat ambiguous. In fact, in some educational structures, students are not even aware of being tracked (Lucas, 1999). For the present study, we use a classification scheme that describes different forms of tracking in terms of three features: the institutional level, the role of achievement, and the impact of placement on future academic careers (Trautwein et al., 2005). First, tracking can occur at a minimum of three institutional levels. Teachers in the early grades often use some form of *within-class achievement grouping*, in which learners progressing at the same rate are grouped together. In secondary schooling, *course-level grouping* is widespread, with students choosing from, or being assigned to, classes working at different levels or covering different content (e.g., regular vs. advanced placement courses, emphasis on humanities vs. emphasis on sciences). Grouping also occurs at the *school level.* Here, it is possible to differentiate between *implicit school-level tracking* on the basis of factors such as area of residence, as is typical in state schools in the United States, and *explicit school-level tracking* to different *school types* catering to specific student groups. Prior achievement is frequently the most important determining factor in schooltype tracking, but academic specializations may also play a role. Thus, school types are typically distinguished by average achievement levels and/or a specific curriculum. Whereas within-class, course-level, and implicit school-level tracking are common in the United States, explicit schooltype tracking is more widespread in countries such as Japan, Taiwan, the Netherlands, and Germany.

A second major feature of tracking is the role of achievement in determining placement in a certain track. If placement is based on prior achievement, then the term *achievement grouping* is most appropriate. If, however, other factors such as students' interests or parental educational goals and financial resources influence the placement, then a term such as *opt-in tracking* may be more appropriate. U.S. high schools frequently feature a mix of these two approaches.

Third, different forms of tracking are characterized by the *impact of placement on future academic careers*. In several Asian countries (e.g., Taiwan, Japan), placement in a low-achievement school type at secondary school level reduces or eliminates a student's chances of obtaining a university degree. In the United States, tracking effects on later educational outcomes are less visible and rigid, although high school students' choices of advanced placement courses are associated with their future educational trajectories.

The three features of tracking described above contribute to the status of a track and, accordingly, to the status of a student (or his or her status in a specific subject). From a theoretical point of view, status differences are most prominent in explicit, highly visible forms of achievement tracking and less prominent in more implicit, less visible forms of tracking or opt-in courses. In the case of explicit tracking programs that use achievement as an entry characteristic and have a profound impact on later educational opportunities, the tracking status of a student is clear to the student, to his or her parents and peers, and to teachers. Explicit tracking often involves labels such as *advanced placement course* or *special education*. Implicit tracking is less visible and less clearly defined; for example, formally equivalent schools may vary considerably in terms of instructional quality. Despite being less visible, if these differences are known to students or parents, then there may well be an awareness of status differences associated with implicit tracking.

Tracking and Student Motivation: Contrast and Assimilation Effects

What effects does tracking have on student motivation? Some argue that explicit tracking in the United States "was designed to sort and pacify students" (see Lucas, 1999, p. 11). According to this view, students were put in various tracks to accustom them to their future positions in society and the economy. In the same vein, many sociologists, social psychologists, and educational researchers assume that ability grouping has positive effects on student motivation in higher tracks and detrimental effects in lower tracks (e.g., Berends, 1994; Oakes, 1985). In his influential work on tracking in the United States, Lucas (1999) reported evidence suggesting that students in low-track classes received low-quality teaching in unsupportive learning climates and cited research indicating that placement in low tracks was associated with less favorable outcomes than placement in high tracks.

A considerable number of psychological studies on ability grouping have focused on students' self-concept as the outcome variable. In recent years, researchers have increasingly begun to differentiate between global evaluations of the self, typically called *self-esteem*, and evaluations of specific abilities or qualities, such as academic self-concept and social self-concept (see Bracken, 1996; Shavelson, Hubner, & Stanton, 1976). The academic self-concept has further been differentiated into domain-specific academic self-concepts. Domain-specific academic self-concepts reflect a person's self-evaluation regarding a specific academic domain or ability. Typical domain-specific academic self-concept items are "I am quite good at mathematics" (math self-concept) and "I have a poor vocabulary" (verbal self-concept). The majority of studies that have probed for reference group effects on self-concept have used academic or domain-specific academic self-concept as the dependent variable (see Marsh & Craven, 2002).

Several social psychologists have argued that self-concept may be enhanced by membership of groups that are positively valued by the individual (Diener & Fujita, 1997; Tesser, 1988). Similarly, Marsh, Kong, and Hau (2000; see also Felson, 1984; Marsh, 1984) have suggested that the academic self-concepts of students in academically selective classes may be enhanced by their basking in the reflected glory of the accomplishments or qualities of other group members. In this sense, placement in a high-achievement group may be expected to positively affect students' global and domain-specific self-concepts by means of "assimilation effects" (see Marsh et al., 2000; Oakes, 1985). Figure 1a provides a graphical illustration of the assimilation effect. As shown, a student's academic self-concept is positively predicted by his or her individual achievement (on average, high-achieving students evidence higher academic self-concepts), but it is also positively affected by membership of a high-status school or track.

However, a second research tradition has postulated that so-called contrast effects weaken the academic self-concepts of students in high-achieving groups. According to this research tradition, accomplishments are evaluated in relation to frames of reference (Festinger, 1954; Marsh, 1987). Marsh (1984; Marsh & Parker, 1984) proposed the big-fish-little-pond effect (BFLPE) to account for the frame-of-reference effects typically observed in educational settings, hypothesizing that students primarily compare their own academic achievement with that of their schoolmates or classmates and use this social comparison information as the basis for their own academic self-concept (see also Schwarzer, Lange, & Jerusalem, 1982). In other words, rather than forming an impression of their own abilities in comparison with a "typical student" of their age, students focus only on their immediate environment. The BFLPE occurs when students with a similar level of academic achievement, as measured by a standardized achievement test, have lower academic self-concepts when they are placed in a high-achieving environment and higher academic self-concepts when they are placed in a low-achieving environment. Thus, according to this model, a student's academic self-concept depends on both his or her own academic accomplishments and those of the other students in the class or school that he or she attends.

Figure 1b provides a graphical depiction of the contrast effect, according to which academic self-concept is positively influenced by individual achievement, but negatively affected by school- or class-average achievement when controlling for individual

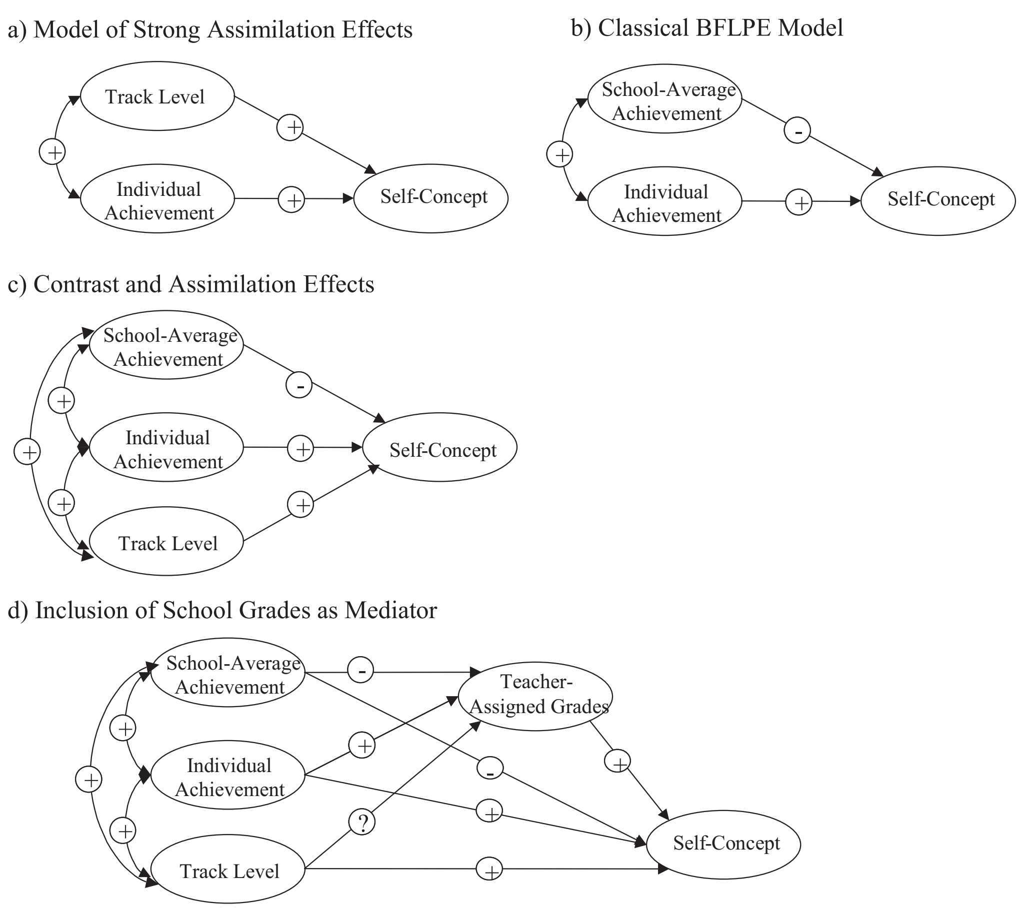


Figure 1. Graphical depiction of (a) a model that postulates strong assimilation effects, (b) the classical big-fish-little-pond (BFLPE) model and (c) and (d) two extended BFLPE models. Plus signs indicate a positive effect; minus signs indicate a negative effect.

achievement. In other words, according to the contrast hypothesis, the same student will have a lower academic self-concept in a school with high average achievement and a higher academic self-concept in a school with low average achievement because students compare and contrast their own achievement with that of their schoolmates or classmates. It is important to note that the contrast hypothesis can only be tested empirically when a data set contains an "objective" measure (i.e., a standardized achievement test) of the relative achievement of all students in the sample on a common metric. Although teacher-assigned grades typically give an accurate estimate of the position of each student within a class, because of differential grading standards they do not typically provide a valid basis for gauging achievement across classes or schools (see The Mediating Role of Teacher-Assigned Grades section for more information on the role of grades in the formation of academic self-concepts).

On the face of it, Figure 1a and 1b would seem to allow straightforward hypotheses to be formulated about the association between track level and school-average achievement, on the one hand, and self-concept, on the other. However, the situation is more complex than it appears at a casual glance: In any given school, assimilation and contrast effects may occur concurrently (Marsh, 1984) so that observed effects are the net effects of these two counter-balancing processes. Because high track level is typically associated with high school-average or stream-average achievement, the assimilation effects of being a member of a high-status group may compete with the contrast effects experienced by students in academically selective schools or streams. In other words, a contrast effect (e.g., "A lot of students are better than me, so I can't be as good as I thought") may counterbalance an assimilation effect (e.g., "I must be smart because I'm in a selective program"). If only one of the two school-level variables (i.e., track level or school-average achievement) is used to predict self-concept, then the beta coefficient of this variable is likely to conflate assimilation and contrast effects. To the extent that the contrast effect associated with track status is stronger than the assimilation effect, for example, the regression coefficient of track status (see Figure 1a) may become negative.

