Happy Fish in Little Ponds:

Testing a Reference Group Model of Achievement and Emotion

Reinhard Pekrun

University of Munich and Australian Catholic University

Kou Murayama

University of Reading and Kochi University of Technology

Herbert W. Marsh

Australian Catholic University and University of Oxford

Thomas Goetz

University of Konstanz and Thurgau University of Teacher Education

Anne C. Frenzel

University of Munich

September 2018

Journal of Personality and Social Psychology, in press

Author Note

Reinhard Pekrun, Department of Psychology, University of Munich, Munich, Germany, and Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia; Kou Murayma, School of Psychology and Clinical Language Sciences, University of Reading, United Kingdom, and Research Institute, Kochi University of Technology, Kami City, Kochi, Japan; Herbert W. Marsh, Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia, and Department of Education, University of Oxford, Oxford, United Kingdom; Thomas Goetz, Department of Empirical Educational Research, University of Konstanz, Konstanz, Germany, and Department of Empirical Educational Research, Thurgau University of Teacher Education, Kreuzlingen, Switzerland; Anne C. Frenzel, Department of Psychology, University of Munich, Munich, Germany.

This research was supported by four grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) awarded to Reinhard Pekrun (PE 320/11-1, PE 320/11-2, PE 320/11-3, PE 320/11-4), and a grant from the Japan Society for the Promotion of Science awarded to Kou Murayama (Kakenhi 16H06406).

Correspondence concerning this article should be addressed to Reinhard Pekrun, Department of Psychology, University of Munich, 80802 Munich, Germany. E-mail: pekrun@lmu.de

Abstract

A theoretical model linking achievement and emotions is proposed. The model posits that individual achievement promotes positive achievement emotions and reduces negative achievement emotions. In contrast, group-level achievement is thought to *reduce* individuals' positive emotions and *increase* their negative emotions. The model was tested using one cross-sectional and two longitudinal datasets on 5^{th} to 10^{th} grade students' achievement emotions in mathematics (Studies 1-3: *N*s = 1,610, 1,759, and 4,353, respectively). Multi-level latent structural equation modeling confirmed that individual achievement had positive predictive effects on positive emotions (enjoyment, pride) and negative predictive effects on negative emotions (anger, anxiety, shame, and hopelessness), controlling for prior achievement, autoregressive effects, reciprocal effects, gender, and socio-economic status. Class-level achievement had negative compositional effects on the positive emotions and positive compositional effects on the negative emotions. Additional analyses suggested that self-concept of ability is a possible mediator of these effects. Furthermore, there were positive compositional effects of class-level achievement on individual achievement in Study 2 but not in Study 3, indicating that negative compositional effects on emotion are not reliably counteracted by positive effects on performance. The results were robust across studies, age groups, synchronous versus longitudinal analysis, and latent-manifest versus doubly-latent modeling. These findings imply that individual success drives emotional well-being, whereas placing individuals in high-achieving groups can undermine well-being. Thus, the findings challenge policy and practice decisions on achievement-contingent allocation of individuals to groups.

Keywords: achievement emotion, compositional effect, contextual effect, frame-of-reference effect, control-value theory

Achievement emotions are an integral part of human personality, and they are ubiquitous across settings. Traditionally, these emotions have received little attention by personality researchers, except for studies on test anxiety (Zeidner, 2014) and the link between causal attributions and achievement emotions (Weiner, 1985, in press). More recently, however, there has been growing recognition that achievement emotions are central to human well-being and achievement strivings. These emotions are no longer regarded as epiphenomena that may occur in achievement settings but lack any instrumental relevance. Rather, achievement emotions are recognized as being of critical importance for psychological health, personality development, and the productivity of individuals, organizations, and cultures (Ashkanasy & Humphrey, 2011; Boehm & Lyubomirsky, 2008; Efklides & Volet, 2005; Elfenbein, 2007; Linnenbrink, 2006; Pekrun & Linnenbrink-Garcia, 2014; Warr, 2007). In fact, across disciplines researchers investigating achievement have noted that there is an affective turn in their fields (see, e.g., Ashkanasy, Härtel, & Daus, 2002; Pekrun, 2017).

Given the importance of achievement emotions, more knowledge about their antecedents is needed. Prior research has considered individual success and failure as possible antecedents but has neglected the achievement of others. In this research, we argue that the level of achievement in one's reference group also plays a major role, over and above the influence of individual achievement. As such, we consider both individual and group-level achievement. We examine the role of these variables as antecedents to six major positive and negative achievement emotions, including enjoyment, pride, anger, anxiety, shame, and hopelessness.

The analysis is based on a theoretical model that disentangles the effects of individual and group-level achievement. In this model, we consider the compositional effects of group-level achievement on emotions. The term compositional effect is generally used to denote effects of the group-level aggregate of individual variables, such as achievement, gender, or socio-economic status, over and above the contribution of these variables at the individual level (see Harker & Tymms, 2004; Raudenbush & Bryk, 2002).[1] To test the model in ecologically valid ways, we used data from three large-scale studies of students' achievement and emotions in mathematics, a field that is known to generate intense emotions (Chang & Beilock, 2016; Pekrun et al., 2017). Although the model was tested using data from an educational context, it provides a general framework for understanding compositional effects of group-level achievement on emotion, whatever the specific context.

Our model explains how both individual success and failure and the achievement of others contribute to the arousal of achievement emotions. Specifically, we argue that the level of achievement in one's reference group determines opportunities to succeed relative to others, and that others' achievement prompts social comparison processes that influence achievement emotions and the self-concepts underlying these emotions (Huguet et al., 2009; Pekrun, 2006; Pekrun & Perry, 2014). As such, the model proposes that groups serve as frames of reference for the arousal of achievement emotions, and that the composition of groups prompts frame-of-reference effects on these emotions.

By considering groups as frames of reference shaping emotions, the model builds on social comparison theory and the literature on frame-of-reference effects (Davis, 1966; Festinger, 1954; Stapel & Blanton, 2007; Zell & Alicke, 2009). This literature has shown that beliefs about oneself can be strongly influenced by social comparison, but has neglected the impact of social comparison on emotion. By targeting effects of group composition, the model also is related to existing research on compositional effects. Compositional effects are foundational to the impact of groups on individual differences and personality development (Raudenbush & Bryk, 2002).

Prior research in this field has focused on the effects of ability composition on cognitive performance (e.g., Baratta & McManus, 1992; De Fraine, van Damme, van Landeghem, Opdenakker, & Onghena, 2003; Dicke et al., 2018; Harker & Thymms, 2004; Thrupp, Lauder, & Robinson, 2002) but has also neglected possible effects on emotion.

The present research addresses this gap in personality and social psychological research by integrating perspectives from the achievement emotion, frame-of-reference, and compositional effects literatures. In the following sections we first summarize prior research. Next, we describe our reference group model and provide an overview of the present studies. We then present the findings of our three empirical studies.

Prior Research

Prior research has focused on the relationship between individual achievement and emotion but has neglected the possible influence of group-level achievement. Furthermore, most studies used cross-sectional designs and failed to examine the impact of achievement on emotions over time.

Individual Achievement and Emotions

Test anxiety research has found negative correlations between individual achievement and achievement-related anxiety (Zeidner, 2014). These correlations can be explained by the effects of achievement on the development of anxiety. Specifically, success and failure shape perceptions of competence and expectancies of failure underlying achievement anxiety (Lang & Lang, 2010; Pekrun, 1992). Alternatively, the relationship may be caused by effects of anxiety on task-irrelevant thinking that interferes with cognitive performance and contributes to low achievement (interference and attentional deficit models; see Chang & Beilock, 2016; Eysenck, 1997). In fact, the scarce longitudinal evidence available suggests that achievement and anxiety are linked by reciprocal effects over time. However, the influence of achievement on anxiety typically is stronger than reverse effects of anxiety on achievement (Meece, Wigfield, & Eccles, 1990; Pekrun, 1992; Schnabel, 1998; Steinmayr, Credel, McElvany, & Wirthwein, 2016).

For emotions other than anxiety, the evidence is limited. Positive correlations with achievement have been observed for students' enjoyment of learning (Larson, Hecker, & Norem, 1985; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Pekrun et al., 2017). Similarly, general positive affect has been found to correlate positively with students' achievement as well as employees' workplace performance (Linnenbrink, 2007; Boehm & Lyubomirsky, 2008). However, null findings have been reported as well (Linnenbrink, 2007; Pekrun et al., 2009). For anger, shame, and general negative affect, negative correlations have been found (Boekaerts, 1993; Pekrun et al., 2009, 2011, 2017), although not in all instances (Linnenbrink, 2007; Turner & Schallert, 2001). For hopelessness, the findings suggest that relations with achievement are uniformly negative (Pekrun et al., 2011, 2017). However, longitudinal studies of achievement and emotions other than anxiety are largely lacking (for an exception, see Pekrun et al., 2017).

In summary, to date research on the relationship between individual achievement and achievement emotions is limited. Most studies addressed achievement anxiety and have demonstrated that anxiety is negatively related to achievement. A smaller number of studies have focused on other emotions, and found that enjoyment in achievement contexts tends to be positively associated with achievement, whereas anger, shame, and hopelessness tend to be negatively associated. More research is clearly needed on the relation between individual achievement and emotions (other than anxiety).

Group-Level Achievement and Emotions

Evidence on the possible relation between group-level achievement and emotions is lacking, except for two studies of test anxiety in gifted students. These studies included gifted students who were either members of regular classes or of special classes for gifted students within Israel's national program for gifted students. In the first investigation, students from gifted (i.e., high average achievement) classes reported higher anxiety than students from regular, mixed-achievement classes (Zeidner & Schleyer, 1998). In the second study, gifted students enrolled in gifted classes had higher anxiety scores than gifted students in mixed-achievement classes (Zeidner & Schleyer, 1999). The dataset of this study was reanalyzed by Goetz, Preckel, Zeidner, and Schleyer (2008) to directly examine the relations of individual and class-level achievement with gifted students' test anxiety. The results of hierarchical linear modeling showed that individual achievement was negatively related to students' test anxiety, whereas class-level achievement was positively related to their anxiety.

These findings suggest that the group-average level of achievement can be linked to group members' anxiety. However, this evidence is limited by the cross-sectional nature of the analysis, use of small convenience samples, and inclusion of one negative emotion only. Conclusions on the temporal ordering of variables and generalizability to other emotions cannot be drawn from these findings.

Reference Group Model of Achievement and Emotion

It seems straightforward to assume that individual success strengthens positive achievement emotions whereas failure prompts negative emotions (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Zeidner, 1998). Similar assumptions are widely held for group-level achievement. Being a member of a high-achieving group is thought to be beneficial for well-being and performance (see, e.g., Stäbler, Dumont, Becker, & Baumert, 2017). For example, in education parents seek to have their children accepted to the "best" schools where they are surrounded by able peers who can serve as role models for learning and performance. Similarly, at work people strive to join successful companies and institutions, and in sports athletes dream of joining top-scoring teams. Being a part of high-achieving groups is perceived to convey high social status, thus presumably promoting pride and positive emotions, and reducing negative emotions such as shame.

In line with the test anxiety studies cited earlier, our model challenges this view by considering the emotional costs of being among high achievers. The model is based on Pekrun's (2006; Pekrun & Perry, 2014) control-value theory of achievement emotions. In this theory, achievement emotions are defined as emotions related to achievement activities or achievement outcomes. Examples are enjoyment, pride, anger, anxiety, shame, or hopelessness related to studying, working, or athletic performance and to the success and failure outcomes of these activities. According to the theory, these emotions depend on the perceived controllability of achievement activities and outcomes that are subjectively important. Achievement emotions are thought to be aroused if an individual feels in control over, or out of control of, these activities and outcomes. For example, anxiety should be triggered if an employee feels out of control over performance on an important project. Because perceptions of control are shaped by past performance, the theory can be used to derive hypotheses about the effects of individual and group-level achievement on emotions.

The theory considers different types of control appraisals including action- and outcome-related expectations and attributions. As all of these appraisals hinge on perceptions of one's ability, self-concept of ability is thought to be of primary relevance among different constructs of control. Self-concepts of ability are defined as cognitive representations of one's abilities that are stored in long-term memory. Typically, these self-concepts are organized in domain-specific ways (e.g., math vs. verbal self-concepts; Möller & Marsh, 2013). High self-concept of ability is expected to promote positive achievement emotions and to reduce negative emotions (for a summary of supporting evidence, see Pekrun & Perry, 2014). By implication, factors influencing self-concept of ability, which include individual as well as group-level achievement, are expected to also influence these emotions.

By targeting the effects of both individual and group-level achievement, the model is similar to previous models of compositional effects (see Dicke et al., in press; Harker & Tymms, 2004; Thrupp, Lauder, & Robinson, 2002). The model has a multi-level structure that is consistent with the general logic of models considering compositional effects. It differs from other models by explaining emotions as outcomes. More specifically, the model proposes the following.

Effects of Individual Achievement

The evidence cited earlier suggests that individual success and failure are prime antecedents of achievement emotions. Success prompts positive achievement emotions, including enjoyment and pride; failure triggers negative achievement emotions, including anger, anxiety, shame, and hopelessness. From the control-value theory, it follows that these effects can in part be explained by the influence of individual achievement on self-concept of ability. Individual achievement is known to have positive effects on self-concept of ability (see Arens, Marsh, Pekrun, Lichtenfeld, & vom Hofe, 2017). Self-concept of ability, in turn, is thought to have positive effects on positive achievement emotions (enjoyment, pride) and negative effects on negative achievement emotions (anger, anxiety, shame, hopelessness).