Taken together, the predictive effects of school-average achievement and track level on academic self-concept reflect, to differing extents, the counterbalancing contrast and assimilation effects. To separate these two effects, both track level and school-average achievement need to be considered in a single model, as illustrated in Figure 1c. Marsh et al. (2000) argued that the BFLPE represents the net effects of these two opposing processes. When translated into statistical predictions for the present nonexperimental study, which allows track level and school-average achievement to be separated, the model postulates a complex pattern of results. The negative contrast effect is expected to be reflected in a (negative) regression coefficient of school-average achievement, in particular, whereas the assimilation effect is expected to be stronger for the track level to which the student is assigned, given that tracking is explicit, highly visible, and closely linked to status in the German school system (see related distinctions by Marsh et al., 2000). When controlling for school-average achievement, assimilation effects are expected to be expressed in positive path coefficient for the track-level variable. Because positive assimilation effects of attending the Gymnasium (the highest track school type in the German system) are confounded with negative contrast effects associated with the same variable, the regression coefficient for attending a high-status track should be more positive after partialing out the negative effect of school-average achievement. Hence, attending a high-status track should be associated with high academic self-concept (an assimilation effect) once any contrast effects associated with membership of a high-achieving group are controlled for. Similarly, because negative contrast effects of school-average achievement are confounded with positive assimilation effects associated with the same variable, the regression coefficient for school-average achievement should be more negative after partialing out the positive assimilation effect of attending a high-track school. Hence, when both school-average achievement and school type are considered simultaneously, a mutual suppression effect is expected.

Empirical Evidence for the BFLPE

Most educational research has found a negative regression coefficient of school- or class-average achievement, as measured by standardized achievement tests on academic self-concept (e.g., Lüdtke, Köller, Marsh, & Trautwein, 2005; Marsh & Hau, 2003; Marsh, Köller, & Baumert, 2001; see review by Marsh & Craven, 2002), congruent with the interpretation that the net effect of the opposing contrast and assimilation effects is negative in these studies. Marsh and Hau (2003) conducted a large cross-cultural test of the BFLPE using data from the Programme for International Student Assessment (PISA; Organisation for Economic Cooperation and Development [OECD], 2001). Nationally representative samples of approximately 4,000 students from each of the 26 participating countries (total $N = 103{,}558$ students in 3,851 schools) completed standardized achievement tests and a self-concept questionnaire. Consistent with a priori predictions, the predictive effects of individual student achievement were substantial and positive, whereas the regression coefficients for school-average achievement were negative.

In the study by Marsh and Hau (2003), the mean achievement of a group was analyzed, but tracking information from the different countries—if available at all—was not used. Thus, the study only partially addresses the issue of explicit tracking. The same holds for several other investigations of the BFLPE. Especially in the U.S. context, schools are ostensibly of the same "type" and open to the whole student population. In fact, their average achievement levels differ widely. In other words, there is implicit tracking at the school level. In such cases, assimilation effects—if they exist—can only be detected if additional measures of school prestige are used or if positive assimilation effects are stronger than negative contrast effects.

Some researchers dealing with different school systems have been able to take explicit tracking information into account. Schwarzer et al. (1982) examined students in the German school system, in which students are tracked according to their achievement at about age 10, in a longitudinal study observing the development of academic self-concept after transition to secondary school. In line with the contrast effect hypothesis, the academic self-concept of high-achieving students tended to decrease after transition to the tracked secondary schools, whereas the academic self-concept of low-achieving students tended to increase. Rheinberg and Enstrup (1977) compared the academic self-concept, test anxiety, and achievement motivation of 165 students with mild to moderate learning disabilities ($70 < IQ \leq 85$). Those who attended special schools were found to have higher academic self-concepts and achievement motivation and lower test anxiety than those who attended regular schools. This study supports the predictions of the contrast hypothesis (see also Marsh, Tracey, & Craven, in press; Tracey, Marsh, & Craven, 2003). Unfortunately, Schwarzer et al. and Rheinberg and Enstrup did not try to separate the effects of school type and school mean achievement levels statistically.

The Mediating Role of Teacher-Assigned Grades

As noted above, research on contrast and assimilation effects typically uses standardized achievement tests as indicators of students' underlying level of achievement. Crucially, these tests can be used to compare the achievement of students in different classrooms, an essential prerequisite for testing the assumptions of the BFLPE.

However, achievement on a standardized achievement test is only one of the possible operationalizations of students' underlying level of achievement. Teacher-assigned grades are another indicator of achievement. It is interesting to note that teacher-assigned grades have rarely been considered in research on the BFLPE, despite their potential relevance (for an exception, see Marsh, 1987). Clearly, teacher-assigned grades cannot replace standardized achievement tests in research on the BFLPE unless they are externally moderated relative to a common achievement test such that teacher-assigned grades in different schools vary along a common metric. Nevertheless, classroom grades are likely to have a substantial effect on domain-specific academic self-concept, and indeed to be affected by self-concept (e.g., Trautwein, Lüdtke, Köller, & Baumert, 2006; see also Marsh & Craven, 1997). Numerous studies have shown that grades are of pivotal importance to students' academic self-concepts (see Hansford & Hattie, 1982; Skaalvik & Skaalvik, 2002). Moreover, grades have direct effects on students' academic careers, whereas low-stakes tests, as typically used in research on the BFLPE, have no direct implications for students' present or future educational goals.

It is plausible to expect a positive, but by no means perfect, correlation between the two indicators of achievement. In many educational systems, the majority of teachers do not use an absolute criterion for achievement (as is the case in standardized achievement tests) when assigning grades. Rather, they primarily grade on a norm-referenced basis (Hodge & Coladarci, 1989; Ingenkamp, 1971; Marsh, 1987), with the best student in the class receiving an A or a B grade and the weakest student a D or an E grade. The same applies to Germany (e.g., Baumert, Trautwein, & Artelt, 2003), where "grading-on-a-curve" effects can be observed in most schools. As in many other countries, alternative grading practices (e.g., approaches based strictly on a priori criteria) in which grades from different teachers vary along a common underlying metric are exceptions to the rule.

What role do teacher-assigned grades play in the context of reference group effects on self-concept? For the contrast effect postulated by the BFLPE model to occur, students must base their evaluation of their own achievement primarily on a comparison with their classmates or schoolmates (i.e., they must primarily use information about their relative standing in their class or "pond"). For the assimilation effect to occur, students must have and incorporate information about the achievement level or status of their group (i.e., the relative standing of their group). In this article, we propose an extended BFLPE model that includes teacher-assigned grades as an important mediator variable. The central assumption here is that these classroom grades provide students with easily accessible, reliable, clear, and relevant feedback information about their relative standing within their group, but they are less indicative of the overall standing of the group, and that students are likely to base their self-evaluations on this information, which should thus produce or reinforce the contrast effect.

The extended BFLPE model is illustrated in Figure 1d. According to this model, a student's academic self-concept is closely related to his or her teacher-assigned grades. These grades, in turn, are postulated to be affected by both individual achievement and school-average achievement. Assuming that teachers assign grades on a norm-referenced basis, with the class or school providing the frame of reference, an individual student's teacher-assigned grade will reflect his or her own achievement relative to the achievement of the other students in the same class or school. Accordingly, contrast effects (which are assumed to be the central mechanisms underlying the BFLPE) may readily apply to teachers' grading of students' achievement. In terms of statistical predictions, a negative regression coefficient is expected for school-average standardized achievement on teacher-assigned grades when controlling for individual standardized achievement. In this sense, teacher-assigned grades are likely to be a key mediator variable intervening between a student's underlying level of achievement (as indicated by his or her score on a standardized test) and his or her membership of a specific reference group (as indicated by school-average scores on standardized tests and track status), on the one hand, and his or her academic self-concept, on the other.

What predictions can be made for the association between standardized achievement and track level, on the one hand, and academic self-concept, on the other, when teacher-assigned grades are incorporated into the model? Because teacher-assigned grades represent a main source of performance feedback for students, they may statistically mediate much of the impact of the underlying level of achievement on academic self-concept. At the same time, the proposed model postulates a significant effect of school-average achievement and individual achievement, even after controlling for teacher-assigned grades. This assumption is based on the observation that students use various sources of social comparison information to form their self-concept (Skaalvik & Skaalvik, 2002) and on the finding that the reliability of teacher-assigned grades tends to be moderate, thus leaving room for the effects of other variables. Given the lack of prior research or a coherent theoretical rationale, we do not make specific predictions in regard to the association between track level and teacher-assigned grades when controlling for school-average achievement.

Extending the BFLPE: Effects on Interest

Some researchers have recently questioned the educational importance and implications of the BFLPE, arguing that BFLPE studies focus almost exclusively on academic self-concept as the outcome variable and that their operationalization of self-concept is too narrow (e.g., Plucker et al., 2004; Rindermann & Heller, 2005; see also Trautwein & Lüdtke, 2005). For instance, Plucker et al. (2004) stated that research on the BFLPE "does not speak to 'other psychological benefits' or deficits, only to decreases in self-concept of an indeterminate length" (p. 268). Indeed, most studies on the BFLPE use academic self-concept as the dependent variable (for a notable exception, see Marsh, 1991). In the present study, we aim to extend prior research by probing for direct and indirect reference group effects on students' interest. We believe that research into reference group effects on students' domain-specific interest is of great relevance to educational theory and practice. It is important to note that interest has always been a key variable in the debate about ability grouping (e.g., Ireson & Hallam, 2001; Oakes, 1985). In fact, proponents of ability grouping typically argue that it is a means of enhancing the academic interests of high-achieving students (e.g., Plucker et al., 2004) and—in turn—their academic achievement. Moreover, interest has been shown to have a huge impact on achievement-related decisions (e.g., Eccles & Wigfield, 2002; Trautwein, Lüdtke, Kastens, & Köller, 2006).

It has become increasingly common to differentiate two forms of interest (see Hidi & Harackiewicz, 2000; Krapp, 2000; Schraw & Lehman, 2001): situational interest and individual (or personal) interest. Situational interest can be characterized as transient, context-dependent enjoyment that is triggered by environmental factors. It is an important—and sometimes difficult—task for teachers to design learning environments that allow situational interest to develop. Situational interest is often a necessary first step in the development of more stable individual interest (Hidi & Renninger, 2006; Krapp, 2000).

In contrast to situational interest, individual or personal interest is hypothesized to be a relatively enduring predisposition to attend to certain objects and activities (Hidi & Ainley, 2002; Köller, Baumert, & Schnabel, 2001; Krapp, 2000; Schraw & Lehman, 2001). For instance, individual math interest reflects an ongoing, rather stable affinity for math. There is a theoretical distinction between the feeling-related (also called *intrinsic value*) and the value-related (*attainment value/commitment-related*) components of personal interest (see Eccles et al., 1983; Krapp, 2000), but researchers have often been unable to distinguish between these components empirically (Köller et al., 2001; see also Schraw & Lehman, 2001). When a student engages in an interest-related activity (i.e., when his or her individual interest is actualized), the activity is typically associated with positive affect, persistence, and favorable learning outcomes. Actualized interest-driven activities are characterized by the experience of competence and personal control, feelings of autonomy and self-determination, positive emotional states and, in the best-case scenario, by an experience of flow in which the person and the object of interest merge (Csikszentmihalyi & Schiefele, 1993).