The effects of achievement on emotions and underlying self-concepts are thought to be universal, as individuals naturally judge their ability in relation to their performance on achievement tasks. The effects may be especially pronounced if the available information about achievement is salient, important, and consistent. These conditions are met if individuals receive distinct information about their performance, if performance has consequences that are important for one's current life or future prospects, and if performance information from different sources converges, as is typically the case with grades or test scores in educational institutions, ratings by supervisors at the workplace, or the results of competitions in professional athletes.

Effects of Group-Level Achievement

The level of achievement within reference groups determines one's chances to succeed relative to others. All else being equal, being in a high-achieving group makes it more difficult to be successful as compared with others and increases the likelihood of failure. Conversely, being in a low-achieving group makes it easier to succeed and reduces the likelihood of failure relative to others. Because success and failure drive achievement emotions, high group-level achievement is thought to reduce positive achievement emotions (enjoyment, pride) and to increase negative achievement emotions (anger, anxiety, shame, hopelessness).

Self-concepts of ability are thought to play a role in these effects as well. High group-level achievement is known to have negative effects on individuals' self-concept of ability, all else being equal (frog-pond effect or big-fish-little-pond effect; Davis, 1966; Marsh, 1987; Marsh et al., 2008, 2015; Marsh, Parker, & Pekrun, 2018). When group-average achievement is high and individual opportunities to succeed relative to others are reduced, social comparison with others results in less favorable self-perceptions. As such, being in a high-achieving group undermines one's self-concept of ability, whereas being in a low-achieving group promotes a positive self-concept, all else being equal. The big-fish-little-pond effect is also expected to be, and has been shown to be, universal as individuals rank-order themselves based on perceptions of their ability relative to others, in line with social comparison theory (Marsh et al., 2015). The effect is remarkably robust across genders, age groups, student characteristics, school settings, and cultures (Marsh, Hau, & Craven, 2004; Marsh et al., 2008, 2015; Seaton, Marsh, & Craven, 2010).

The effects of group-level achievement on emotions may also be especially pronounced if the available information about the achievement of others is salient, relevant, and consistent. As such, it is expected, and has been shown empirically, that salient “local” groups generate stronger frog-pond effects than more distant reference groups (“local dominance effect;” Alicke, Zell, & Bloom, 2010; Zell & Alicke, 2009). For example, the level of achievement in a students’ class is likely to exert stronger effects than state-wide achievement in the students’ country, and the level of achievement in one’s local sports team may be more important for one’s emotions than the national team’s performance. As such, to properly examine the possibility of compositional effects on emotion, it is important to adequately select the units of observation at the group level.

In sum, we hypothesize that individual achievement is positively linked to emotional well-being (i.e., associated with more positive emotions and less negative emotions). In contrast, group-level achievement is expected to be negatively linked to well-being (i.e., associated with less positive emotions and more negative emotions), due to the negative effects of group-level achievement on opportunities to succeed and to develop favorable self-perceptions of ability. As such, our model implies that being in a low-achieving group relates to better emotional well-being compared with being in a group of high achievers. We suggest calling this relation the "happy-fish-little-pond effect" (it's better to be a happy fish in a little pond than an unhappy fish in a big pond of high achievers), using a term that expresses the consistency of the relation with the big-fish-little-pond effect on self-concept, while also denoting that the target of the effect is emotions rather than self-appraisals.

The Present Research

Overview of Studies

Datasets from one cross-sectional and two longitudinal studies were used to test the model. The three studies were located in an educational context and examined students' achievement and emotions in mathematics, including math-related enjoyment, pride, anger, anxiety, hopelessness, and shame. Subject-specific emotions were assessed rather than generalized emotional traits, because achievement emotions have been shown to be organized in domain-specific ways (Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007). Standardized mathematics tests and self-report were employed to measure achievement and emotions, respectively. We used test scores rather than grades as an indicator of achievement because grades are derived from grading-on-the curve within classes, implying that they cannot be used to assess differences in group-average achievement between classes. Studies 2 and 3 additionally included students' math self-concept.

The studies included representative student samples from the German secondary public school system which uses achievement-based tracking. Progression through the system is made contingent on achievement, whereby low-achieving students are retained or transferred to a lower school track, indicating that achievement is salient and critically important for students' educational career. As such, we expected to discover the proposed relations of achievement to emotions in these samples.

Study 1 used a cross-sectional design for an initial exploration of the proposed achievement-emotion links. Studies 2 and 3 used longitudinal designs. One major obstacle to examining effects of compositional variables, such as group-level achievement, over time is change in the composition of groups. To minimize this problem, two-wave designs were used that provided stability of classroom composition from wave to wave (grade 5 to 6 in Study 2, and grade 9 to 10 in Study 3). Achievement and emotion were assessed at both waves, making it possible to control for autoregressive and reciprocal effects.

Multi-level modeling with latent variables was employed to analyze the data. To consider the multicollinearity between emotions (see Table 1), separate models were estimated for the different emotions. As students are nested in classes and performance assessment is class-based in German public schools, classes were considered as the relevant unit of analysis at the group level (also see Marsh, Köller, & Baumert, 2001; Marsh, Kuyper, Morin, Parker, & Seaton, 2014). Classes represent students' immediate reference groups; as such, this choice is consistent with considering the local dominance effect as discussed earlier (Alicke et al., 2010; Zell & Alicke, 2009). Students' gender and family socio-economic status (SES) were controlled in the analysis (for the influence of gender on math emotions, see Beilock, Gunderson, Ramirez, & Levine, 2010; Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013).

Study 1

Study 1 explored the links between achievement and four major achievement emotions in mathematics, including enjoyment, anger, anxiety, and hopelessness. The study used data from a cross-sectional investigation which is part of the multi-study *Project for the Analysis of Learning and Achievement in Mathematics*, PALMA (see Marsh et al., 2017, 2018;

Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Murayama, Pekrun, Suzuki, Marsh, & Lichtenfeld, 2016; Pekrun et al., 2017).[2]

Method

The study is a secondary analysis of an existing dataset (see Pekrun et al., 2007). The studies of the PALMA project received Institutional Review Board approval from the Bavarian State Ministry for Education, Science, and the Arts (reference III/5-S4200/4-6/68 908).

Sample and procedure. The sample consisted of 1,610 students (grades 5-10; 47.5% female) from 69 classrooms in the state of Bavaria, Germany, and comprised a wide range of socioeconomic backgrounds, including both rural and urban areas and students from all three schools tracks of the public school system in this state (low-achievement track: Hauptschule; medium-achievement track: Realschule; high-achievement track: Gymnasium). The mean number of participating students per class was 23.3 (minimum = 11, maximum = 33, *SD* = 4.86). The sample included 292 students (50.7% female) from grade 5 (mean age = 11.24 years, *SD* = .48), 253 students (43.9% female) from grade 6 (mean age = 12.41 years, *SD* = 0.77), 271 students (45.4% female) from grade 7 (mean age = 13.31 years, *SD* = 0.55), 311 students (46.3% female) from grade 8 (mean age = 14.37 years, *SD* = 0.63), 292 students (49.7% female) from grade 9 (mean age = 15.44 years, *SD* = 0.73), and 191 students (45.5% female) from grade 10 (mean age = 16.45 years, *SD* = 0.77). The assessment took place towards the end of the school year. All instruments were administered in the students' classrooms by trained external test administrators. Student participation was based on written parental consent, and students' responses were kept strictly confidential.

Variables and measures. The study variables included students' achievement, enjoyment, anger, anxiety, and hopelessness in mathematics as well as their gender and grade level. Pride, shame, self-concept, and SES were not assessed in this study.

Achievement. The PALMA Mathematical Achievement Test (Murayama et al., 2013; Pekrun et al., 2007) was used to measure students' achievement. This is a standardized test using item response theory (IRT) scaling to assess competencies in arithmetic, algebra, and geometry across a wide range of ability.

Emotions. Short scales from the Achievement Emotions Questionnaire-Mathematics (AEQ-M; Frenzel, Thrash, Pekrun, & Goetz, 2007; Pekrun et al., 2011) were used to assess students' math emotions. The instructions asked respondents to describe how they typically felt when attending class, doing homework, and taking tests in math. The scales assessed enjoyment (8 items; e.g., "I enjoy my math class"), anger (6 items; e.g., "My mathematics homework makes me angry"), anxiety (9 items; e.g., "I worry if the material is much too difficult for me"), and hopelessness (6 items; e.g., "During the math test, I feel hopeless"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale, and the scores were averaged to form the emotion indexes (α range = .87 - .90; see supplemental materials, Table 1).

Demographic variables. We controlled for gender and grade level in the analysis. Gender was coded 1 = female, 2 = male. Grade level was coded as it is defined in schools (i.e., grade 5 = 5, grade 6 = 6, etc.).

Multi-level modeling. Studies on compositional effects have moved from employing single-level models and ordinary least-square regression analysis to using multi-level modeling (Gonzáles-Romá & Hernández, 2017; Marsh et al., 2009). Single-level analysis renders conflated estimates of individual-level and group-level parameters (Hox, 2002; Raudenbush & Bryk, 2002), making it difficult to interpret the results. As such, we used multi-level modeling (Mplus, Version 8, Muthén & Muthén, 2017; students at Level 1, classes at Level 2).

In the estimated models, we corrected for measurement error. Not correcting for measurement error can lead to over- or underestimation of compositional effects as well as false compositional effects (phantom effects; Dicke et al., 2018; Harker & Tymms, 2004; Hauser, 1974; Marsh et al., 2009; Televantou et al., 2015). In supplementary analyses, we additionally corrected for sampling error. Sampling error refers to error in estimating Level-2 parameters caused by insufficient sampling of individuals from Level-2 units (e.g., if a project team comprises twenty members but only five have been assessed; see Lüdtke et al., 2008; Marsh et al., 2009; Morin, Marsh, Nagengast, & Scalas, 2014). Sampling error can also lead to biased parameter estimates. The type and amount of bias caused by sampling error depends on the sampling ratio within Level-2 units. Not correcting for sampling error can lead to an underestimation of compositional effects, whereas correcting for such error can lead to an overestimation when sampling ratios are high (Marsh et al., 2009). Procedures for deriving point estimates for the true effect are not available; therefore, a reasonable strategy is to calculate both types of coefficients to obtain estimates for the range of possible true effects. Combined with a correction for measurement error, and using terms from Marsh et al.'s (2009; Lüdtke et al., 2008) taxonomy of multi-level models, this implies using both latent-manifest modeling (correcting for measurement error but not sampling error) and doubly-latent modeling (correcting for both types of errors) to estimate compositional effects.

The IRT scores from the mathematics tests were used as latent estimates for achievement. For the emotions, we followed a two-step procedure. We first used confirmatory factor analysis (CFA) to derive latent factor score estimates for the emotions. Following recommendations by

Pekrun et al. (2011), a correlated uniqueness approach was employed by including correlations between residuals for items representing the same setting (attending class, doing homework, and taking tests in mathematics). Subsequently, we used the factor score estimates in multi-level modeling. We did not use multi-level modeling to estimate the latent emotion variables due to insufficient between-class variance for the emotion items. With insufficient Level-2 item variances, item-based multi-level modeling is not suited to produce stable and meaningful solutions (Marsh et al., 2009). As such, the two-step procedure was used to obtain latent measures while minimizing problems of non-convergence and model identification.

Based on these measures, we constructed latent-manifest models in which emotion is predicted by student-level and class-level achievement scores (see Supplemental Materials for the equations used to estimate these models). These models do not correct for sampling error; class-level achievement scores are computed by aggregating individual achievement scores within each class rather than estimating them as class-level latent variables (Marsh et al., 2009). The student-level achievement scores are group-mean centered. For each classroom, most of the students participated in the study (i.e., the sampling ratios within classes were high). As such, we considered these models as adequate (see Lüdtke et al., 2008; Marsh et al., 2009). However, to explore the possible impact of sampling error even under the present conditions of high sampling ratios, and following the methodology described by Marsh et al. (2009), we additionally estimated doubly-latent models calculating the Level-2 achievement scores as class-level latent variables.

To estimate the model parameters, the robust maximum likelihood estimator (MLR) was employed which is robust to non-normality of the observed variables. The unstandardized compositional effects were calculated by subtracting the individual-level effects of achievement on emotion from the class-level effects (see Kreft & De Leeuw, 1998). Subsequently, the standardized compositional effects (i.e., effect sizes for the compositional effects) were estimated using Effect Size 2 (ES 2; Marsh et al., 2009; also see Morin et al., 2014). ES 2 is comparable to Cohen's (1988) *d*: ES = (2 * B * $SD_{\text{predictor}}$) / SD_{outcome} where B is the unstandardized coefficient for the compositional effect, $SD_{\text{predictor}}$ is the standard deviation of the predictor, and SD_{outcome} is the Level 1 standard deviation of the outcome.

There were a few missing data (0.79%). To make full use of the data from students with missing data, we applied the full information maximum likelihood method (FIML; Enders, 2010). FIML has been found to result in trustworthy, unbiased estimates for missing values even in the case of large numbers of missing values (Enders, 2010) and to be an adequate method to manage missing data in studies with large samples (Jeličič, Phelps, & Lerner, 2009). All of the models were saturated.

Results

Preliminary analysis. Descriptive statistics and Pearson product-moment correlations for the study variables are presented in Table 1. Enjoyment correlated negatively with anger, anxiety, and hopelessness, and the three negative emotions showed positive intercorrelations, as is typical for achievement emotions (e.g., Pekrun et al., 2011). In addition, enjoyment correlated positively with achievement, and all three negative emotions correlated negatively with achievement. Furthermore, the results of CFA support the measurement quality of the emotion constructs (see supplemental materials Table 1 for factor loadings, factor determinacy scores, and correlations of factor score estimates). All factor determinacy scores were > .93, demonstrating that the factor score estimates correlated highly with their respective factors, and the correlations among the score estimates attest to their validity. The CFA model showed a good fit to the data, χ^2 (254) =

1108.139, $p < .001$; comparative fit index (CFI) = .959; root mean square error of approximation (RMSEA) = .046; and standardized root mean square residual (SRMR) = .042.