In the present research, we focus on individual rather than on situational interest. We address the question of how reference group effects are associated with the level of individual interest. There seem to be at least two plausible associations. First, as in the case of domain-specific self-concepts, assimilation effects may be associated with negative effects on the interest of students in low-achieving tracks. Some critics of tracking (e.g., Oakes, 1985) claim that tracking may convey the message to low-track students that they are not expected to excel or to engage in "academic" subjects such as mathematics. If this hypothesis is correct, then low-track status (but not necessarily low school- or class-average achievement) should be associated with low interest, even when controlling for individual achievement.

Second, individual interest may be substantially influenced by academic self-concept (see Eccles et al., 1983; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). To the extent that there are negative reference group effects on self-concept, similar effects may be expected on interest. Several authors have proposed that academic self-concept affects interest (e.g., Krapp, 2000; Köller et al., 2001). For instance, Baumert, Schnabel, and Lehrke (1998) suggested that the effect of achievement on interest may be mediated by academic self-concept. Likewise, in their original expectancy-value model, Eccles et al. (1983) hypothesized academic self-concept to affect expectations of success and task value directly and to affect achievement-related choices indirectly via its influences on expectancy beliefs and task value. Empirical research (e.g., Marsh et al., 2005; Wigfield et al., 1997) has supported the assumption that self-concept has positive effects on interest. If this line of reasoning is correct, then a similar regression coefficient should result if either academic self-concept or individual interest are regressed on achievement variables. Moreover, the predictive effect of individual and school-average standardized achievement on interest may be strongly mediated by academic self-concept.

The Present Investigation

In recent years, a wealth of empirical studies have explored frame-of-reference effects on academic self-concept (Marsh & Craven, 2002; Marsh & Hau, 2003). The empirical evidence suggests that—when controlling for individual achievement in a standardized test—self-concept development is more favorable if students are placed in low-achieving environments than in high-achieving environments. This has been interpreted as support for the existence of contrast effects (see Marsh & Craven, 2002). Much less research (but see Marsh et al., 2001, 2000) has examined the role of assimilation effects.

The present study aims at extending the existing knowledge base about tracking and the role of assimilation and contrast effects in three important respects. First, prior studies have found only scant evidence for assimilation effects. This may, however, have been because there were no marked status differences between the tracks in question. With its large samples of students in two forms of tracking, both of which involve clear differences in status, the present study permits a strong test of assimilation effects. In terms of our classification of tracking, outlined above, both forms of tracking are explicit. Because tracking is highly visible in Germany and the different tracks have very different prestige, we expected to find assimilation effects running counter to the contrast effects of school-average achievement. Empirically, this should be reflected in negative regression coefficients of school-average achievement and positive regression coefficients of track status when predicting academic self-concept.

Second, prior research has drawn attention to the importance of teacher-assigned grades (e.g., Trautwein, Lüdtke, Köller, & Baumert, 2006) in the development of academic self-concept. Although it is widely accepted that teacher-assigned grades contribute to establishing frame-of-reference effects (Skaalvik & Skaalvik, 2002), they have rarely been considered in this type of research (see Trautwein & Lüdtke, 2005). In all likelihood, the inclusion of grades will help to predict a student's domain-specific self-concept. Therefore, the effects of teacher-assigned grades are very relevant to research on assimilation and contrast effects. As illustrated in Figure 1d, we expected to find that grades substantially mediate the predictive effects of school-average achievement and individual achievement on academic self-concept.

Third, prior research on the BFLPE has tended to focus on academic self-concept as the main dependent variable, and several authors have called for tests of the broader implications of the BFLPE (e.g., Plucker et al., 2004; Trautwein & Lüdtke, 2005). In the present study, we introduced individual interest as a second dependent variable. We tested two opposing hypotheses on how frame-of-reference effects influence individual interest. The assimilation hypothesis predicts that low-achieving students lose interest in typical "academic" subjects when placed in a nonacademic track. On the basis of this hypothesis, a negative effect of track status on interest must be expected. However, given the close relationship between academic self-concept and individual interest, one may expect to find similar frame-of-reference effects on interest as on academic self-concept. On the basis of this reasoning, frame-of-reference effects on interest should decrease markedly or disappear altogether when controlling for academic self-concept.

Study 1

The German school system is well-known for its early and selective differentiation of students in different school types. Selection takes place after Grade 4 (in a few states after Grade 6), when students are about age 10. Although there is considerable variation across the German states in terms of the number and quality of tracks (Baumert et al., 2003), the "tripartite system" of "Hauptschule," "Realschule," and "Gymnasium" is most widespread. Hauptschule is the academically least demanding track, Realschule the intermediate track, and Gymnasium the highest track. Hauptschule students graduate after Grades 9 or 10 and then enter the dual system, which combines part-time education at vocational school with on-the-job training. Realschule students graduate after Grade 10; most of them also enter the dual system, usually aspiring to more skilled occupations than Hauptschule graduates. Gymnasium students graduate after Grades 12 or 13. A pass in the final Gymnasium examination ("Abitur") is a prerequisite for university entrance, but some of the more attractive jobs in the dual system (e.g., management positions in banks, high-level police officers) also require this qualification. Comprehensive secondary schools ("Gesamtschulen") play a minor role in Germany. They are the focus of Study 2.

To return to our classification scheme outlined above, the *tripartite system* can be defined as (a) a form of school-type tracking that (b) is based on achievement differences and (c) has profound implications for student careers. The status of the tracks differs markedly; Gymnasium has the highest status, followed by Realschule and then Hauptschule. Therefore, if assimilation processes take place at the school-type level, then the tripartite system should entail disadvantages for Hauptschule students and advantages for Gymnasium students (relative to Realschule students) in terms of academic self-concept.

In Study 1, we used a large sample of ninth-grade students to examine assimilation and contrast effects. Three hypotheses were tested. First, we assumed that assimilation processes at the level of school type may counteract the contrast effects typically found in BFLPE research at the between-school level. To test this assumption empirically, we simultaneously included both individual- and school-level math achievement and track level in the analyses. We expected to find negative regression coefficients for school type and school-average math achievement on math self-concept when just one of these variables was used as a predictor variable in addition to individual achievement; such a pattern of results could be interpreted as support for the contrast effect postulated by the BFLPE model. Given that school-average math achievement and track level are likely to be substantially correlated, however, we also hypothesized that, once the effect of school-average achievement was controlled, students in high-status school types would evidence a higher math self-concept (which would yield support for the assimilation hypothesis), whereas the negative regression coefficient of school-average achievement (the assumed contrast effect) would become even more negative. Second, we hypothesized that teacher-assigned grades would at least partly statistically mediate the predictive effects of both individual and school-average math achievement on math self-concept (see Figure 1d). Third, we expected to find frame-of-reference effects to be similarly related to math interest as they are to math self-concept. Furthermore, based on prior theory and research (e.g., Eccles et al., 1983; Marsh et al., 2005), we assumed that in a mediator analysis, math self-concept would mediate much of the predictive power of achievement and track status on math interest.

Method

Sample

The analyses presented are based on data from the German extension (Baumert et al., 2002) to the year 2000 cycle of the PISA study (OECD, 2001). The main goal of this extension was to make it possible to analyze and compare the results of the 16 German states. To this end, data were collected from an expanded, nationally representative sample of 34,765 ninth-grade students rather than from 15-year-olds (who formed the sample for the international PISA study). A multistage sampling procedure was implemented to ensure high representativity of the data, and participation rates of 85% and above were achieved in all states. All classes were tested on two consecutive days in May and June 2000.

As mentioned above, the implementation of the traditional tripartite system varies across the German states. In some states, for instance, very few students attend Hauptschule. Because the juxtaposition of tracking at the *school-type level* and *school-level achievement* is central to the present study, only students from the eight states with a high proportion of students ($> 20\%$) in each of the three school types were included. Furthermore, only schools that do not use within-school tracking were included. The final sample thus consisted of 14,341 students from 621 schools (49.8% girls).

Instruments

Standardized math achievement test. The national math achievement test implemented in the German extension to PISA, 2000, was successfully tailored (see Klieme, Neubrand, & Lüdtke, 2001) to students in German schools and had high curricular validity. It was administered to all students in the sampled population. Math achievement scores for individual students were generated using item response technique. The resulting test score distribution had a mean of 100 ($SD = 30$); the reliability of the test was .89 (formula by Rost, 1996). Various content areas—arithmetic, geometry, elementary statistics, and algebra—were covered. As in the International Association for the Evaluation of Educational Achievement's third international mathematics and science (TIMS) study (Beaton et al., 1996), the items were assigned to different performance categories (e.g., applying knowledge, solving mathematical problems). Although the content areas and performance categories of the items varied, analyses based on item response theory confirmed that a unidimensional model was appropriate for describing the latent variable underlying the test results. The correlation between students' scores on the German national math assessment and the international math assessment was .85. A similar curriculum is used in all three school tracks, but the level of difficulty varies.

Teacher-assigned grades. Students reported the mathematics grades they had obtained on their midterm report card (in February 2000) using the six-level grading system implemented throughout Germany. We reverse coded these grades, resulting in the following six rating levels: *excellent* (6), *good* (5), *satisfactory* (4), *sufficient* (3), *poor* (2), and *very poor* (1).

Mathematics self-concept. The math self-concept instrument administered in the PISA study (OECD, 2001) consists of items from the Self-Description Questionnaire (SDQ; Marsh & O'Neill, 1984), which is considered to be one of the best self-concept instruments available (Byrne, 1996; see also Marsh, 1990, 1993; Marsh & Craven, 1997). Because of limitations on the length of the overall instrument, the three best items from the 10-item Math Self-Concept scale were selected by OECD statistical experts after extensive pilot testing. The internal consistency of this scale was consistently reasonable across all 26 countries, and was high in our sample ($\alpha = .90$). In order to test the representativity of the three items, we used the SDQ archive data ($N = 2{,}436$; see Marsh & O'Neill, 1984) to compute a part-whole correlation between the three selected items and the complete set of SDQ II math self-concept items. The correlation was .94, demonstrating that the three items ("I have always done well in mathematics"; "I get good marks in mathematics"; "Mathematics is one of my best subjects") used in the present study are a good representation of the full scale (for similar conceptualizations of math self-concept, see Harter, 1985, and Wigfield & Eccles, 2002).

Personal interest in mathematics. The PISA 2000 Interest scale consists of three items tapping a rather stable, enduring disposition to engage in math-related activities. Hence, the measure focuses on personal interest and not on situational interest. Two items focus on the feeling-related, affective quality (see Krapp, 2000) or intrinsic value (Eccles & Wigfield, 2002) of engaging in math-related activities ("When I do mathematics, I sometimes get totally absorbed"; "Because doing mathematics is fun, I wouldn't want to give it up"). The third item taps the personal importance (Krapp, 2000) or attainment value (Eccles & Wigfield, 2002) of engaging in math ("Mathematics is important to me personally"). The scale is a modified version of a domain-specific interest scale that has been successfully administered in several school achievement studies (Baumert et al., 1997; Köller et al., 2001; Marsh et al., 2005; Trautwein, Lüdtke, Kastens, & Köller, 2006). Like the Self-Concept scale, the Interest scale was subjected to intensive pilot testing and proved to be reliable across countries and for high- and low-achieving students. The scale exhibited sufficient internal consistency in our sample (Cronbach's $\alpha = .76$).

School type. Dummy variables were created to examine the effects of school type. The middle track (Realschule) was used as the reference category.

Statistical Analyses

We conducted multilevel regression analyses to predict mathematics self-concept and interest. In most studies conducted in school settings, individual student characteristics are confounded with classroom or school characteristics because individuals are not randomly assigned to groups. This clustering effect introduces problems related to appropriate levels of analysis, aggregation bias, and heterogeneity of regression (Raudenbush & Bryk, 2002). In the present investigation, for example, the meaning of a variable at the student level does not necessarily bear any straightforward relation to its meaning at the classroom level. The negative BFLPE is a dramatic example of this problem—achievement at the individual level is positively related to academic self-concept, whereas achievement at the school- or class-average level may be unrelated or negatively related to academic self-concept. The juxtaposition of the effects of individual achievement and class-average achievement is inherently a multilevel issue that cannot be represented adequately at either the individual or the classroom level. Particularly when major variables represent different levels, it is important to use appropriate multilevel statistical procedures for data analysis. Multilevel modeling, a special form of regression analysis, provides a powerful methodology for handling hierarchical data of this kind. A detailed presentation of multilevel modeling (also known as hierarchical linear modeling [HLM]) is beyond the scope of the present investigation and is available elsewhere (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

In the present study, all multilevel analyses were computed with the computer program HLM 5 (Raudenbush, Bryk, Cheong, & Congdon, 2000). HLM 5 does not report standardized regression coefficients. In order to enhance the interpretability of the regression coefficients produced, we standardized ($M = 0$, $SD = 1$) all continuous variables before performing the multilevel analyses (see Marsh & Rowe, 1996). Dichotomous variables were retained in their original metric. Academic achievement was aggregated at the school level to form an index of the overall level of achievement in the school and was not restandardized; thus, school-average achievement effects are presented in the metric of individual student achievement tests. All models reported are random-intercept models estimated by restricted maximum likelihood.

Missing data represent a potentially serious methodological problem in many empirical studies. For each of the items and scales considered here, the percentage of missing data was below 10%, but listwise deletion would have reduced the total sample size by 18%. In the methodological literature on missing data (Little & Rubin, 1987; Schafer, 1997), there is growing consensus that multiple imputation of missing data is superior to traditional pairwise and listwise deletion methods. Hence, we opted for the multiple imputation procedure (Schafer, 1997). The NORM software (Version 2.03, see Schafer & Graham, 2002) was used to generate five data sets in which all missing data were replaced by estimated values. All subsequent statistical analyses were conducted separately for each of the five data sets. Parameters and their standard errors were then combined, using procedures described by Schafer and Graham (2002), to calculate overall estimates and standard errors that take into account the uncertainty of missing data.

Results

We first present results of descriptive analyses in which the intraclass correlation coefficients for math achievement, math grades, math self-concept, and math interest were calculated, and group mean differences in these variables between high-, middle-, and low-track students were established (see Table 1). The high intraclass correlation coefficient for math achievement shows that schools in the present sample differed substantially according to their achievement levels. This is not surprising, of course, as prior achievement was the primary basis for students being allocated to different tracks after Grade 4. The mean of the middle track was approximately one standard deviation higher than the mean of the lower track; likewise, the mean score of students in the upper track was more than one standard deviation higher than that of their peers in the middle track (both differences significant at $p < .001$). When school type was introduced to the HLM analysis, 77% of the Level 2 variance was explained. This finding indicates that the average achievement across school types differs widely; however,

Table 1
Descriptive Statistics for Study 1 Variables

		Hauptschule (lower track)		Realschule (middle track)		Gymnasium (upper track)	
Variable	ICC	*M*	*SD*	*M*	*SD*	*M*	*SD*
Math achievement	0.63	79.73	17.87	98.71	18.32	120.98	18.39
Math school grade	0.05	4.07	1.09	3.82	1.09	3.94	1.07
Math self-concept	0.02	2.52	0.90	2.45	0.98	2.48	0.99
Math interest	0.02	2.48	0.78	2.37	0.85	2.33	0.86

Note. ICC = Intraclass correlation (estimated in a baseline model using HLM 5). Math school grades are reverse scored so that higher values represent higher achievement.

the residual variance at Level 2 also indicates that there are meaningful differences between schools of the same type. The differences in achievement within and between school types are illustrated in Figure 2, which depicts average achievement in the standardized achievement test in the 621 participating schools. As can be seen, although school type accounts for a considerable amount of the between-school difference, there are still clear differences among schools of the same school type.

With respect to teacher-assigned grades, an intraclass correlation of .05 was found in the baseline model. This result indicates that most of the variance in grades was located within schools, not across schools, and indirectly supports the notion that grading occurs "on the curve" and does not vary along a metric that is common to all schools. In fact, the average grade assigned in the lower track was higher than in the middle track, $t(618) = 7.38$, $p < .001$ (the statistical test was performed in HLM 5 to account for the hierarchical data structure) and the upper track, $t(618) = 3.40$, $p < .01$, but higher in the upper track than in the middle track, $t(618) = 4.18$, $p < .01$. Thus, differences in grades were not monotonic with respect to track. Taken together, students were primarily graded according to norm-referenced criteria, with students of the same track forming the reference group; additionally, slight school-type differences at the mean level were observed.

Students' math self-concept and math interest were close to the midpoint of the scale (2.50). The intraclass correlation of the two scales was very small (less than .05; see Table 1), and school-type differences did not exceed .10 standard deviations. Math self-concept and interest at the individual level were substantially correlated ($r = .67$, $p < .05$).

We next performed hierarchical linear modeling with teacher-assigned grades as the dependent variable (see Table 2). Individual and school-average math achievement were introduced in the first model (Model M1). As expected, individual math achievement predicted the teacher-assigned grade of the individual student, whereas—once individual achievement was controlled—a negative regression coefficient was found for school-average achievement. In other words, given the same level of achievement, students' grades were likely to be higher in schools in which the average achievement was low than in high-achieving schools. In the second model (M2), school-average achievement was replaced by the track dummy variables (Hauptschule, Gymnasium). This analysis showed that students whose performance on the standard-

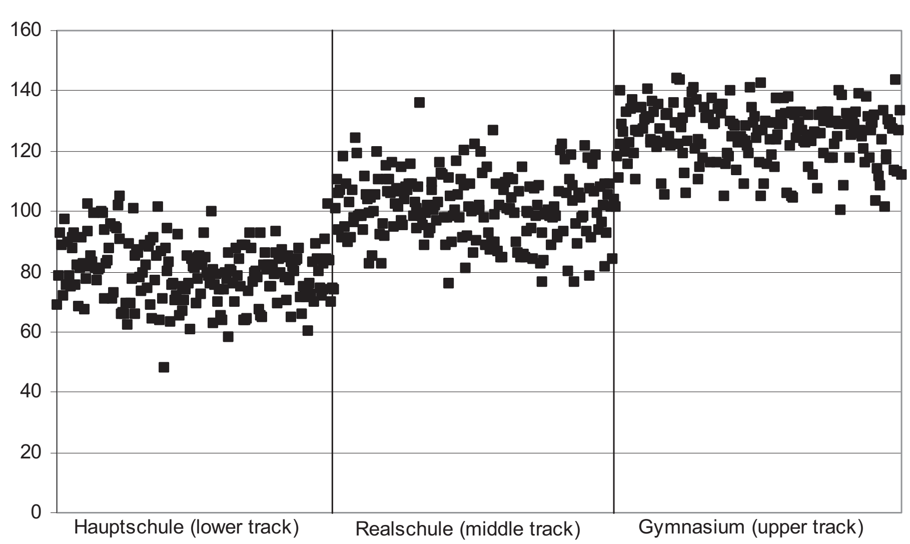


Figure 2. Achievement differences within and between school types: The mean school-average achievement in the 621 participating schools in Study 1.

Table 2
Regressing School Grades on Achievement and Track Level: Results From Multilevel Modeling (in Study 1)

	M1		M2		M3	
	b	*SE*	*b*	*SE*	*b*	*SE*
Intercept	.00	.01	−.04	0.03	.23***	.03
Hauptschule			.80***	0.04	.33***	.04
Gymnasium			−.55***	0.04	−.01	.04
M-Ach school	−.78***	.02			−.64***	.03
M-Ach individual	.78***	.01	.72***	0.01	.78***	.01
Residual variance						
Level 2	.06		.10		.05	
Level 1	.72		.72		.72	

Note. School achievement and school grades were standardized ($M = 0$, $SD = 1$) at the individual level before data analyses and aggregation. Hauptschule = lower track; Gymnasium = upper track (reference category: Realschule = middle track). M-Ach school = school average math achievement in standardized math test; M-Ach individual = individual math achievement in standardized math test.
*** $p < .001$.

ized achievement test was comparable were assigned rather different grades depending on the type of school they attended. Relative to Realschule students with the same math achievement score, the math grades of Gymnasium (high-track) students were 0.55 standard deviations lower, and those of the Hauptschule (low-track) students were 0.8 standard deviations higher.

It is quite possible that assimilation and contrast effects are confounded in either one of the school-level variables (average math achievement and school type) in M1 and M2. Hence, M1 and M2 yield the uncorrected total predictive effect for the two school-level variables. To obtain the net effects of the two variables, they have to be considered simultaneously. This was done in Model 3 (M3). The predictive power of school-average achievement decreased only slightly (from $b = -.78$ to $b = -.64$). The Gymnasium effect, however, vanished when controlling for school-average achievement. It is interesting to note that there was a significant positive effect of Hauptschule, indicating that—when controlling for achievement at the individual and school level—Hauptschule students received higher grades than Realschule students. Taken together, although the grading level is somewhat higher in Hauptschule, the analyses with teacher-assigned grades as the dependent variable confirm that, on the whole, schools tend to grade on a curve.

In the next step, multilevel models were specified, with math self-concept as the dependent variable. Results are reported in Table 3. In M1, we followed the classical approach of BFLPE research and included only two predictor variables: individual math achievement and school-average math achievement. As expected, whereas individual achievement positively predicted math self-concept, school-average achievement evidenced a negative regression weight, indicating that, given the same math achievement, students placed in a high-achieving environment had lower math self-concepts.