Multi-level modeling. The results of latent-manifest modeling support the proposed effects of achievement on emotions (see Table 2 and Figure 1; unstandardized coefficients in supplemental materials Table 4). Student-level achievement was positively linked to enjoyment and negatively linked to anger, anxiety, and hopelessness. In contrast, class-level achievement had a negative compositional effect on enjoyment and positive compositional effects on anger, anxiety, and hopelessness. In doubly-latent modeling, the compositional effect estimates were essentially the same, although slightly higher (supplemental materials, Table 8). In addition, gender was positively linked to enjoyment and negatively linked to anger, anxiety, and hopelessness (Table 2), indicating that male students reported more enjoyment and less anger, anxiety, and hopelessness than female students in mathematics. Grade level was positively related to anger and hopelessness, showing that older students reported more math-related anger and hopelessness.

Discussion

Based on a large-sample cross-sectional dataset, the findings confirm the expected links between achievement and emotions. Individual achievement was positively related to students' enjoyment and negatively related to their anger, anxiety, and hopelessness. In contrast, class-level achievement had a negative compositional effect on enjoyment and positive compositional effects on anger, anxiety, and hopelessness, supporting the happy-fish-little-pond effect proposed in our theoretical model. These relations were robust when controlling for students' gender and grade level, indicating that they were not mere artifacts of these variables. Moreover, the relations were quite substantial. Effect sizes for the compositional effects ranged from .392 to .962 in the latent-manifest models and from .416 to 1.032 in the doubly-latent models. Across all four emotions included, these findings provide strong initial support for our theoretical propositions. However, interpretation of the findings is clearly limited by the cross-sectional design of the study, a limitation we attend to in Studies 2 and 3.

Study 2

Study 2 replicated and extended the Study 1 findings by using a longitudinal design and adding mathematics pride, shame, and self-concept. The study used grade 5 and 6 data from the longitudinal study that is part of the same multi-study project as Study 1 (PALMA; Arens et al., 2017; Marsh et al., 2017; Murayama et al., 2013, 2016; Pekrun et al., 2007, 2017). To explain, the PALMA project contains multiple independent studies. Study 2 used a different sample of students and classes from different schools than Study 1. As such, the data from the two studies are independent. We also used the data from this longitudinal study to conduct supplemental cross-sectional analyses that investigated compositional effects of achievement on emotion in the remaining waves of the study (grades 7-9), and to conduct supplemental analyses of compositional effects while controlling for teacher behavior (for a summary of the supplemental analyses, see the Results section; for details, see the Supplemental Materials available online).

Method

Samples and procedure. The study included annual assessments in grades 5 and 6 that took place towards the end of the school year. Intact classes were sampled. Sampling and the assessments were conducted by the Data Processing and Research Center (DPC) of the International Association for the Evaluation of Educational Achievement (IEA). The full student sample included 2,070 students from 83 classrooms in grade 5, and 2,059 students from 81 classrooms in grade 6. For the present analysis the sample was restricted to students from 78 classrooms in which at least ten students participated in both assessments, thus establishing a reasonable stability of classroom composition as well as sampling ratio within classrooms (for similar procedures, see Marsh, Köller, & Baumert, 2001; Prenzel, Carstensen, et al., 2006).

The resulting sample included 1,759 students (50.0% female; mean starting age = 11.73 years, *SD* = 0.49) from 78 classrooms in 42 schools. The sample was drawn from the same three tracks of the secondary school system in Bavaria as in Study 1 (*n*s = 579, 500, and 680 students from 32, 19, and 27 classrooms in Hauptschule, Realschule, and Gymnasium, respectively) and was representative of the student population in this state in terms of students' school track as well as characteristics such as gender, urban versus rural location, and family background (SES; see Pekrun et al., 2007). The mean number of participating students per class was 22.6 (minimum = 12, maximum = 32, *SD* = 5.60). All instruments were administered in the students' classrooms by trained external test administrators. Student participation was based on written parental consent, and students' responses were kept strictly confidential.

Variables and measures. The study variables included students' achievement, enjoyment, pride, anger, anxiety, shame, hopelessness, and self-concept of ability in mathematics and demographic variables. Achievement, emotions, and self-concept were assessed at both waves.

Achievement. IRT scores from the same PALMA Mathematical Achievement Test as in Study 1 (Pekrun et al., 2007) were used to assess students' mathematics achievement.

Emotions. Scales of the same Achievement Emotions Questionnaire-Mathematics (AEQ-M; Frenzel et al., 2007; Pekrun et al., 2011) as in Study 1 were used to assess math-related enjoyment, pride, anger, anxiety, shame, and hopelessness (9, 6, 8, 15, 8, and 6 items, respectively). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale, and the scores were averaged to form the emotion indexes (α range = .86–.90; supplemental materials Table 1).

Self-concept. The PALMA mathematics self-concept scale (Arens et al., 2017; Marsh et al., 2017; Pekrun et al., 2007) was used to measure self-concept (6 items; e.g., "In mathematics, I am a good student"; "It is easy for me to learn mathematics"). Participants responded on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale, and the scores were averaged to form the math self-concept index (αs = .88 and .90 at grades 5 and 6, respectively; supplemental materials Table 1).

Demographic variables. Gender and family SES were controlled in the analysis. Gender was coded 1 = female, 2 = male. SES was assessed by parent report using the EGP classification (Erikson, Goldthorpe, & Portocarero, 1979), which consists of six ordered categories of parental occupational status. Higher values represent higher SES.

Multi-level modeling. The same multi-level modeling approach with Mplus (Version 8, Muthén & Muthén, 2017) as in Study 1 was used. CFA was used to derive estimates for latent emotion and self-concept measures and to test their measurement invariance over time. Subsequently, we employed these measures to estimate longitudinal multi-level models. In these models, we examined the predictive effects of achievement, emotion, and self-concept at grade 5 on the same variables at grade 6. In each of the models, we controlled for autoregressive and reciprocal effects over time. Furthermore, we controlled for effects of gender and SES on all study variables at both waves. We estimated models of achievement and emotion not including self-concept as well as models that included self-concept. As in Study 1, the MLR estimator was used. There were a few missing data (1.12%), primarily because information about family SES was not available for some students (1.36% missed information about SES but had no other missing data). The full information maximum likelihood method (FIML; Enders, 2010) was used to deal with these missing data. All models were saturated.

In the models not including self-concept (Model 1), Time 1 achievement and emotion served as predictors and Time 2 achievement and emotion as outcome variables. These models estimated the predictive effects of achievement on emotion over time while controlling for autoregressive and reciprocal effects. In the models including self-concept (Model 2), Time 1 achievement, self-concept, and emotion served as predictors, and the same variables at Time 2 served as outcome variables (see Figure 2). These models estimated predictive effects of achievement on self-concept, of self-concept on emotion, and of achievement on emotion over time while controlling for prior levels and reciprocal effects of all three variables.

Given the two-wave design of the analysis, it is not possible to derive estimates for the indirect effects of achievement on emotion as mediated by self-concept within the one-year time span of the study. However, if both the effect of Time 1 achievement on Time 2 self-concept and the effect of Time 1 self-concept on Time 2 emotion are significant and substantial, the evidence would be consistent with the hypothesis that self-concept can act as a mediator in the relation between achievement and emotion. Even if not sufficient to infer firm conclusions about mediation, effects of the predictor on the mediator, and of the mediator on the outcome, are two important requirements that need to be met for mediation (MacKinnon, Fairchild, & Fritz, 2007). As such, we inspected the significance and size of these effects to judge support for mediation.

Results

Preliminary analysis. Descriptive statistics and Pearson product-moment correlations for the study variables are presented in Table 1. Enjoyment and pride correlated positively, as did the negative emotions. The correlations between these two groups of emotions were negative. In addition, both positive emotions as well as self-concept correlated positively with achievement, and all of the negative emotions correlated negatively with achievement.

The results of CFA document the measurement quality of the emotion and self-concept constructs. As in Study 1, correlations between residuals for items representing the same setting (attending class, doing homework, and taking tests) were included. In addition, residuals of the same items over time were allowed to correlate (see Tables S1 and S2 for factor loadings, factor determinacy scores, correlations of factor score estimates, and goodness-of-fit indexes). All factor determinacy scores were > .91, demonstrating that the factor score estimates correlated highly with their respective factors, and the correlations among the score estimates attest to their validity.

To test measurement invariance over time, we sequentially evaluated models of configural, metric, scalar, and residual invariance. Configural invariance is defined by equal patterns of factor loadings; metric invariance additionally requires equal size of factor loadings; scalar invariance requires equal factor loadings and intercepts; and residual invariance requires equal factor loadings, intercepts, and residual variances. To establish construct equivalence for analyzing correlations and path coefficients over time, metric invariance is the minimum needed (Chen, 2007; Steenkamp & Baumgartner, 1998).

To evaluate invariance, we followed recommendations by Chen (2007). Provided adequate sample size, for testing metric invariance, a change of ≥ -.010 in CFI, supplemented by a change of ≥ .015 in RMSEA or a change of ≥ .030 in SRMR would indicate noninvariance; for testing scalar or residual invariance, a change of ≥ -.010 in CFI, supplemented by a change of ≥ .015 in RMSEA or a change of ≥ .010 in SRMR would indicate noninvariance. In addition, we inspected the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the root deterioration per restriction index (RDR; Browne & Du Toit, 1992). As suggested by Raykov and

Penev (1998), we interpreted RDR in relation to recommendations for the RMSEA. Following this rationale, RDR < .05 would indicate strong equivalence of nested models, and RDR < .08 an acceptable level of equivalence. As recommended, we did not use the χ^2 difference test because it is overly sensitive to sample size (Marsh, Balla, & McDonald, 1988).

The results show that configural, metric, scalar, and residual invariance can be assumed for all of the emotion and self-concept constructs (see supplemental materials Table 2). All of the models showed a good fit to the data. In the metric invariance models, fit indexes were CFI > .963, RMSEA < .042, and SRMR < .038 for all of the emotion and self-concept constructs. Fit was good even for the most restrictive model (residual invariance), with CFI > .959, RMSEA < .043, and SRMR < .058 for all constructs. For all sequential comparisons of models with increasing restrictions on invariance, Δ CFI was < .005, Δ RMSEA was < .006, Δ SRMR was < .020, and RDR was < .056 (supplemental materials Table 2).

Multi-level modeling. Replicating the Study 1 findings, the model estimates supported our hypotheses (Tables 3 and 4 and supplemental materials Table 9; for the unstandardized coefficients, see supplemental materials Tables 5 and 6).

Predictive effects of achievement on emotion and self-concept. Individual achievement was a positive predictor of enjoyment and pride and a negative predictor of anger, anxiety, shame, and hopelessness. In contrast, class-level achievement had negative compositional effects on enjoyment and pride and positive compositional effects on the negative emotions. These effects were robust across models despite substantial autoregressive effects of emotion over time. An exception was the positive compositional effect on shame which was not significant (p =.096). Students' shame in an academic context may primarily be driven by social interaction in the classroom rather than social comparison with others' achievement. In doubly-latent models, estimates for compositional effects were essentially the same, although slightly higher (supplemental materials Table 9).

In the models including self-concept (Table 4), student-level achievement had positive predictive effects on self-concept, whereas class-level achievement had negative effects, despite substantial autoregressive effects of self-concept over time. These findings support the big-fish-little-pond effect on self-concept within a controlled longitudinal design. Again, the compositional effects were essentially the same in doubly-latent modeling (supplemental materials Table 9). Self-concept, in turn, had positive predictive effects on the positive emotions and negative effects on the negative emotions, despite substantial autoregressive effects of emotions over time. Taken together, these findings are consistent with the hypothesis that self-concept might be a mediator in the predictive effects of achievement on emotion.

Reciprocal effects. The emotions had reciprocal student-level effects on achievement over time. Enjoyment had positive predictive effects on achievement, and anger, anxiety, shame, and hopelessness had negative effects (Table 3). These effects suggest that enjoyment can promote students' performance whereas negative emotions undermine performance. Combined with the predictive effects of achievement on emotion, the pattern of findings suggests that individual achievement and emotion were linked by reciprocal causation over time. Similarly, self-concept had positive reciprocal effects on achievement (Table 4); combined with the predictive effects of achievement on self-concept, these findings suggest that achievement and self-concept were also linked by reciprocal causation.

Same-variable compositional effects. Class-level achievement and self-concept had positive predictive effects on the same variables at the student level over time. Specifically, class-level achievement at Time 1 had positive compositional effects on individual achievement at

Time 2 (“peer achievement spill-over effect”; Fruehwirth, 2013). Unstandardized coefficients ranged from .370 to .431, and ES 2 from .691 to .804 across models (all *p*s < .001; the unstandardized coefficients are derived by subtracting the student-level effects from the class-level effects of Time 1 achievement on Time 2 achievement as depicted in Tables S5 and S6). These effects suggest that class-level achievement positively influenced individual achievement controlling for prior achievement, in line with some previous studies on the effects of the achievement composition of classrooms (e.g., De Fraine et al., 2003; Stäbler et al., 2017; but see Dicke et al., 2018, Marsh, 1991, and Televantou et al., 2015, for negative effects).

Similarly, class-average self-concept had positive compositional effects on individual self-concept over time. This is a new finding. Unstandardized coefficients for this effect ranged from .490 to .954, and ES 2 from .231 to .449 (the unstandardized coefficients are derived by subtracting the student-level effects from the class-level effects of Time 1 self-concept on Time 2 self-concept as depicted in supplemental materials Table 6; all *p*s < .05). The effect may be attributable to social transmission of self-confidence among members of the same classroom. For emotion, there also were positive compositional effects over time, but none of these effects was significant due to lack of variance of the emotion scores between classes.