We also examined a model in which Hauptschule and Gymnasium were introduced in the form of two school-type dummy variables (M2) in place of school-average achievement. Theoreti-

Table 3
Regressing Math Self-Concept and Math Interest on Achievement, Track, and School Grades (in Study 1): Results From Multilevel Modeling

	Self-concept								Interest									
	M1		M2		M3		M4		M1		M2		M3		M4		M5	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	.00	.01	.04	.03	.12***	.03	.00	.02	.00	.01	.03	.02	.15***	.03	.06*	.02	−.00	.01
Hauptschule			.65***	.04	.15***	.03	−.02	.02			.52***	.03	.17***	.03	.05	.03	.06**	.02
Gymnasium			−.61***	.04	−.05	.04	−.04	.03			−.48***	.03	−.08*	.03	−.08*	.03	−.05*	.02
M-Ach school	−.76***	.02			−.68***	.03	−.35***	.02	−.62***	.02			−.51***	.03	−.28***	.03	−.02	.02
M-Ach individual	.77***	.01	.70***	.01	.77***	.01	.37***	.01	.55***	.01	.47***	.01	.55***	.01	.27***	.02	−.00	.01
M-Grade individual							.51***	.01							.36***	.01	−.02**	.01
Math self-concept																	.74***	.01
Residual variance																		
Level 2	.03		.09		.03		.01		.03		.05		.02		.02		.01	
Level 1	.76		.76		.76		.57		.86		.87		.86		.77		.46	

Note. Math self-concept, school achievement, and school grades were standardized ($M = 0$, $SD = 1$) at the individual level before data analyses and aggregation. Hauptschule = lower track; Gymnasium = upper track (reference category: Realschule = middle track); M-Ach school = school average math achievement in standardized math test; M-Ach individual = individual math achievement in standardized math test; M-Grade individual = teacher-assigned math grade.
* $p < .05$. ** $p < .01$. *** $p < .001$.

cally, the school-type dummies represent both contrast effects (students in Gymnasium [Hauptschule] encounter comparably high- [low-] achieving peers) and assimilation effects (membership of a high- [low-] status track). The positive regression coefficient found for Hauptschule and the negative regression coefficient found for Gymnasium support the assumption that the contrast effects were much stronger than any assimilation effects. What is of interest is that when controlling for individual achievement, the Hauptschule ($b = .65$) and Gymnasium ($b = -.61$) effects were of similar magnitude, but in opposing directions. Moreover, as indicated by the reduction of the residual variance—a common indicator of explained variance in multilevel models (see Raudenbush & Bryk, 2002)—M2, which included school types, explained considerably less variance than M1, which included school-average achievement as the Level 2 variable.

In the next model, M3, both the school-average achievement and the Hauptschule and Gymnasium dummy variables were included. In line with the hypothesis that being placed in a high-achieving group engenders contrast effects, the regression coefficient of school-average achievement was significantly and considerably negative. Somewhat unexpectedly, however, the lower track still had a *positive* and the upper track a *negative* regression coefficient, although these were considerably reduced by the inclusion of school-average achievement. This finding clearly contradicts the assumption that, once individual and school-average achievement are controlled, domain-specific self-concepts are lower in students attending a low-prestige school type and higher in students attending the high-prestige Gymnasium.

Teacher-assigned math grades were entered in Model 4 (M4). As predicted, math self-concept was more closely associated with individual grades than with individual achievement. This finding provides support for our mediation hypothesis. The regression coefficient of both individual and school-average achievement remained significant and of meaningful magnitude, although the size of the respective regression weights diminished considerably. We formally tested (see Sobel, 1982) the indirect, mediated predictive effect of school-average achievement on math self-concept within a multilevel approach (see Krull & MacKinnon, 2001). In support of our mediation hypothesis, the mediated predictive effect of school-average achievement on self-concept was significant ($-.64 \times .51 = -.33$; $p < .001$). Likewise, the predictive effect of individual achievement on self-concept was also mediated by grades ($.78 \times .51 = .40$; $p < .001$). It is important to note, however, when controlling for teacher-assigned grade, the regression coefficient for Hauptschule was no longer statistically significant. More specifically, the predictive effect of this variable was significantly and fully (see Baron & Kenny, 1986) mediated by grades. This indicates that the somewhat harsher grading practices at Gymnasium and Realschule may have affected domain-specific math self-concept. As indicated by the drop in residual variance at both Level 1 and Level 2 (see Table 3), the inclusion of teacher-assigned grades in M4 contributed substantially to the prediction of math self-concept.

We repeated the set of analyses, with math interest as the dependent variable. As shown in Table 3, the results closely mirrored the findings for math self-concept, although the size of the resulting regression coefficients was smaller. Again, the negative predictive effect of school-average achievement ($-.64 \times .36 = -.23$; $p < .001$) and the positive predictive effect of individual achievement ($.78 \times .27 = .21$, $p < .001$) on interest were partly mediated by grades. However, two differences were observed. First, Gymnasium attendance was associated with lower math interest, even after controlling for the other variables in M4. This indicates a lack of any powerful assimilation effects with respect to math interest. Second, as indicated by the higher level of residual variance across the models, the percentage of variance explained in math interest was smaller than for math self-concept; this may be partly attributable to the lower internal consistency of the Interest scale.

We finally conducted an additional analysis to examine whether self-concept mediated the frame-of-reference effects on math interest. In Model 5 (M5), math self-concept was introduced as an additional predictor variable. As reported in Table 3, the regression coefficient for math self-concept was significant and of considerable size, whereas the direct predictive effects of school- and student-level achievement on interest vanished. This pattern of results is in line with the idea that the BFLPE affects important educational outcomes through its impact on academic self-concept.

Discussion

In Study 1, we examined a highly visible, explicit form of tracking that entails clear differences in student status. Students, parents, and teachers are all well aware (LeTendre et al., 2003) that the Gymnasium is the most valued school type in the tripartite German system, and the one that offers its students the best prospects. On average, Gymnasium students score more than two standard deviations higher on a standardized math test than their peers in the least prestigious school type, Hauptschule. Given the magnitude of these achievement differences, the visibility of the tracking structure, and its manifest impact on students' future careers, one might assume that assimilation effects in the form of labeling/stigmatization would occur in the lower track. It is an important result of Study 1 that no empirical support for such an assumption was found. On the contrary, lower track students had higher math self-concepts and math interest when individual math achievement was controlled. In line with our predictions, the beta weights of track level were substantially reduced when we controlled for school-average math achievement. Unexpectedly, however, there was still a positive predictive effect of lower track versus middle track when we controlled for school-average achievement and track level (but not teacher-assigned grades). Thus, contrary to the predictions of Figure 1c, we did not find support for a positive association between track level and math self-concept. We did, however, observe the expected regression coefficients of individual achievement and school-average achievement (positive and negative, respectively), and the same pattern of results was found for math interest. When we tested our mediator hypothesis (see Figure 1d), the predictive effect of individual and school-average math achievement were partially mediated by teacher-assigned grades; their predictive power was reduced, but still clearly apparent. Teacher-assigned grades mediated the predictive effects of both achievement and track level. In our analyses, higher math self-concept in lower track students seems to have been primarily a consequence of differential grading practices (full mediation). Furthermore, math self-concept mediated much of the effects of the other variables on math interest. Thus, taken together, we found strong support for our a priori predictions concerning contrast effects but no support at all for the predicted effects of assimilation.

Study 2

Although the tripartite system is the most common arrangement in German secondary education, there is considerable cross-state variation in the structures of secondary schooling (Baumert et al., 2003). Comprehensive secondary schools ("Gesamtschulen") play a minor role in Germany and are attended by just 10% of all students. However, because of their approach to tracking, these comprehensive schools are of particular interest in the present context. In the terms of our classification scheme, comprehensive schools typically implement (a) a highly structured form of *within-school* tracking, with students being assigned to two (sometimes three) streams—in mathematics, foreign languages, and German, at least. This form of tracking (b) is primarily based on prior achievement and (c) has important implications for students who aim to move on to a preuniversity course because these students are expected to have been in the upper streams in the core subjects. Taken together, this form of within-school ability grouping is thus another explicit, highly visible form of tracking that may be expected to entail labeling effects with respect to domain-specific achievement. It bears some similarities to within-school ability grouping in the U.S. system, although achievement is more relevant to streaming in Germany than in the United States, where course selection is also based on students' interests.

The hypotheses tested were similar to those examined in Study 1. Given that this form of within-school tracking is highly visible, we expected to find empirical evidence for assimilation effects, with students in higher streams exhibiting higher math self-concept. At the same time, we again expected to find statistical support for contrast effects, primarily in the form of negative regression coefficients relating stream-average math achievement to math self-concept. Moreover, we hypothesized that teacher-assigned grades would mediate the effects of individual and school-average achievement on math self-concept. Finally, we expected a similar pattern of results for both math self-concept and math interest; the regression coefficient was expected to be somewhat larger for math self-concept, and math self-concept was expected to substantially mediate frame-of-reference effects on math interest.

Method

Sample

As in Study 1, data were drawn from the German extension (Baumert et al., 2001) to the year 2000 cycle of the PISA study (OECD, 2001); 3,243 ninth-grade students (47.7% girls) from 177 comprehensive schools qualified for inclusion in the present study by providing information on their track level.

Instruments

The same instruments (standardized math achievement test, teacher-assigned math grades, math self-concept, and math interest) were administered as in Study 1. Math self-concept (α = .88) and math interest (α = .78) exhibited good internal consistency. Track level was dichotomized (0 = lower track, 1 = higher track). Students reported their track level in the student questionnaire.

Statistical Analyses

We closely followed the approach implemented in Study 1. Again, we performed multilevel modeling using HLM 5 (Raudenbush et al., 2000), standardizing ($M = 0$, $SD = 1$) all continuous variables before performing the analyses. Dichotomous variables were retained in their original metric. All models reported are random-intercept models, estimated by restricted maximum likelihood. Models were specified with individual students at Level 1, tracks within school at Level 2, and schools as Level 3 variables. Individual math achievement and school grades were used as Level 1 variables; track level and aggregated stream achievement were used as Level 2 variables. "Stream achievement" is the average achievement of all students in the same track of a school (and was not restandardized). With aggregated stream achievement and track level being juxtaposed, the analytical strategy was very similar to that used in Study 1. However, in Study 2, it was also possible to take into account a third level, the school. We therefore specified three-level models, with school as the third level, but we did not include school-average achievement as an additional predictor. There are two reasons for this. First, stream-average and school-average achievement exhibited a substantial intercorrelation that may produce multicollinearity if both variables were introduced simultaneously. Second, from a theoretical point of view, stream-average achievement is more central to self-concept development, given that it provides the most immediate frame of reference. As a set of exploratory analyses indicated, however, when stream-average achievement was substituted by school-average achievement, the regression coefficients were similar. Missing data (not exceeding 8% for any single variable) were again dealt with using multiple imputation.

Results

Results are presented in much the same way as for Study 1. We first consider the descriptive statistics (see Table 4). Not unexpectedly, we found a high intraclass correlation for math achievement, and the achievement differences between lower and upper track students were considerable, with a mean difference of about one standard deviation. Similar to the procedure in Study 1, we next included track level as a predictor variable. Track level explained 47% of the variance at the between-stream/school level; this finding indicates that track level and stream-average achievement are related but by no means are identical, predictor variables, which justifies their simultaneous use in the HLM models, reported below.