Relations for gender. In line with the Study 1 findings, gender related positively to math-related enjoyment and pride and negatively to anxiety, shame, and hopelessness, indicating that male students reported more enjoyment and pride, and less negative emotions, than female students (supplemental materials Table 10). In addition, gender related positively to achievement and self-concept, showing that male students had higher achievement scores and reported higher self-concepts in mathematics.

Supplemental analyses. The PALMA longitudinal study comprises annual assessments from grades 5 to 9 (Pekrun et al., 2007). To control for the robustness of the results in the remaining waves (i.e., grades 7 to 9), we replicated the analysis for each of these waves in supplemental analyses (see Supplemental Materials). Because the change in class compositions from year to year after grade 6 precluded a longitudinal analysis, we conducted separate analyses for these waves. The results show the same pattern as in the grade 5-6 analysis (supplemental materials Table 11). An exception is the coefficients for shame which were not significant in the longitudinal grade 5-6 analysis but were significant in each of the cross-sectional analyses for grades 7 to 9. In the latent-manifest models, the standardized compositional effects of class-average achievement on enjoyment (ES 2) were -.711, -.865, -.951 for grades 7, 8, and 9, respectively (all ps < .001). For pride, the coefficients were -.587, -.719, -.770, respectively. For anger, they were .573, .668, and .624; for anxiety, .521, .571, and .568; for shame, .268, .296, and .263; and for hopelessness .572, .596, and .659 (all ps < .001, except for shame ps < .01). In latent-latent analysis, the results were essentially the same.

To meet a concern that the compositional effects of class-average achievement may be caused by differences in teacher behavior across classes rather than achievement, we controlled for teacher behavior in additional analyses. Specifically, it could be that teachers of high-achieving classes are more enthusiastic about teaching, or that they are stricter and exert more pressure on students than teachers of low-achieving classes. As such, we controlled for student-perceived teacher enthusiasm and pressure for achievement. These two variables have been found to be prime drivers of students' achievement emotions in the classroom. Teacher enthusiasm likely increases both perceived control and the perceived value of achievement, which explains their positive relations with students' enjoyment of learning (see, e.g., Frenzel et al., 2009, 2017).

In contrast, achievement pressure reduces control and has been found to exacerbate anxiety (Zeidner, 1998, 2014).

When including these variables as covariates in the latent-manifest longitudinal models for grades 5 to 6, the compositional effects of class-average achievement on emotions remained fully robust (see supplementary materials, Table 12). The standardized compositional effects (ES 2) were -.280 ($p < .01$), -.240 ($p < .01$), .327 ($p < .01$), .297 ($p < .01$), and .333 ($p < .001$) for enjoyment, pride, anger, anxiety, and hopelessness, respectively. As in the main analyses, the effect for shame was positive but not significant, ES 2 = .123 ($p = .161$). In the latent-latent models, the results were essentially the same. Similarly, when controlling for student-perceived teacher enthusiasm and teacher pressure for achievement in the cross-sectional grade 7, 8, and 9 analyses, the findings replicated the results not including these variables (supplemental materials Table 12). For enjoyment and pride, the compositional effects were again negative and significant (range of standardized effects [ES 2] = -.449 to -.713, all *p*s < .001). For anger, anxiety, shame, and hopelessness, the effects were again positive and significant (range of standardized effects [ES 2] = .198 to .540; all *p*s < .001, except for shame *p*s < .05). In the latent-latent models, the effects were essentially the same.

Discussion

The study findings replicate the results of Study 1 and fully support our theoretical model. The findings again confirm that individual achievement positively predicts achievement-related enjoyment and negatively predicts anger, anxiety, and hopelessness. They also confirm that group-level achievement has a negative compositional effect on enjoyment and positive compositional effects on the negative emotions, supporting the happy-fish-little-pond effect on emotion.

Beyond replication, the findings extend the results of Study 1 in several important ways. Based on a longitudinal dataset, the findings document that the predictive effects of achievement on emotion extend across an entire school year. This finding attests to the longevity of both individual-level and compositional effects of achievement on emotion. Furthermore, the findings suggest that achievement also influences individuals' pride and shame, two emotions that have been regarded as particularly important for achievement strivings (Heckhausen, 1991). The findings indicate that individual achievement promotes pride and reduces shame, whereas group-level achievement reduces individuals' pride and exacerbates their shame when controlling for individual achievement. The results of supplemental analyses showed that these results were robust when controlling for teachers' enthusiasm and pressure for achievement, indicating that the predictive effects of achievement were not confounded with effects of these teacher behaviors.

The findings from the models including self-concept also support the big-fish-little-pond effect of group-level achievement on self-concept. Furthermore, the results show that self-concept of ability, in turn, positively predicted students' enjoyment and pride, and negatively predicted their anger, anxiety, shame, and hopelessness. This pattern of findings is consistent with propositions of Pekrun's (2006) control-value theory of achievement emotions and documents the importance of self-concept of ability for achievement emotions. Both the predictive effects of achievement on self-concept, and the predictive effects of self-concept on emotion, were substantial and significant for all of the emotions. Taken together, these effects are in line with the hypothesis that self-concept can act as a mediator in the relation between achievement and emotion.

Study 3

Study 3 used a longitudinal dataset (grade 9 to 10) with 15-year old students to examine the generalizability of the findings across age groups. The study was based on the German longitudinal extension of the 2003 cycle of the *Programme for International Student Assessment* (PISA), the PISA-I-Plus study (Prenzel, Baumert, et al., 2006).[3] As mathematics was the major domain in PISA 2003, the PISA-I-Plus study examined students' development in mathematics.

Method

The study is a secondary analysis of an existing dataset. The PISA-I-Plus study received Institutional Review Board approval from the Organization for Economic Co-operation and Development (OECD) and the Standing Conference of the Ministers of Education and Cultural Affairs of the "Länder" in the Federal Republic of Germany (Kultusministerkonferenz; see Prenzel, Carstensen, Schöps, & Maurischat, 2006).

Samples and procedure. For the PISA-I-Plus study, intact classes including at least 10 students who participated in both assessments (grades 9 and 10) were selected from the German schools participating in the international PISA 2003 study (Prenzel, Baumert, et al., 2006). As in Study 2, sampling and the student assessments were performed by the Data Processing Center (DPC) of the International Association for the Evaluation of Educational Achievement (IEA). The sample consisted of N = 4,353 students (57.2% female; mean starting age = 15.58 years, SD = .55) from 194 classrooms located in 119 schools representing all major school types that exist in Germany, except the lower-track Hauptschule from which students graduated at the end of grade 9. Schools included the medium-track Realschule, the high-track Gymnasium, the comprehensive Integrierte Gesamtschule, and the Mittelschule which integrates low and medium tracks (Ns = 1,859, 1,864, 228, and 402 students from 81, 80, 13, and 20 classes in Realschule, Gymnasium, Gesamtschule, and Mittelschule, respectively). The mean number of participating students per class was 22.4 (minimum = 12, maximum = 30, *SD* = 4.26). The assessments took place towards the end of the school year. All instruments were administered in the students' classrooms by trained external test administrators. Student participation was based on written parental consent, and students' responses were kept strictly confidential.

Variables and measures. The study variables included students' achievement, enjoyment, anger, anxiety, hopelessness, and self-concept of ability in mathematics and demographic variables. Pride and shame had not been assessed in the PISA-I-Plus study.

Achievement. The PISA mathematics test was used to assess achievement. The PISA test is a standardized achievement test assessing the competencies of 15-year olds across a wide range of ability. The database of the PISA study provides five plausible values based on item response theory scaling, rather than one single score, to estimate a student's mathematics achievement; this procedure was used to avoid biased population estimates of achievement (Davier, Gonzalez, & Mislevy, 2009; Prenzel, Carstensen, et al.. 2006). In accordance with the recommendations of the PISA documentation (Organization for Economic Cooperation and Development, 2005), we did not average plausible values within students, but conducted the analyses separately for the five plausible values and then averaged the resulting parameters.

Emotions. The PISA-I-Plus study used short scales from the same Achievement Emotions Questionnaire-Mathematics (AEQ-M; Frenzel et al., 2007; Pekrun et al., 2011) that was employed in Studies 1 and 2. The scales assessed enjoyment, anger, anxiety, and hopelessness (6, 4, 9, and 6 items, respectively). Participants responded on a 1 (*strongly disagree*) to 4 (*strongly agree*) scale, and the scores were averaged to form the emotion indexes (α range = .83–.93; Table 1).

Self-concept. The PISA 2003 mathematics self-concept scale was used (5 items; e.g., "I get good marks in mathematics"; "I learn mathematics quickly"). Participants responded on a 1 (*strongly agree*) to 4 (*strongly disagree*) scale. The scores were averaged to form the self-concept index (αs = .91 and .92 at grades 9 and 10, respectively; Table 1).

Demographic variables. Students' gender and family SES were included in the analysis. Gender was coded 1 = female, 2 = male. As in Study 2, SES was assessed using the EGP index (Erikson, Goldthorpe, & Portocarero, 1979), with higher values representing higher SES.

Multi-level modeling. We used the same data analysis procedures as in Study 2. Again, the MLR estimator was used. There were a few missing data (3.32%), primarily because some students had not participated in either the grade 9 emotion assessment (1.47%) or the grade 10 assessment (4.20%). The full information likelihood method (FIML; Enders, 2010) was employed to deal with missing data. All models were saturated.

Results

Preliminary analysis. Descriptive statistics and Pearson product-moment correlations for the study variables are presented in Table 1. As in Studies 1 and 2, enjoyment correlated negatively with anger, anxiety, and hopelessness, and the intercorrelations of the negative emotions were positive. Enjoyment and self-concept correlated positively with achievement, and the negative emotions correlated negatively with achievement. CFAs were used to test the measurement properties of the emotion and self-concept constructs. Again, we included correlations between residuals for items representing the same setting and between residuals for the same items over time (see Tables S1 and S3 for factor loadings, factor determinacy scores, correlations of factor score estimates, and goodness-of-fit indexes). All factor determinacy scores were > .90, demonstrating that the factor score estimates correlated highly with their respective factors, and the correlations among the score estimates attest to their validity.

Furthermore, the results show that configural, metric, scalar, and residual invariance can be assumed for all of the constructs (supplemental materials Table 3). All of the models showed a good fit to the data. In the metric invariance models, fit indexes were CFI > .974, RMSEA < .051, and SRMR < .036 for all constructs. Fit was also good for the most restrictive model (residual invariance), with CFI > .971, RMSEA < .048, and SRMR < .037 for all constructs. For all comparisons of models with increasing restrictions on invariance, Δ CFI and Δ RMSEA were < .004, Δ SRMR was < .007, and RDR was < .049.

Multi-level modeling. Replicating the findings of Studies 1 and 2, the estimates again supported the proposed effects (Tables 5 and 6 and Supplemental Materials, Tables S7–S8). In longitudinal modeling, individual achievement positively predicted enjoyment and negatively predicted anger, anxiety, and hopelessness, whereas class-average achievement had negative compositional effects on enjoyment and positive compositional effects on the negative emotions. These results were robust across models despite substantial autoregressive effects of emotion over time. In doubly-latent modeling, estimates for compositional effects were essentially the same, although slightly higher (supplemental materials Table 9).

The results of the models including self-concept also replicated the Study 2 findings, with student-level and class-level achievement predicting self-concept, and self-concept predicting the emotions. The compositional effects of class-average achievement on the emotions were substantially reduced, and the effects on enjoyment, anger, and hopelessness were no longer significant with self-concept of ability in the equation (Table 6). These findings are consistent with the hypothesis that self-concept might be a mediator in the predictive effects of achievement on emotion.

In addition, there were again reciprocal effects of emotion on achievement, in line with reciprocal effects models of these linkages (Pekrun et al., 2017). In contrast to the Study 2 findings, class-level achievement had no compositional effects on individual achievement over time (unstandardized coefficients < .071; ES 2 < .118; *p*s > .372; see Tables 5, 6, S7, and S8). As in Study 2, there were positive compositional effects of class-level self-concept on individual self-concept, but these effects were not significant either. Finally, consistent with the Studies 1 and 2 findings, gender related positively to math-related enjoyment and pride, and negatively to anger, anxiety and hopelessness (supplemental materials Table 10), indicating that male students reported more positive emotions and less negative emotions than female students. Gender also related positively to achievement and self-concept, showing that male students had higher performance scores and more self-confidence in mathematics.

Discussion

Using a large-scale longitudinal dataset, the findings replicated the results of Study 2. Extending Study 2, they show that these results hold for older students as well. The findings again confirm that individual achievement is a positive predictor of achievement-related enjoyment and a negative predictor of anger, anxiety, and hopelessness, whereas the compositional effects of group-level achievement are negative for enjoyment and positive for the negative emotions, in line with the proposed happy-fish-little-pond effect on emotion. As in Study 2, these relations proved not to be short-lived, but to extend over the course of an entire school year. Furthermore, as in Study 2, the findings document the big-fish-little-pond effect on self-concept, and they support the hypothesis that self-concept might be a mediator both for the individual-level effects and the compositional effects of achievement on emotions.

General Discussion

Achievement emotions are critically important for engagement, learning, and performance in achievement contexts, and for personal growth and well-being more generally (Ashkanasy & Humphrey, 2011; Elfenbein, 2007; Pekrun, 2017; Pekrun & Linnenbrink-Garcia, 2014; Zeidner, 2014). Given the relevance of these emotions, it is important to acquire information about their origins. In this research, we considered two particularly promising candidates to explain the development of achievement emotions, namely, individual and group-level achievement. We developed a reference group model describing the possible effects of achievement on emotion, and we tested the model in three separate large-scale empirical studies.

By including both individual and contextual variables, this research contributes to our understanding of individual differences as well as the impact of social groups. The results suggest that personality and social psychological research on the influence of groups should attend to the emotional consequences of group composition. In addition, understanding how group composition influences emotion can also inform other disciplines. To the extent that the present findings are generalizable, they are relevant for sociology and management research as well as applied fields within psychology, such as educational, occupational, performing arts, and sports psychology.

Relations of Achievement to Emotion

Our theoretical model posits that individual achievement is related to more positive emotions and less negative emotions, whereas group-level achievement is related to less positive emotions and more negative emotions. Both of these propositions were supported across all three studies. From a total of 28 relevant path coefficients (student-level and compositional effects of achievement on four emotions in Studies 1 and 3 and six emotions in Study 2), all were in the expected direction, and all but one were significant. The findings were robust when controlling for autoregressive effects, reciprocal effects, gender, SES, and teacher behavior, and they were consistent across studies, analytical designs (synchronous vs. longitudinal), modeling procedures (latent-manifest vs. doubly-latent), and a range of age groups and grade levels. The size of the path coefficients was substantial and increased further across age, as can be seen from the Studies 2 and 3 results (Tables 3–6 and S9). Overall, these findings attest to the importance of both individual achievement and group-level achievement as antecedents of achievement emotions.

The Role of Self-Concept

Based on our theoretical model, we expected that achievement predicts self-concept of ability, and that self-concept, in turn, predicts emotions. In two of the three studies, measures of self-concept were included to test these propositions. By linking the constructs of achievement emotion and self-concept, the present research contributes to the integration of two research traditions that have worked in relative isolation to date. The results support the proposed relations of self-concept of ability to achievement and emotions.

First, the results confirmed prior evidence that individual achievement is positively related to students' self-concept of ability, whereas group-average achievement is negatively related to self-concept when controlling for individual achievement (Marsh et al., 2008). Whereas most previous studies used school-level achievement to examine compositional effects on self-concept, the present research demonstrates that these effects also occur when considering classes as reference groups, in line with a few previous studies that considered class-level achievement (see Marsh, Köller, & Baumert, 2001; Trautwein, Gerlach, & Lüdtke, 2008). In fact, to the extent that the peers in one's class represent the immediate environment for evaluating achievement relative to others, studies considering class-level achievement may be particularly well suited to examine compositional effects of achievement, in line with the local dominance effect (Alicke et al., 2010; Zell & Alicke, 2009).

Self-concept, in turn, was a positive predictor of positive emotions and a negative predictor of negative emotions. This finding supports Pekrun's (2006) control-value theory of achievement emotions which posits that achievement emotions are shaped by competence beliefs. The two-wave design of our longitudinal studies (i.e., Studies 2 and 3) precluded longitudinal tests of mediation. However, taken together, the predictive effects of achievement on self-concept, and of self-concept on emotion, support the hypothesis that self-concept is a mediator in the effects of both individual and group-level achievement on emotion.

Reciprocal Effects and Spill-Over Effects

We found that achievement and emotion were linked by reciprocal relations over time, which confirms the functional importance of emotions for performance and supports reciprocal causation models of emotion (Pekrun et al., 2017). Achievement was a positive predictor of positive emotions and a negative predictor of negative emotions controlling for prior emotion. Positive emotions, in turn, were positive predictors of achievement, and negative emotions were negative predictors, controlling for prior achievement. These findings suggest that achievement and emotions are linked by virtuous and vicious cycles, with success and positive emotions mutually reinforcing each other, and failure being reciprocally linked with negative emotions.

Can the negative emotional effects of being in a high-performing group be outweighed by beneficial effects on achievement (peer achievement spill-over effects; Fruehwirth, 2013)? The findings show that class-level achievement had positive compositional effects on individual achievement over time in Study 2 but not in Study 3. Such a lack of consistently beneficial effects is counter to traditional expectations, but mirrors the mixed findings from other studies. Some studies found positive effects of group-average achievement (e.g., Burns & Mason, 2002; De Fraine et al., 2003; Opdenakker & van Damme, 2001; Stäbler et al., 2017), whereas others found null effects or even negative effects (Dicke et al., in press; Marsh, 1991; Televantou et al., 2015). As argued by Dicke et al. (in press; also see Hauser, 1970; Harker & Tymms, 2004), it is important to use longitudinal designs and to control for autoregressive effects, measurement error, and pre-existing differences between students when examining these effects, because not including these controls can generate positive estimates ("phantom effects;" Harker & Tymms, 2004) when the true effect is in fact negative or zero. Overall, the evidence suggests that being in a high-achieving group is not a reliable predictor of benefits for individual achievement.

In fact, given the negative compositional effects of group-level achievement on individual emotions, it is plausible to assume that group-level achievement can also have negative effects on individual achievement. The present findings document negative compositional effects of class-level achievement on self-concept and positive emotions, and positive compositional effects on negative emotions. Self-concept and positive emotions, in turn, had positive reciprocal effects on individual achievement, and negative emotions had negative effects. Taken together, this pattern of findings suggests that group-level achievement can have negative effects on individual achievement mediated by reduced self-confidence, reduced enjoyment, and increased negative emotions such as anxiety and hopelessness.

However, to more comprehensively judge the effects of being in a high- versus low-achieving group, it would also be necessary to consider effects on outcomes other than emotions or performance as measured in traditional cognitive assessments. For example, students in high-achieving groups may be exposed to higher-level thinking or to more advanced methods and ideas. Furthermore, across education, work, and sports domains being in high-achieving groups may help individuals connect to unique social networks that can assist them socially and professionally. It is a task for future research to consider a broader range of outcomes and to disentangle effects of group-level achievement from associated variables that are extraneous to achievement itself, such as institutional resources or quality of leadership.

Limitations and Future Directions

The findings of this research provide robust support for our proposed reference group model. Nevertheless, a number of limitations may be noted and used to suggest directions for future research. First, as compared with experimental studies, the power of non-experimental field studies to derive causal conclusions is limited. In the absence of random assignment with experimental manipulation, or even when there is random assignment, typically there are alternative explanations of the effects. As such, although the present research used longitudinal structural equation modeling and controlled for related variables, autoregressive effects, and reciprocal effects, the possibility still exists that our findings are attributable to other variables that were not included in the study. A related limitation is the use of latent factor score estimates which implies some degree of factor indeterminacy, although small in the present research according to the factor determinacy scores (supplemental materials Table 1).

On the other hand, field studies may be more ecologically valid than experimental emotion studies, which are limited in terms of situational representativeness and ethical concerns about experimentally manipulating emotions. Furthermore, statistical power is higher in field studies such as those presented here due to large sample size. The design of our research, including large-scale multi-level longitudinal analyses controlling for measurement error and sampling error, is arguably stronger than most previous nonexperimental research on compositional effects (Gonzáles-Romá & Hernández, 2017).

To balance the benefits and drawbacks of different methodologies and make headway in this avenue of research, future investigations should further pursue the approach taken herein while complementing this approach with experimental studies. For future studies, it would also be important to consider additional third variables that could not be controlled using the present datasets, such as genetic dispositions that could influence the covariation between variables at the individual level. Studies using co-twin control or discordant-twin designs, for example, could help bolster the strength of conclusions drawn from the present investigation (McGue, Osler, & Christensen, 2010). In addition, longitudinal research should consider designs that involve more than two waves and more fine-grained temporal resolution to investigate the generalizability of the findings across different time scales, and to make it possible to directly analyze mediation by variables such as self-concept.

The present results pertain to the achievement emotions experienced by students in mathematics. It is open to question whether the same pattern of relationships holds for other academic domains. In addition, it also is open to question whether similar effects will be found for achievement contexts other than education. Our theoretical model suggests that individual and group-level achievement should exert similar effects across various contexts (in work, sports, performing arts, etc.). Future research should examine the proposed links in these non-academic contexts.

Furthermore, the present research included a range of age groups in adolescence. Given that self-concepts of ability and related achievement emotions develop prior to entering formal education, it is plausible to assume that the happy-fish-little-pond effect emerges early in the preschool years. In addition, as implied by social comparison theory and control-value theory, it is plausible to except that this effect also plays a major role in adulthood. The possible generalizability of the findings to younger as well as older age groups has yet to be examined, however. Similarly, the present research analyzed achievement emotions in one cultural context only. The links between individual achievement, appraisals, and achievement emotions have been found to be robust across cultural contexts (Pekrun, 2009, 2018), suggesting cross-cultural universality. However, the generalizability of the effects of group-level achievement on emotions across cultures has yet to be established.

Finally, while achievement and self-concept of ability are important antecedents of achievement emotions, other factors are likely to be relevant as well. The links between achievement, self-concept, and achievement emotions were far from perfect in the present data, suggesting the need to attend to additional processes to more fully account for the origins of achievement emotions. It is also important to note that both self-concept and emotion were assessed using self-report in this research. When using other measures of emotion (e.g., behavioral and physiological indicators), the link between the two variables may be less strong and leave more room for detecting other possible mediators in the achievement-emotion relation. From the perspective of control-value theory, important additional candidates include achievement goals and the perceived value of achievement (Pekrun, 2006; Pekrun et al., 2009). Future research should examine the interplay of these various individual antecedents in shaping achievement emotions.

Implications for Practice

From a policy and practice perspective, the findings should be of interest to individuals who make decisions about their membership in groups and participation in group-based programs. They also should be of interest to policy-makers, administrators, supervisors, coaches, and teachers shaping the composition of organizations, teams, and classes. As aptly argued by Harker and Tymms (2004), compositional effects are important to policy-making. The way in which organizations group people matters to both individual and collective outcomes. Trying to explain and understand compositional effects is therefore vital. This is likely to be true not only for organizational and educational settings, the domains for which recent advances in the study of compositional effects may have been strongest (Gonzáles-Romá & Hernández, 2017; Marsh et al., 2009). Compositional effects are equally relevant for institutions and groups in other contexts (economy, sports, the family, etc.; for similar arguments, see Jex & Bliese, 1999; Kriwy, Gross, & Gottburgsen, 2013; Maltby, Wood, Vlaev, Taylor, & Brown, 2012; Raudenbush & Bryk, 2002; Rose 2018; Rousseau, 1986).

More specifically, two important messages follow from the present research. First, the findings suggest that individual achievement may benefit emotional well-being, with successful performance facilitating the development of positive emotions and failure contributing to negative emotions. Accordingly, providing individuals with opportunities to experience success may help to promote their emotional well-being (e.g., by using intrapersonal standards to evaluate achievement and emphasizing mastery over competition goals; Pekrun, Cusack, Murayama, Elliot, & Thomas, 2014).

Second, the findings suggest that placing individuals in high-achievement groups may incur emotional costs. For future research and practice, it will be important to develop strategies to reduce the costs for emotional well-being that may be associated with membership in a high-achieving group. Alternatively, policy-makers could consider establishing systems with lower levels of achievement-based stratification (also see Parker, Jerrim, Schoon, & Marsh, 2016). In sum, by documenting the predictive power of individual and group-level achievement for explaining emotions, the present findings elucidate one important factor that can be targeted by practitioners and policy-makers to reduce individuals' negative affect and facilitate the development of their emotional well-being and mental health.

References

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

Arens, A. K., Marsh, H., W., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology, 109,* 621-634. http://dx.doi/10.1037/edu0000163

Ashkanasy, N. M., & Humphrey, R. H. (2011). Current emotion research in organizational behavior. *Emotion Review, 3,* 214-224. http://dx.doi.org/10.3390/ijerph10126215

Ashkanasy, N. M., Härtel, C. E. J., & Daus, C. S. (2002). Diversity and emotion: The new frontiers in organizational behavior research. *Journal of Management, 28,* 307-338.

Baratta, J., & McManus, M. A. (1992). The effect of contextual factors on individuals' job performance. *Journal of Applied Social Psychology, 22,* 1702-1710.

Beilock, S. L., Gunderson, E. A., Ramirez, G., Levine, S.C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107,* 1860-1863. http://dx.doi.org/10.1073/pnas.0910967107

Boehm, J. K., & Lyubomirsky, S. (2008). Does happiness promote career success? *Journal of Career Assessment, 16,* 101-116. http://dx.doi/10.1177/1069072707308140

Boekaerts, M. (1993). Anger in relation to school learning. *Learning and Instruction, 3,* 269-280. http://dx.doi/10.1016/0959-4752(93)90019-V

Browne, M. W., & Du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research, 27,* 269-300.

Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. *American Educational Research Journal, 39,* 207-233.

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: A review of current behavioral and psychophysiological research. *Current Opinion in Behavioral Science, 10,* 33-38. http://dx.doi.org/10.1016/j.cobeha.2016.04.011

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modelling, 14,* 464-504. http://dx.doi.org/10.1080/10705510701301834

Cohen, D. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale NJ: Lawrence Erlbaum.

Davier, M. von, Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series, Issues and Methodologies in Large-Scale Assessments* (Volume 2, pp. 9-36). Princeton, NJ: IEA-ETS Research Institute.

Davier, M. von, Xu, X., & Carstensen, C. H. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika, 76,* 318-336. http://dx.doi.org/10.1007/s11335-011-9202-z

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72,* 17-31.

De Fraine, B., van Damme, J., van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29,* 841-859. http://dx.doi.org/10.1080/0141192032000137330

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (in press). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*.