What we had not anticipated was the marked difference in grading practices observed in the upper and lower tracks. Without controlling for student achievement, the mean difference in teacher-assigned math grades between the two tracks amounted to

Table 4
Descriptive Statistics for Study 2 Variables

Variable	ICC	Lower track *M*	Lower track *SD*	Upper track *M*	Upper track *SD*
Math achievement	.38	78.62	15.70	96.72	18.74
Math school grade	.12	3.61	1.11	4.22	1.02
Math self-concept	.06	2.29	0.87	2.63	0.88
Math interest	.05	2.28	0.79	2.46	0.79

Note. Math school grades are reverse scored so that higher values represent higher achievement. ICC = Intraclass correlation for Level 2/track level (estimated in a baseline model using HLM 5).

more than half a standard deviation. This significant difference, $t(350) = 12.88$, $p < .001$ (statistical test performed with HLM), indicates that grades were implicitly or explicitly moderated by track level. Students were not simply graded on a curve. Math self-concept and math interest were higher in the upper track ($p < .001$), although there was little to no difference in the standard deviations, and between-track differences in prior achievement levels were not controlled at this point.

We next performed a series of HLM models. Again, we started with the prediction of teacher-assigned grades. Results of these analyses are documented in Table 5. In all models, individual math achievement positively predicted math grades. Unexpectedly, when controlling for individual math achievement, stream-average math achievement was nonsignificant in M1. This indicates that there was some kind of grading continuum across the two tracks, a finding that runs counter to the idea of grading on a curve. When we substituted track level for stream-average math achievement (M2), we found a positive regression coefficient for track level. In addition, when we included both track status and stream-average achievement (M3), results indicated that students with the same achievement were assigned lower grades in high-achieving groups but higher grades in the high-status track. Thus, reflecting the grading differences noted in the descriptive analyses, higher track status was associated with better grades. Taken together, these models do not fully support the idea that students are graded on a curve. Although—when track level was controlled—better grades were assigned in low-achieving streams (in line with our predictions and the results of Study 1), track status was also associated with teacher-assigned grades. Students in the lower track received lower grades than students in the upper track, even when controlling for individual and stream-average achievement.

We next performed a series of models with math self-concept as the dependent variable (see Table 6). In M1, math achievement was introduced at Levels 1 and 2. Consistent with the contrast hypothesis, stream-average achievement negatively predicted self-concept. In the next model, individual math achievement and track level, but not stream-average achievement, were used as predictors. Math track level did not significantly predict math self-concept in this model. Thus, unlike Study 1, there was no support for a contrast effect of track level in this model. In M3, track level and stream-average achievement were simultaneously introduced as Level 2 variables in addition to individual achievement. As expected, these two variables showed mutual suppression effects. The beta weight of track level became statistically significantly positive, indicating that, given the same individual and stream-average achievement, students in the higher tracks had a higher math self-concept. This is in line with predictions made on the basis of the assimilation hypothesis and suggests that membership of a high-status group has positive consequences for the self-concept. Moreover, consistent with our notion that the BFLPE is the net effect of counterbalancing contrast and assimilation effects, the regression weight for stream-average achievement in M3 ($b = -.29$) was larger than the regression weight reported for M1, in which track level was not taken into account (M1, $b = -.14$). In M4, teacher-assigned math grades were added. At the individual level, both math grades and math achievement significantly predicted math self-concept, but math grades had the larger regression weight. The introduction of school grades reduced the negative regression coefficient of stream-average achievement; in line with our mediation hypothesis, both the predictive effect of stream-average achievement ($-.28 \times .46 = -.13$, $p < .001$) and individual achievement ($.36 \times .46 = .17$, $p < .001$) were significantly mediated by school grades. Moreover, the positive regression coefficient of being a member of the upper track—which can be interpreted as evidence for an assimilation effect—was greatly reduced and no longer statistically significant. This indicates that the predictive power of track status was fully and significantly ($.47 \times .46 = .22$, $p < .001$) mediated by teacher-assigned grades. In support of the predictive power of school grades, the residual variance at Level 1 decreased markedly from M3 to M4 (see Table 6).

The results can be summarized as follows: First, both individual achievement and teacher-assigned grades (if included) positively predicted math self-concept in all models. When teacher-assigned grades were included, the regression coefficient of achievement was diminished but remained positive. Second, higher stream-average achievement was associated with lower math self-concept. However, the strength of this relationship was clearly affected by the inclusion of both grades and track level. Third, the predictive effect of track type varied across the models. The positive predictive effect of track type that would be expected given strong assimilation processes was only found when stream-average achievement was controlled and individual teacher-assigned grades were not included in the model. This suggests that the apparent support for assimilation processes provided by this study can be attributed to higher grades being assigned to students in the high track.

The pattern of results for math interest as the dependent variable (see Table 6) was again similar, although there are three important qualifications. First, the predictive effects of individual achievement (standardized math achievement test and math grades) on math interest were not as big as they were on math self-concept as the dependent variable. Second, when math school grades were taken into account (M4), stream-average achievement no longer contributed significantly to the prediction of math interest. Third,

Table 5
Regressing School Grades on Achievement and Track Level: Results From Multilevel Modeling (in Study 2)

	M1		M2		M3	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	−.01	.02	−.16***	.04	−.30***	.04
Track (0 = low, 1 = high)			.26***	.05	.47***	.06
M-Ach stream	−.01	.04			−.28***	.05
M-Ach individual	.36***	.02	.32***	.02	.36***	.02
Residual variance						
Level 3	.03		.03		.02	
Level 2	.05		.03		.04	
Level 1	.80		.80		.80	

Note. School achievement and school grades were standardized ($M = 0$, $SD = 1$) at the individual level before data analyses and aggregation. M-Ach stream = stream average math achievement in standardized math test; M-Ach individual = individual math achievement in standardized math test.
*** $p < .001$.

Table 6
Regressing Math Self-Concept and Math Interest on Achievement, Track, and School Grades (in Study 2): Results From Multilevel Modeling

	Self-concept								Interest									
	M1		M2		M3		M4		M1		M2		M3		M4		M5	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	.00	.02	−.01	.03	−.16***	.04	−.02	.03	.00	.02	−.01	.03	−.08*	.04	.01	.04	.03	.03
Track (0 = low, 1 = high)			.03	.04	.25***	.05	.04	.05			.01	.04	.13*	.06	−.02	.06	−.04	.05
M-Ach stream	−.14***	.04			−.29***	.05	−.17***	.04	−.08*	.04			−.15**	.05	−.07	.05	.05	.04
M-Ach individual	.42***	.02	.36***	.02	.42***	.02	.25***	.02	.25***	.02	.22***	.02	.25***	.02	.14***	.02	−.03	.02
M-Grade individual							.46***	.02							.30***	.02	−.01	.02
Math self-concept																	.69***	.02
Residual variance																		
Level 3	.03		.04		.02		.01		.01		.02		.01		.01		.00	
Level 2	.01		.01		.01		.00		.03		.03		.03		.02		.01	
Level 1	.84		.84		.84		.67		.91		.91		.91		.84		.53	

Note. Math self-concept, school achievement, and school grades were standardized ($M = 0$, $SD = 1$) at the individual level before data analyses and aggregation. M-Ach stream = stream-average math achievement in standardized math test; M-Ach individual = individual math achievement in standardized math test; M-Grade individual = teacher-assigned math grade.
* $p < .05$. ** $p < .01$. *** $p < .001$.

the residual variance in math interest explained was higher than for self-concept, which may be attributable in part to the lower internal consistency of the Interest scale relative to the Self-Concept scale. In an additional model (M5), we probed for the mediating effects of math self-concept. It emerged that math self-concept strongly predicted math interest, whereas the predictive effects of the other variables were no longer statistically significant. Overall, these results again support the assumption that frame-of-reference effects are not restricted to math self-concept but, mediated by math self-concept, also seem to apply to math interest.

Discussion

Within-school tracking in German comprehensive schools is a highly visible, explicit form of tracking that can have major implications for students' future educational careers. Given the large achievement differences between the tracks, it seems reasonable to expect corresponding differences in math self-concept. Such differences were indeed found on the mean level and in the form of a positive beta weight of track level in the multilevel models. It is tempting to interpret this pattern of results as indicative of assimilation effects. However, it was only when we controlled for both track and stream-average achievement at Level 2 that assimilation effects could be discerned (and not when only track level and individual achievement were controlled), and this "assimilation" effect disappeared when school grades were taken into account. This finding implies that differential grading practices, and not psychological assimilation effects per se, were the driving force behind the track differences observed in math self-concept. In comprehensive schools, students were not graded on a curve within their streams; rather, grades are moderated by some absolute measure of achievement. This pattern of results differs from that found in Study 1, in which students were graded on a curve, with a tendency for lower grades—relative to levels of achievement in a standardized test—to be assigned in higher tracks. The most probable reason for this difference is that Study 1 is a between-school study and Study 2 a within-school study.

General Discussion

In the second half of the last century, educational and political concerns caused explicit and rigid tracking to be eradicated almost completely from U.S. secondary schools. The driving forces for this change were concerns about overall achievement levels and the equality of educational opportunities for children of different races and family backgrounds, but tracking was also seen to impact negatively on the self-concepts, interests, and educational aspirations of students in lower tracks (Lucas, 1999). However, there is now growing awareness that being placed in lower achieving groups may actually enhance aspects of students' motivation, such as domain-specific self-concept. Because assimilation and contrast effects can best be examined in highly structured educational systems with tracks of different status, we focused on the German school system and tested the opposing predictions made of contrast and assimilation effects. Going beyond the work done in prior studies, we took school grades into account and postulated that the effects of tracking may be mediated by grading standards. Our two studies provide strong support for contrast effects and only limited support for assimilation effects. Moreover, our findings underline that grading practices play a key role in this arena.

Frames of Reference

Our findings substantiate the notion that students use others in their immediate environment (rather than all students of their age) as their reference group when constructing math self-concepts. As indicated by their rather high mean math self-concept scores, students in the lower track (Hauptschule) do not seem to excessively compare their achievement with that of students in the upper track (Gymnasium). Moreover, Gymnasium students do not seem to bask in the reflected glory of being selected to this track. Hence, the results of our study lend support to the view that Gymnasium students' math self-concepts tend to be negatively influenced by tracking, whereas Hauptschule students are likely to experience a positive impact on their self-views.

Our empirical results are much in line with predictions of Marsh's (1987) BFLPE model. Perhaps the most important finding is the absence of any consistent empirical support for assimilation effects in either study once school grades were controlled. This is a remarkable finding given the obvious achievement and status differences between the tracks considered in the present study. First, tracking was very explicit in our samples, and students were well aware of the track to which they were assigned. Second, track membership was based on prior achievement. Although some research findings indicate that students from families with a high socioeconomic and educational status are more likely to gain a place at Gymnasium than their peers who perform just as well at school but come from a less favorable family background (e.g., Baumert et al., 2001), prior achievement level remains the most important predictor of track membership (see also Table 2). Third, track membership has a direct impact on students' future educational and vocational outcomes. Despite these obvious differences in the prestige and average achievement levels of the three tracks, the mean level of math self-concept was similar across the three tracks in Study 1. This indicates that assimilation effects ("I'm in a high-status track, so I must be a good student") play a surprisingly minor role in the psychosocial development of adolescents, at least where domain-specific academic self-concepts are concerned.