Efklides, A., & Volet, S. (Eds.). (2005). Feelings and emotions in the learning process [Special issue]. *Learning and Instruction, 15*(5). http://dx.doi.org/10.1016/ j.learninstruc.2005.07.006

Elfenbein, H. A. (2007). Emotion in organizations: A review and theoretical integration. *Academy of Management Annals,* 1, 371-457. http://dx.doi.org/10.1080/078559812

Enders, C.K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-38. http://dx.doi.org/10.1037/1082-989X.12.2.121

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France, and Sweden. *British Journal of Sociology, 30,* 341-415. http://dx.doi.org/10.1111/j.1468-4446.2009.01246.x

Eysenck, M. W. (1997). *Anxiety and cognition.* Hove, East Sussex, UK: Psychology Press.

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

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics – a “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22,* 497-514.

Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology, 101,* 705-716. http://dx.doi.org/10.1037/a0014695

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2017). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology.* Advance online publication. http://dx.doi.org/ 10.1037/edu0000228

Frenzel, A. C., Thrash, T. M., Pekrun, R., Goetz, T. (2007). Achievement emotions in Germany and China: A cross-cultural validation of the Academic Emotions Questionnaire-Mathematics (AEQ-M). *Journal of Cross-Cultural Psychology, 38,* 302-309. http://dx.doi.org/10.1177/0022022107300276

Fruehwirth, J. C. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, *4*, 85-124. http://dx.doi.org/ 10.3982/QE93

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science, 24,* 2079-2087. http://dx.doi.org/ 10.1177/0956797613486989

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99,* 715-733. https://dx.doi.org/10.1037/0022-0663.99.4.715

Goetz, T., Preckel, F., Zeidner, M., & Schleyer, E. (2008). Big fish in big ponds: A multilevel analysis of test anxiety and achievement in special gifted classes. *Anxiety, Stress and Coping, 21*, 185-198. http://dx.doi.org/abs/10.1080/10615800701628827

Gonzáles-Romá, V., & Hernández, A. (2017). Multilevel modeling: Research-based lessons for substantive researchers. *Annual Review of Organizational Psychology and Organizational Behavior, 4,* 183-210. http://dx.doi.org/10.1146/annurev-orgpsych-041015-062407

Hauser, R. M. (1974). Contextual analysis revisited. *Sociological Methods & Research, 2,* 365-75.

Harker, R., & Tymms, P. (2004). The effect of student composition on school outcomes. *School Effectiveness and Improvement, 15*, 177–199.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology, 75,* 645-654.

Heckhausen, H. (1991). *Motivation and action.* New York: Springer.

Hox, J. (2002). *Multilevel analysis techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the Big-Fish-Little-Pond Effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97,* 156-170. http://dx.doi.org/10.1037/a0015558

Jeličič, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology, 45*, 1195-1199. http://dx.doi.org/10.1037/a0015665

Jex, S. M., & Bliese, P. D. (1999). Efficacy beliefs as a moderator of the impact of work-related stressors: A multilevel study. *Journal of Applied Psychology, 84,* 349-361.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Kriwy, P., Gross, C., & Gottburgsen, A. (2013). Look who's talking: Compositional effects of gender and status on verbal contributions at sociology conferences. *Gender, Work and Organization, 20,* 545-560. http://dx.doi.org/10.1111/j.1468-0432.2012.00603.x

Lang, J. W. B., & Lang, J. (2010). Priming competence diminishes the link between cognitive test anxiety and test performance: Implications for the interpretation of test scores. *Psychological Science, 21,* 811-819. http://dx.doi.org/10.1177/0956797610369492

Larson, R., Hecker, B., & Norem, J. (1985). Students' experience with research projects: Pains, enjoyment and success. *High School Journal, 69,* 61-69.

Linnenbrink, E. A. (Ed.). (2006) Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition [Special issue]. *Educational Psychology Review, 18*(4). http://dx.doi.org/10.107/s10648-006-9028-x

Linnenbrink, E. A. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 107-124).

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203-229. http://dx.doi.org/ 10.1037/a0012869

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593-614. http://dx.doi.org/10.1146/annurev.psych.58.110405.085542

Maltby, J., Wood, A. M., Vlaev, I., Taylor, M. J., & Brown, G. D. A. (2012). Contextual effects on the perceived health benefits of exercise: The exercise rank hypothesis. *Journal of Sport and Exercise Psychology, 34,* 828-841. http://dx.doi.org/10.1123/jsep.34.6.828

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

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

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

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988), Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391-410.

Marsh, H. W., Hau, K. T., & Craven, R. (2004). The big-fish-little-pond effect stands up to scrutiny. *American Psychologist, 59,* 269-271. http://dx.doi.org/10.1037/0003-066X.59.4.269

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., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764-802.

Marsh, H. W., Parker, P.D., & Pekrun, R. (2018). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist.* Advance online publication. http://dx.doi.org/10.1027/ 1016-9040/a000332

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

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109,* 425-438. http://dx.doi.org/10.1037/edu0000144

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

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science, 5,* 546-56. http://dx.doi.org/ 10.1177/1745691610383511.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrolment intentions and performance in mathematics. *Journal of Educational Psychology, 82,* 60-70. http://dx.doi.org/10.1037/0022-0663.82.1.60

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*, 544-560. http://dx.doi: 10.1037/a0032459

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education, 82,* 143-167. http://dx.doi:10.1080/00220973.2013.769412

Murayama, K., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84,* 1475-1490. http://dx.doi.org/10.1111/cdev.12036

Murayama, K., Pekrun, R., Suzuki, M., Marsh, H. W., & Lichtenfeld, S. (2016). Don't aim too high for your kids: Parental over-aspiration undermines students' learning in mathematics. *Journal of Personality and Social Psychology, 111,* 166-179. http://dx.doi.org/10.1037/pspp0000079

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide.* Los Angeles, CA: Author.

Opdenakker, M.-C., & van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effects on mathematics achievement. *British Educational Research Journal, 27,* 407-432. http://dx.doi.org/10.1080/01411920120071434

Organization of Economic Cooperation and Development (OECD). (2005). *PISA 2003 data analysis manual: SPSS users.* Paris: OECD.

Parker, P. D., Jerrim, J., Schoon, I., Marsh, H. W. (2016). A multination study of socioeconomic inequality in expectations for progression to higher education. *American Educational Research Journal, 53,* 6-32. http://dx.doi.org/10.3102/0002831215621786

Pekrun, R. (1992). Expectancy-value theory of anxiety: Overview and implications. In D.G. Forgays, T. Sosnowski, & K. Wrzesniewski (Eds.), *Anxiety: Recent developments in self-appraisal, psychophysiological and health research* (pp. 23-41). Washington, DC: Hemisphere.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18,* 315-341. http://dx.doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2009). Global and local perspectives on human affect: Implications of the control-value theory of achievement emotions. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 97-115). Toronto, Canada: Hogrefe.

Pekrun, R. (2017). Achievement emotions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation (2nd edition): Theory and application* (pp. 251-271). New York: Guilford Press.

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. In G. A. D. Liem & D. M. McInerney (Eds.), *Big theories revisited 2* (pp. 165-190). Charlotte, NC: Information Age Publishing.

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction, 29,* 115-124. http://dx.doi.org/10.1016/j.learninstruc.2013.09.002

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101,* 115-135. http://dx.doi.org/10.1037/a0013383

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36,* 36-48. http://dx.doi.org/10.1016/ j.cedpsych.2010.10.002

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T. & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17-37). Münster, Germany: Waxmann.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development.* Published online first, http://dx.doi.org/10.1111/cdev12704

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotions in education*. New York: Taylor & Francis.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120-141). New York: Taylor & Francis.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J.. & Schiefele, U. (Eds.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* [PISA 2003. Analysis of competence development over one school year]. Münster, Germany: Waxmann.

Prenzel, M., Carstensen, C., Schöps, K., & Maurischat, C. (2006). Die Anlage des Längsschnitts bei PISA 2003 [The design of the PISA 2003 longitudinal study]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost., & U. Schiefele (Ed.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 29-62). Münster: Waxmann.

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

Raykov, T., & Penev, S. (1998). Nested structural equation models: Noncentrality and power of restriction test. *Structural Equation Modeling: A Multidisciplinary Journal, 5,* 229-246. http://dx.doi.org/10.1080/10705519809540103

Rose, R. A. (2018). Multilevel modelling in family violence research. *Journal of Family Violence, 33,* 109–122. http://dx.doi.org/10.1007/s10896-017-9938-z

Rousseau, D. M. (1986). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7,* 1-37.

Schnabel, K. (1998). *Prüfungsangst und schulisches Lernen* [Test anxiety and academic learning]. Münster: Waxmann.

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and moderation – two sides of the same coin. *American Educational Research Journal, 47,* 390-433. http://dx.doi.org/10.3102/0002831209350493

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

Stapel, D. A., & Blanton, H. (Eds.). (2007). *Social comparison theories.* New York: Psychology Press.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78–90. http://dx.doi.org/10.1086/209528

Steinmayr, R., Credel, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology, 6*: 1994. http://dx.doi.org/10.3389/fpsyg.2015.01994

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

Thrupp, M., Lauder, H. and Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*, 483-504. http://dx.doi.org/10.1016/S0883-0355(03)00016-8

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

Turner, J. E., & Schallert, D. L. (2001). Expectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology, 93,* 320-329. http://dx.doi.org/10.1037/0022-0663.93.2.320

Warr, P. (2007). *Work, happiness, and unhappiness.* Mahwah, NJ: Lawrence Erlbaum.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92,* 548-573.

Weiner, B. (2018). The legacy of an attribution approach to motivation and emotion: A no-crisis zone. *Motivation Science, 4,* 4-14. http://dx.doi.org/10.1037/mot0000082

Zeidner, M. (1998). Test anxiety: the state of the art. New York: Plenum.

Zeidner, M. (2014). Anxiety in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 265-288). New York: Taylor & Francis.

Zeidner, M., & Schleyer, E. J. (1998). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology, 24,* 305-329. http://dx.doi.org/10.1006/ceps.1998.0985

Zeidner, M., & Schleyer, E. J. (1999). Test anxiety in intellectually gifted school students. *Anxiety, Stress, and Coping, 12,* 163-189. http://dx.doi.org/10.1080/10615809908248328

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

Footnotes

[1] To denote the effects of group composition, we use the term compositional effect. Alternatively, these effects have been called contextual effects in some of the literature. Both terms are used to depict effects of group-level variables on individual outcomes. However, the term "contextual" is often used to describe effects of group-level variables other than group composition, such as size, budget, leadership styles, or governance structures. In contrast, as argued by Harker and Thymms (2004) and Rose (2018), "compositional" specifically refers to effects of the group-level composition of individual variables. As such, this term is best suited for the purposes of the present research.

[2] Findings based on the PALMA datasets have been published in Arens et al., 2017; Frenzel, Goetz, Lüdtke, Pekrun, & Sutton, 2009; Frenzel, Pekrun, & Goetz, 2007; Marsh et al., 2017, 2018; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Murayama, Pekrun, Suzuki, Marsh, & Lichtenfeld, 2016; and Pekrun et al., 2007, 2017. None of these publications has addressed compositional effects of achievement on emotions and self-concept.

[3] Research based on the PISA-I-Plus dataset has been published in Davier, Xu, and Carstensen, 2011, and Prenzel et al., 2006. None of these publications has addressed compositional effects of achievement on emotions and self-concept.

Table 1. *Descriptive Statistics and Pearson Product-Moment Correlations in Studies 1-3*

Variable	*M*	*SD*	Alpha	Enjoy-ment	Pride	Anger	Anxiety	Shame	Hope-lessness	Self-concept
					Study 1 [a]					
Enjoyment	2.875	.919	.90							
Anger	2.300	1.032	.87	.683						
Anxiety	2.442	.900	.87	.489		.663				
Hopelessness	2.175	1.017	.89	-.537		.706	.834			
Achievement [b]	.000	1.000		.088		−.157	−.239		−.220	
					Study 2					
Enjoyment [c]	3.312	.843	.87							
	2.974	.844	.87							
Pride	3.205	.939	.87	.728						
	2.994	.937	.88	.747						
Anger	2.023	.927	.87	−.555	−.350					
	2.209	.967	.88	−.576	−.436					
Anxiety	2.282	.823	.90	−.431	−.302	.748				
	2.305	.845	.90	−.423	−.348	.740				
Shame	1.892	.868	.86	−.281	−.181	.652	.781			
	1.904	.884	.88	−.257	−.212	.616	.776			
Hopelessness	2.032	.967	.86	−.429	−.345	.728	.831	.738		
	2.115	1.014	.89	−.462	−.413	.749	.860	.737		
Self-concept	3.453	.821	.88	. 656	.596	−.457	−.490	−.355	−.497	
	3.301	.895	.90	.661	.603	−.511	−.545	−.389	−.563	
Achievement [b]	.000	1.000		.122	.107	−.216	−.270	−.276	−.286	.345
	.000	1.000		.156	.126	−.241	−.265	−.289	−.250	.308
					Study 3					
Enjoyment	2.180	.771	.91							
	2.133	.755	.91							
Anger	2.020	.780	.83	−.567						
	1.998	.780	.83	−.571						
Anxiety	2.002	.725	.90	−.434		.732				
	2.001	.727	.90	−.443		.748				
Hopelessness	1.915	.817	.89	−.488		.783	.863			
	1.941	.822	.89	−.496		.795	.869			
Self-concept	2.523	.833	.92	.685		−.663	−.629		−.655	
	2.540	.864	.93	.711		−.666	−.636		−.659	
Achievement [b]	.000	1.000		.215		−.317	−.355		−.329	.391
	.000	1.000		.256		−.343	−.347		−.340	.382

Note. The coefficients are manifest correlations. For Studies 2 and 3, upper coefficients are for Time 1, lower coefficients are for Time 2.
[a] Study 1 correlations were averaged across grade levels (5–10) using Fisher's *z* transformation.
[b] Achievement scores were Rasch-scaled and z-standardized ($M = 0$, $SD = 1$) in all studies.
$p < .01$ for $|r| \geq .07$, .07, and .04 in Studies 1, 2, and 3, respectively.