Because students are tracked early in Germany (typically after Grade 4), the ninth-grade students in our sample had already been in their tracks for more than 4 years. With its single measurement point, our study is not suited to answer the question of how quickly reference group effects occur. However, a study by Schwarzer et al. (1982) found that academic self-concept was initially positively related to placement in the Gymnasium (upper track) rather than in the Hauptschule (lower track). Over their first 4 months of secondary education (between September and January), the self-concepts of Gymnasium and Hauptschule students converged considerably, but no further convergence was observed between January and June of the same school year. Hence, seen from a developmental perspective, students seem to switch their frames of reference fairly quickly when placed into a new learning environment.

The Role of Teacher-Assigned Grades

Our studies address an aspect that has rarely been examined in research on the BFLPE: the role of teacher-assigned grades. Consistent with Marsh (1987), we found that including grades in Study 1 substantially reduced the residual (unexplained) variance in math self-concept and somewhat reduced the predictive power of school-average achievement. More important, including grades in Study 2 qualified the empirical support for an assimilation effect associated with track level in within-school achievement grouping. Although track level was found to be associated with self-concept in within-school tracking, the main reason for this association appeared to be the teachers' differential grading practices. As in most other countries, grading in Germany tends to be based on norm-referenced criteria, with the immediate learning group (class or school) constituting the frame of reference. What frame of reference is used in comprehensive schools—the school or the track? It seems that teachers in comprehensive schools take both track information and students' relative position within their track into account, resulting in comparatively low grades in lower tracks. Indeed, taking into account individual and stream-average levels of achievement, the grades awarded to low-track students in Study 2 were lower than we predicted on the basis of group-specific grading practices (grading on a curve); this seems to have accounted for the apparent assimilation effect associated with track in Study 2. In fact, although the mean math achievement of Hauptschule students and students in the lower track of Gesamtschule students was similar, teacher-assigned school grades differed by about half a standard deviation, as did their math self-concept.

Overall, our results emphasize the importance of teacher-assigned grades as a predictor variable. Had grades not been considered, our conclusion would probably have been that—once individual achievement and stream-average achievement was controlled—low-track comprehensive students engage in (self-) labeling processes that are detrimental to their academic self-concepts. We would now argue that they are exposed to harsher grading practices than are similarly achieving students at, say, the low-track Hauptschule. On the basis of these findings, we suggest that research on the assimilation and contrast effects of achievement grouping should routinely examine the role of differential grading practices.

Given the strong statistical mediation effects of school grades, one may speculate about the causal status of grading practices in the development of the BFLPE. Might students use their classmates or schoolmates as the reference group precisely because teachers' grading practices are typically influenced by the average achievement of the students in a class? In other words, would we still find a BFLPE if no school grades were assigned? Results of both Studies 1 and 2 show that school-average (Study 1) and stream-average achievement (Study 2) had a negative effect on self-concept even after controlling for school grades assigned by teachers. Hence, the BFLPE seems to be more than simply an effect associated with grading on a curve. A recent study by Gerlach, Trautwein, and Lüdtke (2005) also sheds light on this issue. Gerlach and colleagues studied the effects of class-average achievement in physical education on the development of physical self-concept in more than 1,000 students in Grades 3 and 4. In their sample, physical education grades were not assigned until the end of Grade 3. The authors were nevertheless able to document a significant negative effect of class-average achievement on physical education self-concept at this grade level prior to the first assignment of school grades. This finding indicates that the BFLPE cannot be reduced to a simple grading-on-the-curve effect. In line with the model proposed in Figure 1d, however, the size of the BFLPE increased between Grades 3 and 4 in the Gerlach et al. study. This indicates that the introduction of school grades is likely to amplify reference group effects on domain-specific self-concept.

The Educational Relevance of the BFLPE: The Impact on Personal Interest

Some researchers (e.g., Plucker et al., 2004) have recently questioned the educational importance and implications of frame-of-references effects, maintaining that BFLPE studies have focused on academic self-concept as the outcome variable and that their conceptualization of self-concept has thus been too narrow (e.g., Plucker et al., 2004; Rindermann & Heller, 2005; see also Trautwein & Lüdtke, 2005). For this reason, the inclusion of individual interest in mathematics as a second dependent variable was an important aspect of our study. Interest is generally seen as a key factor in education that affects both effort and academic choices (e.g., Eccles & Wigfield, 2002; Krapp, 2000). The interest measure used in our research covered both the personal importance and the attainment value of engaging in math, as well as its intrinsic value (affective component).

The results, with regard to math interest attest to the power of the BFLPE, call claims of assimilation effects into question, and emphasize the importance of academic self-concept as an important outcome and mediator variable. First, the pattern of results for interest was very similar to that observed for self-concept. Interest was negatively predicted by school-average achievement and positively predicted by individual achievement. Hence, when controlling for individual achievement, students exhibited higher interest in lower track than in higher track schools. Second, track level was nonsignificantly or negatively related to interest, indicating that assimilation effects are not found in the context of interest. Third, self-concept was a potent predictor of interest and almost completely mediated the effects of achievement and track status. This highlights the role of self-concept as a key variable in educational settings (see also Marsh et al., 2005).

The pattern of results found for interest has important implications—not only for researchers specializing in the study of reference effects but also for those who are specifically interested in the role of individual interest in educational settings. To date, frame-of-reference effects have not attracted much attention from interest researchers. The results of our investigation suggest that this should change. Both studies indicate that an individual student's math interest is associated with his or her academic standing in the school or stream. Students with similar levels of individual math achievement are more likely to evidence high, stable individual interest in math if the achievement level of their reference group is low. It is of importance that this indicates that even students who perform well above average on standardized achievement tests are likely to report low interest in math if most other students in their school or stream perform even better.

What is the mechanism underlying this pattern of results? According to the expectancy-value theory (Eccles et al., 1983), domain-specific academic self-concepts affect the development of attainment value and feeling-related attitudes toward a school subject. Applying a longitudinal reciprocal effects model to a large sample of students in Grades 7 and 8, Marsh et al. (2005) recently found empirical evidence for this hypothesis, with math self-concept predicting later math individual interest. Our analysis provides further support for this mechanism. When we specified a mediator model in which math self-concept mediated between achievement variables and math interest, math self-concept fully mediated the predictive effects of the achievement variables. It is important to bear in mind, however, that our study was not longitudinal in design.

Our results suggest that students are less likely to develop high personal interest in mathematics when placed in a high-achieving classroom or track. For teachers, this finding constitutes a substantial challenge, as their influence on tracking is typically restricted. However, teachers are not powerless when it comes to enhancing students' interest. Research has documented the association between certain teaching characteristics (e.g., cognitive challenge and activation; high-quality homework) and positive development of student interest (e.g., Kunter & Baumert, 2006; Trautwein, Lüdtke, Schnyder, & Niggli, 2006; see also Hidi & Renninger, 2006; Schraw & Lehman, 2001). Moreover, prior research has shown that a teaching style that emphasizes effort and task orientation rather than the absolute level of achievement predicts the positive development of academic self-concept (see Lüdtke et al., 2005). It seems likely that a similar relationship may also be found for personal interest.

It is also important to note that our investigation was restricted to *individual* or *personal interest*, which is defined as an enduring positive affinity for and predisposition to attend to certain objects and activities. We did not examine frame-of-reference effects on *situational interest*, which is defined as transitory, context-dependent enjoyment triggered by environmental factors (see Krapp, 2000; Schraw & Lehman, 2001). Should similar frame-of-reference effects be expected for situational interest? On the one hand, students high on personal interest in a certain domain are more likely to activate their interest in corresponding situations. Thus, similar effects on individual and situational interest may be expected. On the other hand, teachers may purposely select materials with a strong "catch component" (see Hidi & Renninger, 2006; Mitchell, 1993) for the lower achieving students. From this perspective, the extent to which teachers select materials that are tailored to the high- or low-achieving students in a class is likely to determine the strength and direction of frame-of-reference effect on situational interest.

Limitations of the Present Study and Future Research

Although our investigation was based on two very strong data sets, some potential limitations should be addressed in future studies. First, by including math interest as an outcome variable, our study extended prior research on the BFLPE, which is typically restricted to the outcome variable of academic self-concept. Arguably, stronger assimilation effects may be observed when considering an even broader range of outcome variables. Marsh (1991) found strong support for contrast effects—negative effects of school-average achievement that generalized across a wide range of outcomes. However, because this study used data from the

United States, possible assimilation effects deriving from explicit achievement tracking at the school or stream level could not easily be examined. Similar studies with various outcome variables, both domain-specific and general (e.g., overall well-being at school, self-worth, attendance, persistence, coursework selection, educational and occupational aspirations, and accomplishments following graduation from high school), should be conducted in countries with explicit tracking systems to shed light on this important issue.

Second, schools rather than school classes were the basic unit for sampling in PISA, 2000, and, in line with the study's international guidelines, class membership was not documented in the German extension to the PISA, 2000 data set. In this respect, the present studies are comparable with the bulk of research on contrast effects in high-achieving groups (e.g., Marsh, 1991; Marsh & Hau, 2003) in which school-aggregated achievement was typically used as the primary variable. From a theoretical point of view, however, one would probably argue that classes are the natural frame of reference for students (Skaalvik & Skaalvik, 2002). From a methodological point of view, it is preferable to include all meaningful levels of analysis (e.g., Opdenakker & Van Damme, 2000). Thus, our study would have been even stronger (but considerably more complex) if classes could have been used as an additional unit in the sampling process.

Third, our study was nonexperimental and cross-sectional in nature, meaning that causal interpretations require caution. For instance, track differences in achievement should not be interpreted as a result of track effects because prior levels of achievement were the basis for selection into tracks. Moreover, in correlational studies, there is always the possibility that untested variables affected the pattern of results. Teacher-assigned grades are a prime example in the present context—although these grades evidenced a close association with the dependent variables, they have rarely been incorporated in research on the BFLPE. Likewise, we may well have overlooked further important predictor and mediator variables in the present investigation.

Tracking in schools is a worldwide phenomenon, but the nature of tracking and, in all likelihood, its impact on students differ markedly. More studies examining different forms of tracking around the world are needed before researchers can fully understand what characteristics of tracking affect what students, and in what ways. At the beginning of this study, we proposed a classification system to describe forms of tracking in terms of three central features that contribute to a track's status. We suggest that these features are used to differentiate between and compare different forms of tracking. Given the strong support found for contrast effects in the present article, as well as the lack of support for assimilation effects, future research should specifically aim to examine educational structures that are likely to shore up assimilation effects.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., et al. (Eds.). (2002). *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich* [PISA 2000: A comparison of the German states]. Opladen, Germany: Leske and Budrich.

Baumert, J., Klieme, E., Neubrand, M., Schiefele, U., Schneider, W., Stanat, P., et al. (Eds.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* [PISA 2000: Literacy of students in international comparison]. Opladen, Germany: Leske and Budrich.

Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., et al. (1997). *TIMSS: Mathematisch-Naturwissenschaftlicher Unterricht im internationalen Vergleich.* [TIMSS: Mathematics and science instruction in an international comparison]. Opladen, Germany: Leske and Budrich.

Baumert, J., Schnabel, K., & Lehrke, M. (1998). Learning math in school: Does interest really matter? In L. Hoffmann, A. Krapp, K. A. Renninger, & J. Baumert (Eds.), *Interest and learning* (pp. 327–336). Kiel, Germany: IPN.

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten—Institutionelle Bedingungen des Lehrens und Lernens [School environments—Institutional conditions for learning and instruction]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Eds.), *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* [PISA 2000] (pp. 259–330). Opladen, Germany: Leske and Budrich.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's third international mathematics and science study*. Chestnut Hill, MA: Boston College.

Berends, M. (1994). Educational stratification and students' social bonding to school. *British Journal of Sociology of Education, 16,* 327–351.

Bracken, B. A. (Ed.). (1996). *Handbook of self-concept*. New York: Wiley.

Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.

Csikszentmihalyi, M., & Schiefele, U. (1993). Die Qualität des Erlebens und der Prozess des Lernens [The quality of experience and the process of learning]. *Zeitschrift für Pädagogik, 39,* 207–221.

Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Mahwah, NJ: Erlbaum.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53,* 109–132.

Eccles (Parsons), J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.), *Perspectives on achievement and achievement motivation* (pp. 75–146). San Francisco: Freeman.

Felson, R. B. (1984). The effect of self-appraisals on ability of academic performance. *Journal of Personality and Social Psychology, 47,* 944–952.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

Gerlach, E., Trautwein, U., & Lüdtke, O. (2005). *Physical self-concept and physical activity: A longitudinal study on frame of reference effects*. Manuscript submitted for publication.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52,* 123–142.

Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver, CO: University of Denver.

Hidi, S., & Ainley, M. (2002). Interest and adolescence. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (pp. 247–275). Greenwich, CT: Information Age.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70,* 151–179.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41,* 111–127.

Hodge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59,* 297–313.

Ingenkamp, K. (1971). *Die Fragwürdigkeit der Zensurengebung* [The dubiousness of school grades]. Weinheim, Germany: Beltz.

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

Klieme, E., Neubrand, M., & Lüdtke, O. (2001). Mathematische grundbildung: Testkonzeption und ergebnisse [Mathematical literacy: Test construction and results]. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Eds.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 139–190). Opladen: Leske and Budrich.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education, 32,* 448–470.

Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 109–128). London: Elsevier.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277.

Kunter, M., & Baumert, J. (2006). Linking TIMSS to research on learning and instruction: A re-analysis of the German TIMSS and TIMSS video data. In S. J. Howie & T. Plomp (Eds.), *Learning mathematics and science: Lessons learned from TIMSS.* London: Routledge.

LeTendre, G., Hofer, B., & Shimizu, H. (2003). What is tracking? Cultural expectations in the U. S., Germany, and Japan. *American Educational Research Journal, 40,* 43–89.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

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

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.

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28,* 165–181.

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

Marsh, H. W. (1990). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology, 82,* 107–116.

Marsh, H. W. (1991). The failure of high-ability high schools to deliver academic benefits: The importance of academic self-concept and educational aspirations. *American Educational Research Journal, 28,* 445–480.

Marsh, H. W. (1993). Academic self-concept: Theory, measurement and research. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 4, pp. 59–98). Hillsdale, NJ: Erlbaum.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 131–198). San Diego, CA: Academic Press.

Marsh, H. W., & Craven, R. (2002). The pivotal role of frames of reference in academic self-concept formation: The big fish little pond effect. In F. Pajares & T. Urdan (Eds.), *Adolescence and education* (Vol. II, pp. 83–123). Greenwich, CT: Information Age.

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,* 364–376.

Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big fish little pond effect on academic self-concept. *American Educational Research Journal, 38,* 321–350.

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.

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21,* 153–174.

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

Marsh, H. W., & Rowe, K. J. (1996). The negative effects of school-average ability on academic self-concept: An application of multilevel modeling. *Australian Journal of Education, 40,* 65–87.

Marsh, H. W., Tracey, D. K., & Craven, R. G. (in press). Multidimensional self-concept structure for preadolescents with mild intellectual disabilities: A hybrid multigroup-mimic approach to factorial invariance and latent mean differences. *Educational and Psychological Measurement*.

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

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85,* 424–436.

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

Opdenakker, M.-C., & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement, 11,* 103–130.

Organisation for Economic Cooperation and Development. (2001). *Knowledge and skills for life. First results from the OECD programme for international student assessment (PISA) 2000.* Paris: Author.

Plucker, J. A., Robinson, N. M., Greenspon, T. S., Feldhusen, J. F., McCoach, D. B., & Subotnik, R. F. (2004). It's not how the pond makes you feel, but rather how high you can jump. *American Psychologist, 59,* 268–269.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. T. (2000). *HLM 5: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

Rheinberg, F., & Enstrup, B. (1977). Self-concept of intelligence with pupils in special and normal schools: A group effect. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 9,* 171–180.

Rindermann, H., & Heller, K. A. (2005). The benefit of gifted classes and talent schools for developing students' competences and enhancing academic self-concept. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 19,* 133–136.

Rost, J. (1996). *Testtheorie und Testkonstruktion* [Testing theory and the construction of tests]. Bern, Switzerland: Huber.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Wiley.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13,* 23–52.

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

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research, 46,* 407–441.

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

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

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 181–227). San Diego, CA: Academic Press.

Tracey, D. K., Marsh, H. W., & Craven, R. G. (2003). Self-concepts of preadolescent students with mild intellectual disabilities: Issues of measurement and educational placement. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (Vol. 1, pp. 203–230). Greenwich, CT: Information Age.

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. (2005). The big fish little pond effect: Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 19,* 137–140.

Trautwein, U., Lüdtke, O., Kastens, C., & Köller, O. (2006). Effort on homework in grades 5 through 9: Development, motivational antecedents, and the association with effort on classwork. *Child Development, 77,* 1094–1111.

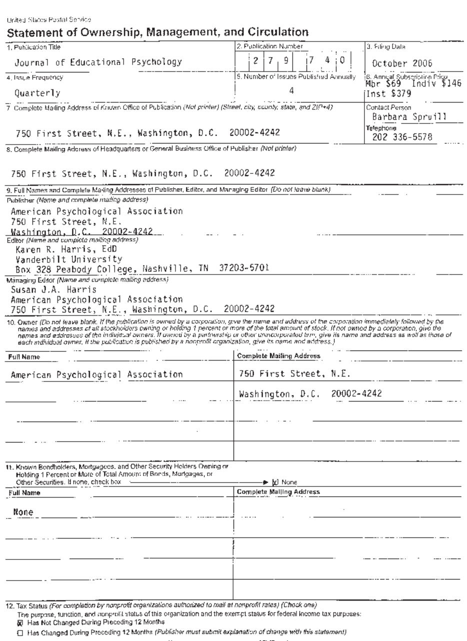
Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90,* 334–349.

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98,* 438–456.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 173–195). San Diego, CA: Academic Press.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A., Freedman-Doan, K., & Blumenfeld, P. C. (1997). Changes in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology, 89,* 451–469.

Received April 22, 2005
Revision received June 15, 2006
Accepted July 6, 2006 ■



United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title: Journal of Educational Psychology
2. Publication Number: 2 7 9 7 4 0
3. Filing Date: October 2006
4. Issue Frequency: Quarterly
5. Number of Issues Published Annually: 4
6. Annual Subscription Price: Mbr $69 Indiv $146 Inst $379
7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*: 750 First Street, N.E., Washington, D.C. 20002-4242

Contact Person: Barbara Spruill
Telephone: 202 336-5578

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*: 750 First Street, N.E., Washington, D.C. 20002-4242
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*

Publisher *(Name and complete mailing address)*
American Psychological Association
750 First Street, N.E.
Washington, D.C. 20002-4242

Editor *(Name and complete mailing address)*
Karen R. Harris, EdD
Vanderbilt University
Box 328 Peabody College, Nashville, TN 37203-5701

Managing Editor *(Name and complete mailing address)*
Susan J.A. Harris
American Psychological Association
750 First Street, N.E., Washington, D.C. 20002-4242

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

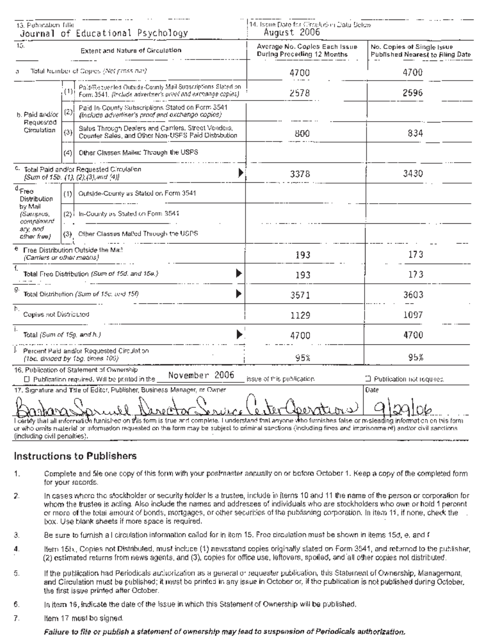
Full Name	Complete Mailing Address
American Psychological Association	750 First Street, N.E.
	Washington, D.C. 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address
None	

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form **3526**, October 1999 *(See Instructions on Reverse)*



13. Publication Title: Journal of Educational Psychology
14. Issue Date for Circulation Data Below: August 2006

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*		4700	4700
b. Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. *(Include advertiser's proof and exchange copies)*	2578	2596
	(2) Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*		
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	800	834
	(4) Other Classes Mailed Through the USPS		
c. Total Paid and/or Requested Circulation *(Sum of 15b. (1), (2),(3),and (4))* ▶		3378	3430
d. Free Distribution by Mail *(Samples, compliment ary, and other free)*	(1) Outside-County as Stated on Form 3541		
	(2) In-County as Stated on Form 3541		
	(3) Other Classes Mailed Through the USPS		
e. Free Distribution Outside the Mail *(Carriers or other means)*		193	173
f. Total Free Distribution *(Sum of 15d. and 15e.)* ▶		193	173
g. Total Distribution *(Sum of 15c. and 15f)* ▶		3571	3603
h. Copies not Distributed		1129	1097
i. Total *(Sum of 15g. and h.)* ▶		4700	4700
j. Percent Paid and/or Requested Circulation *(15c. divided by 15g. times 100)*		95%	95%

16. Publication of Statement of Ownership
☐ Publication required. Will be printed in the November 2006 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Barbara Spruill Director Service Center Operations
Date: 9/29/06

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form **3526**, October 1999 *(Reverse)*