Table 2

Multi-Level Modeling in Study 1

	Enjoyment		Anger		Anxiety		Hopelessness	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects								
Achievement	.182	.027	−.154	.028	−.254	.023	−.200	.024
Gender	.150	.033	−.069	.033	−.121	.029	−.187	.027
R^2	.066	.015	.032	.010	.091	.014	.089	.015
Class-level effects								
Achievement	−.401	.184	−.309	.220	−.557	.251	−.282	.244
Grade level	−.244	.193	.678	.227	.479	.267	.710	.243
R^2	.385	.112	.204	.116	.091	.077	.247	.134
Compositional effects								
Achievement	−.962	.162	.392	.153	.667	.152	.607	.131

Note. Coefficients for student-level and class-level effects are standardized path coefficients for latent-manifest models. Coefficients for compositional effects are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

Table 3

Multi-Level Modeling in Study 2 (Model 1)

	Enjoyment model				Pride model				Anger model			
	Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects												
Achievement t1	.115	.020	.615	.018	.095	.021	.619	.018	−.145	.018	.604	.019
Emotion t1	.510	.024	.054	.020	.468	.021	.039	.021	.445	.026	−.093	.021
R^2	.304	.024	.402	.023	.252	.021	.401	.023	.251	.024	.408	.023
Class−level effects												
Achievement t1	−.069	.098	.942	.019	−.182	.151	.935	.016	.064	.116	.919	.019
Emotion t1	.765	.072	.026	.051	.743	.125	.005	.066	.635	.117	−.075	.048
R^2	.617	.100	.876	.026	.614	.189	.876	.026	.881	.027	.389	.114
Compositional effects												
Achievement t1	−.251	.080	.771	.098	−.235	.074	.748	.097	.278	.081	.742	.091
	Anxiety model				Shame model				Hopelessness model			
Student-level effects												
Achievement t1	−.153	.023	.597	.019	−.114	.020	.600	.019	−.145	.023	.599	.019
Emotion t1	.498	.027	−.098	.020	.486	.029	−.093	.019	.434	.030	−.094	.020
R^2	.320	.024	.408	.023	.281	.027	.408	.023	.252	.024	.408	.023
Class-level effects												
Achievement t1	.113	.140	.898	.033	−.146	.129	.889	.029	.136	.142	.915	.023
Emotion t1	.742	.120	−.082	.060	.680	.111	−.109	.050	.648	.129	−.061	.060
R^2	.486	.134	.881	.027	.568	.132	.885	.027	.380	.143	.879	.027
Compositional effects												
Achievement t1	.331	.093	.715	.111	.130	.078	.691	.105	.322	.090	.745	.104

Note. Emo = emotion; Ach = achievement. Coefficients for student-level and class-level effects are standardized path coefficients for latent-manifest models. Coefficients for compositional effects are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, .01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

Table 4

Multi-Level Modeling in Study 2 (Model 2)

	Enjoyment model						Pride model						Anger model					
	Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects																		
Achievement t1	.093	.022	.205	.022	.567	.019	.071	.022	.205	.022	.566	.019	−.123	.020	.195	.021	.567	.019
Self−concept t1	.078	.030	.433	.035	.171	.027	.072	.032	.447	.023	.176	.024	−.071	.028	.448	.028	.121	.026
Emotion t1	.466	.028	.084	.030	−.043	.024	.432	.022	.068	.026	−.054	.023	.420	.029	−.092	.024	−.048	.026
R^2	.307	.025	.391	.023	.417	.023	.257	.021	.390	.023	.417	.023	.254	.024	.393	.023	.417	.023
Class-level effects																		
Achievement t1	−.243	.110	−.365	.142	.909	.034	−.407	.210	−.377	.178	.876	.052	.091	.120	−.247	.133	.916	.023
Self-concept t1	.427	.195	.988	.171	.081	.070	.714	.425	1.127	.335	.186	.119	−.257	.180	.844	.125	.022	.069
Emotion t1	.454	.173	−.263	.190	−.033	.057	.116	.411	−.396	.369	−.156	.138	.497	.157	.093	.159	−.063	.056
R^2	.699	.099	.638	.126	.879	.025	.689	.228	.652	.149	.689	.228	.432	.141	.608	.122	.881	.026
Compositional effects																		
Achievement t1	−.322	.094	−.578	.118	.788	.108	−.280	.092	−.585	.132	.726	.143	.274	.092	−.490	.105	.803	.094
	Anxiety model						Shame model						Hopelessness model					
Student-level effects																		
Achievement t1	−.132	.023	.188	.022	.564	.019	−.095	.025	.188	.022	.561	.019	−.116	.024	.191	.022	.565	.019
Self-concept t1	−.074	.029	.443	.025	.119	.026	−.058	.029	.464	.027	.123	.023	−.099	.030	.449	.026	.121	.024
Emotion t1	.469	.028	−.103	.022	−.051	.024	.470	.031	−.081	.023	−.060	.021	. 396	.031	−.090	.025	−.047	.023
R^2	.324	.025	.395	.023	.418	.023	.282	.027	.392	.023	.419	.023	.258	.024	.393	.024	.417	.023
Class-level effects																		
Achievement t1	.109	.142	−.287	.153	.898	.033	−.158	.137	−.268	.144	.033	.063	.142	.145	−.257	.134	.914	.025
Self-concept t1	−.234	.190	.757	.110	.025	.070	.109	.184	.789	.090	.885	.032	−.112	.222	.807	.127	.033	.084
Emotion t1	.617	.147	−.079	.180	−.069	.070	.712	.112	−.027	.055	−.101	.053	.586	.185	.016	.173	−.040	.085
R^2	.520	.147	.598	.122	.881	.026	.586	.143	.602	.119	.886	.026	.389	.143	.607	.119	.880	.026
Compositional effects																		
Achievement t1	.292	.094	−.501	.117	.773	.113	.090	.082	−.491	.111	.753	.107	.272	.090	−.489	.108	.804	.105

Note. Emo = emotion; Self = self-concept; Ach = achievement. Coefficients for student-level and class-level effects are standardized path coefficients for latent-manifest models. Coefficients for compositional effects are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

Table 5
Multi-Level Modeling in Study 3 (Model 1)

	Enjoyment model				Anger model				Anxiety model				Hopelessness model			
	Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	SE	Path	*SE*
Student-level effects																
Achievement t1	.118	.014	.843	.007	−.159	.015	.849	.008	−.160	.013	851	.008	−.163	.012	.852	.009
Emotion t1	.652	.011	.056	.009	.556	.015	−.033	.012	.598	.013	−.026	.013	.570	.014	−.022	.014
R^2	.491	.013	.733	.010	.403	.016	.731	.010	.483	.015	.731	.010	.449	.015	.731	.010
Class-level effects																
Achievement t1	.102	.069	.930	.011	−.260	.101	.936	.019	−.209	.092	.958	.021	−.221	.117	.943	.020
Emotion t1	.739	.056	.005	.031	.650	.097	.013	.039	.632	.096	.073	.043	.587	.112	.033	.045
R^2	.562	.081	.866	.020	.615	.079	.866	.021	.542	.094	.871	.021	.490	.085	.867	.021
Compositional effects																
Achievement t1	−.205	.047	.030	.068	.202	.063	.027	.076	.268	.045	.070	.079	.273	.059	.035	.080

Note. Emo = emotion; Ach = achievement. Coefficients for student-level and class-level effects are standardized path coefficients for latent-manifest models. Coefficients for compositional effects are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

Table 6

Multi-Level Modeling in Study 3 (Model 2)

	Enjoyment model						Anger model					
	Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects												
Achievement t1	.059	.015	.141	.013	.825	.012	−.101	.017	.132	.013	.824	.012
Self−concept t1	.177	.023	.604	.021	.055	.020	−.169	.022	.654	.016	.075	.015
Emotion t1	.551	.019	.122	.019	.025	.012	.474	.018	−.067	.015	.003	.011
R^2	.504	.013	.624	.013	.734	.010	.417	.016	.616	.013	.734	.010
Class-level effects												
Achievement t1	.104	.072	.019	.078	.934	.013	−.301	.122	.104	.104	.925	.020
Self-concept t1	−.009	.140	.709	.131	−.025	.051	−.230	.255	1.067	.235	.077	.075
Emotion t1	.746	.127	.088	.135	.023	.036	.437	.297	.350	.267	.093	.070
R^2	.563	.080	.613	.073	.866	.020	.632	.069	.630	.086	.867	.021
Compositional effects												
Achievement t1	−.074	.050	−.304	.042	.076	.075	.051	.077	−.249	.050	.117	.081
	Anxiety model						Hopelessness model					
Student-level effects												
Achievement t1	−.110	.016	.129	.013	.825	.012	−.111	.014	.130	.013	.825	.012
Self-concept t1	−.155	.020	.654	.015	.079	.017	−.159	.020	.660	.016	.083	.015
Emotion t1	.524	.017	−.071	.013	.012	.013	.493	.017	−.060	.015	.019	.014
R^2	.494	.015	.616	.014	.734	.010	.461	.015	.615	.013	.734	.010
Class-level effects												
Achievement t1	−.305	.107	.140	.100	.962	.020	−.389	.147	.136	.119	.960	.024
Self-concept t1	−.522	.217	1.119	.160	.085	.056	−.922	.346	1.166	.275	.099	.083
Emotion t1	.165	.237	.453	.191	.122	.053	−.277	.397	.467	.317	.125	.085
R^2	.626	.099	.674	.083	.872	.021	.677	.103	.654	.091	.869	.021
Compositional effects												
Achievement t1	.118	.054	−.225	.049	.133	.080	.069	.071	−.231	.055	.130	.085

Note. Emo = emotion; Self = self-concept; Ach = achievement. Coefficients for student-level and class-level effects are standardized path coefficients for latent-manifest models. Coefficients for compositional effects are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

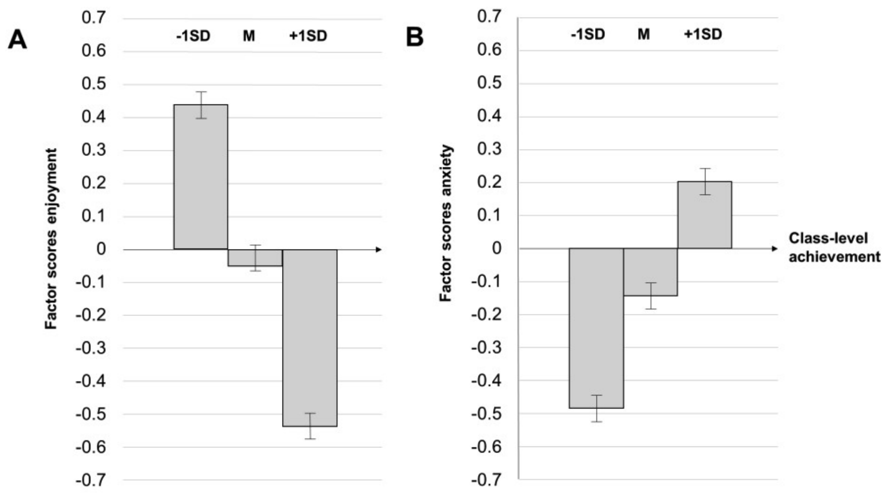

Figure 1. Relations of class-level achievement to enjoyment (**A**) and anxiety (**B**) in Study 1. The bars depict enjoyment and anxiety scores at –1SD, mean, and +1SD class-level achievement, showing that class-level achievement related negatively to enjoyment and positively to anxiety. The error bars indicate 95% confidence intervals. For both enjoyment and anxiety, all pairwise comparisons between the estimates at –1SD, mean, and +1SD class-level achievement were significant at $p < 0.000$.

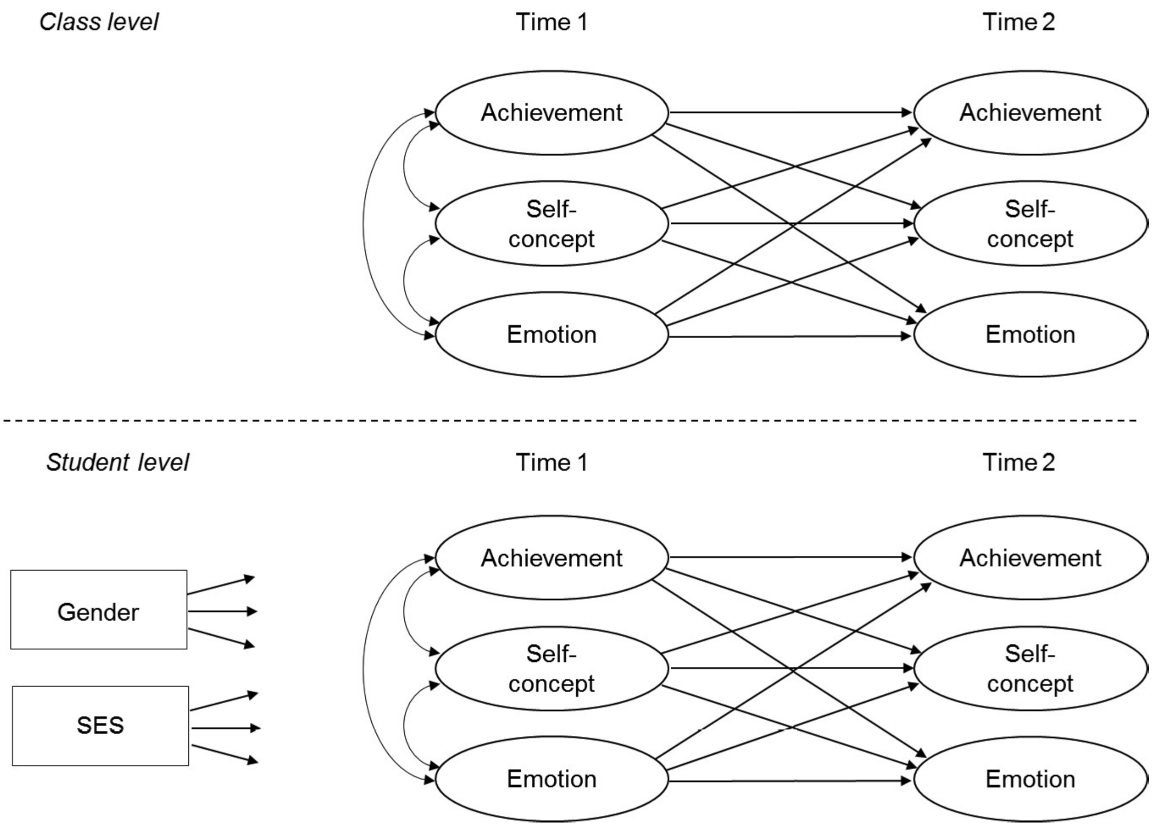

Figure 2. The links between achievement, self-concept, and emotion analyzed in longitudinal multi-level modelling in Studies 2 and 3. The models included effects of covariates on the student-level variables at both Time 1 and Time 2. Correlations between residuals are not depicted. Estimates for compositional effects are derived by subtracting the student-level effects from the class-level effects.

Supplemental Materials for

Happy Fish in Little Ponds:
Testing a Reference Group Model of Achievement and Emotion

This file includes:

Equations for Latent-Manifest Models

In latent-manifest modeling, student-level achievement scores were group-mean centered (Enders & Tofighi, 2007) and were included in the model as a fixed effect predictor. The mathematical expression of the model is as follows.

Level 1 (student level): $$\boldsymbol{Y}_{ij} = \beta_{0j} + \beta_{1j}(\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_{.j}) + \boldsymbol{r}_{ij} \quad (1)$$

Level 2 (class level): $$\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{\boldsymbol{X}}_{.j} + \boldsymbol{u}_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

In the Level 1 equation (1), $\boldsymbol{Y}_{ij}$ is the outcome variable (i.e., achievement emotion) for student *i* in classroom *j* predicted by the intercept β_{0j} of classroom *j* and the regression slope β_{1j} in classroom *j*. The student-level achievement score $\boldsymbol{X}_{ij}$ is centered at the respective classroom mean $\overline{\boldsymbol{X}}_{.j}$. In the level 2 equations (2) and (3), γ_{00} and γ_{10} represent the population means of β_{0j} and β_{1j}, and γ_{01} is the regression coefficient relating class-level achievement scores $\overline{\boldsymbol{X}}_{.j}$ to the intercepts from the level 1 equation. $\boldsymbol{u}_{0j}$ is the class-level residual term. These equations can be combined into one equation as follows.

$$\boldsymbol{Y}_{ij} = \gamma_{00} + \gamma_{01}\overline{\boldsymbol{X}}_{.j} + \gamma_{10}(\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_{.j}) + \boldsymbol{u}_{0j} + \boldsymbol{r}_{ij} \quad (4)$$

With this model, a compositional effect is presumed to be present if $\hat{\gamma}_{01}$ is significantly different from $\hat{\gamma}_{10}$. Therefore, we computed the compositional effect as follows (Kreft & De Leeuw, 1998; Lüdtke et al., 2008; Marsh et al., 2009):

$$\text{Estimated compositional effect} = \hat{\gamma}_{01} - \hat{\gamma}_{10} \quad (5)$$

Our study hypotheses imply that the compositional effect of the class-level achievement score on the outcome variable is opposite to the effect of the student-level achievement score. For positive emotions, we expected the student-level effect $\hat{\gamma}_{10}$ to be positive; the compositional effect of achievement on emotion was expected to be negative. For negative emotions, we expected $\hat{\gamma}_{10}$ to be negative; the compositional effect of achievement on emotion was expected to be positive. The standard error of the compositional effect was computed using the delta method.

Supplemental Analyses for Study 2

To further check the robustness of findings, we first investigated the compositional effects of class-average on emotion in the remaining waves of the PALMA longitudinal study in separate analyses for grades 7, 8, and 9 (see main text). Second, we performed additional analyses controlling for student-perceived teacher enthusiasm and pressure for achievement in the grade 5-6 longitudinal analysis and each of the grades 7, 8, and 9 cross-sectional analyses.

Analyses for Grades 7-9

Method. Samples included the grades 7, 8, and 9 participants of the PALMA longitudinal study (grades 7-9: *N*s = 2,397, 2,410, and 2,528 students, respectively, 50.1%, 50.5%, and 51.1% female; see Pekrun et al., 2007). We used the same measures for achievement and emotion as in the longitudinal grade 5-6 analysis (see main text). For achievement, the IRT scores from the PALMA mathematical achievement test were used. The emotions were measured using the same Achievement Emotions Questionnaire-Mathematics as in the longitudinal analysis. Alpha reliabilities for the enjoyment, pride, anger, anxiety, shame, and hopelessness scales were .88, .88, .87. .91, .87, and .90, respectively, in grade 7; .85, .89, .87, .91, .82, and .90 in grade 8; and .89, .89, .88, .92, .89, and .90 in grade 9. Factor scores for the emotions were derived from separate confirmatory factor analyses for the three waves. As in the longitudinal grade 5-6 analysis, we constructed separate multi-level models for the six emotions, with achievement as the predictor and emotion as the outcome variable. Gender and SES were controlled in the analysis.

Results. The findings replicate the results of the longitudinal grade 5-6 analysis (Table S9). Again, individual achievement related positively to students' math enjoyment and pride, and negatively to their math-related anger, anxiety, shame, and hopelessness. In contrast, there were negative compositional effects of class-average achievement on enjoyment and pride, and positive compositional effects on anger, anxiety, shame, and hopelessness. The effects were stronger than in the longitudinal analysis, likely due to the cross-sectional design of the analysis involving an assessment of both achievement and emotion within the same school year. In contrast to the non-significant compositional effect on shame in the longitudinal analysis, the compositional effects on shame were significant for each of the grade 7, 8, and 9 analyses, which likely is also due to the nature of the design. In sum, the findings attest to the robustness of the student-level and compositional effects of achievement on emotion across each of the remaining waves of the longitudinal study.

Analyses Controlling for Teacher Behavior

Method. To control for teacher behavior, we replicated the grade 5-6 longitudinal analysis (Model 1 including achievement and emotion) and the grades 7, 8, and 9 cross-sectional analyses (see above) while including teachers' student-perceived teacher enthusiasm and pressure for achievement as predictors. For the grade 5-6 analysis, teacher behavior at grade 5 was entered as a predictor, in line with the longitudinal logic of keeping predictor and outcome variables temporally separate. In each of the four analyses, enthusiasm was measured using the PALMA Teacher Engagement and Enthusiasm scale (5 items; e.g., "Our math teacher is enthusiastic about teaching;" Cronbach's α = .84, .85, .87, and .87 at grade 5, 7, 8, and 9, respectively). Pressure for achievement was measured with the PALMA Teacher Pressure for Achievement scale (5 items; e.g., "My teacher expects better grades from me than I can possibly achieve;" Cronbach's α = .72, .75, .68, and .69 at grade 5, 7, 8, and 9, respectively). In multi-level modeling, the two variables were treated in the same way as the achievement predictors, that is, they were group-mean centered and entered in the models at both the student and class levels (at the class level, they represent students' shared perception of teacher behavior; at the student level, they represent students' individual deviation from shared perception).

Results. The results indicate that the study findings were robust when controlling for teachers' enthusiasm and pressure for achievement (for a summary of compositional effects, see Table S10). Specifically, class-average achievement again had significantly negative compositional effects on enjoyment and pride across the grade 5-6 and the grade 7, 8, and 9 analyses. Across these analyses, the compositional effects were significantly negative for anger, anxiety, shame, and hopelessness, except for the effect on shame in the longitudinal grade 5-6 analysis. As in the main analysis, this effect was negative but not significant. Overall, these findings suggest that the study results on the compositional effects of class-average achievement were not confounded with differences in teachers' emotion-relevant behavior across classes.

References

Browne, M. W., & Du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research, 27,* 269-300.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121-38. doi: 10.1037/1082-989X.12.2.121

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203-229. doi: 10.1037/a0012869

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764-802.

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T. & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17-37). Münster, Germany: Waxmann.

Table S1. *Factor Loadings, Factor Determinacy Scores, and Pearson Product-Moment Correlations of Factor Score Estimates in Studies 1-3*

			Correlations					
	Factor loadings	Factor determinacy [a]	Enjoyment	Pride	Anger	Anxiety	Shame	Hopelessness
Enjoyment	.460–.817	.933						
Anger	.626–.804	.931	−.688					
Anxiety	.485–.748	.934	−.477		.638			
Hopelessness	.698–.870	.953	−.529		.681	.766		
Enjoyment [c]	.368–.814	.930						
	.352–.810	.930						
Pride	.572–.789	.922	.702					
	.572–.778	.916	.715					
Anger	.653–.762	.943	−.561	−.359				
	.647–.790	.938	−.583	−.457				
Anxiety	.415–.725	.933	−.430	−.312	.756			
	.410–.747	.933	−.422	−.366	.754			
Shame	.468–.766	.940	−.276	−.209	.640	.784		
	.555–.730	.924	−.238	−.233	.608	.788		
Hopelessness	.600–.835	.943	−.422	−.363	.730	.817	.753	
	.659–.859	.951	−.440	−.424	.747	.849	.745	
Self–concept	.654–.813	.941	.644	.610	−.456	−.496	−.368	−.499
	.669–.813	.949	.649	.617	−.513	−.539	.745	−.554
Enjoyment	.705–.883	.956						
	.735–.897	.956						
Anger	.568–.783	.908	−.558					
	.576–.803	.907	−.555					
Anxiety	.522–.750	.908	−.440		.719			
	.586–.754	.907	−.467		.745			
Hopelessness	.695–.882	.950	−.490		.771	.843		
	.748–.896	.954	−.501		.774	.842		
Self–concept	.790–.876	.955	.689		−.616	−.633		−.640
	.778–.872	.962	.718		−.614	−.644		−.643

Note. For Studies 2 and 3, upper coefficients are for Time 1, lower coefficients are for Time 2.
[a] Factor determinacy scores are the correlations of factor score estimates with their respective factors.

Table S2. *Measurement Equivalence of Emotion and Self-Concept Constructs Across Waves in Study 2*

Type of invariance	CFI	RMSEA	SRMR	AIC	BIC_{adj}	RDR
			Enjoyment			
Configural	.974	.041	.032	85609.210	85797.448	---
Metric	.972	.041	.037	85630.664	85800.537	.043
Scalar	.969	.042	.040	85660.331	85811.840	.044
Residual	.967	.042	.057	85698.218	85829.066	.046
			Pride			
Configural	.979	.041	.025	70598.319	70736.054	
Metric	.978	.040	.028	70595.565	70719.527	.022
Scalar	.977	.040	.029	70604.788	70719.567	.037
Residual	.977	.038	.034	70599.695	70698.405	.010
			Anger			
Configural	.972	.042	.027	80546.973	80714.551	
Metric	.971	.041	.029	80547.233	80698.742	.020
Scalar	.969	.041	.030	80558.022	80693.461	.036
Residual	.968	.040	.032	80570.959	80688.034	.028
			Anxiety			
Configural	.964	.033	.036	151683.238	152078.079	
Metric	.963	.033	.037	151682.130	152044.832	.020
Scalar	.962	.032	.038	151699.047	152029.611	.031
Residual	.960	.032	.039	151718.275	152014.406	.031
			Shame			
Configural	.991	.022	.022	77722.385	77889.962	
Metric	.990	.021	.024	77722.403	77873.911	.011
Scalar	.990	.021	.024	77717.710	77853.150	.013
Residual	.990	.020	.026	77713.100	77830.175	.000
			Hopelessness			
Configural	.995	.020	.017	59339.891	59438.601	
Metric	.994	.020	.020	59340.347	59427.579	.016
Scalar	.993	.021	.020	59342.548	59418.302	.026
Residual	.991	.023	.026	59365.661	59427.642	.038
			Self-concept			
Configural	.996	.021	.015	50949.410	51048.120	
Metric	.992	.026	.034	50979.502	51066.734	.055
Scalar	.989	.030	.037	51009.968	51085.722	.055
Residual	.989	.028	.043	51006.210	51068.191	.000

Note. BIC_{adj} = sample-size adjusted BIC. RDR = root deterioration per restriction index (Browne & Du Toit, 1992). RDR is used to compare more restrictive nested models (e.g., metric invariance) with less restrictive models (e.g., configural invariance).

Table S3. *Measurement Equivalence of Emotion and Self-Concept Constructs Across Waves in Study 3*

Type of invariance	CFI	RMSEA	SRMR	AIC	BIC_{adj}	RDR
			Enjoyment			
Configural	.995	.029	.012	96624.784	96781.386	
Metric	.993	.030	.018	96652.934	96793.556	.038
Scalar	.993	.030	.018	96665.611	96790.253	.026
Residual	.993	.028	.019	96670.040	96775.506	.009
			Anger			
Configural	.983	.045	.026	99976.024	100103.862	
Metric	.982	.043	.029	99991.454	100106.508	.033
Scalar	.979	.044	.029	100039.557	100141.828	.048
Residual	.977	.042	.032	100057.237	100143.528	.027
			Anxiety			
Configural	.975	.045	.033	163421.141	163721.562	
Metric	.974	.044	.035	163442.466	163717.318	.030
Scalar	.972	.043	.034	163484.633	163733.918	.035
Residual	.972	.042	.036	163504.368	163724.889	.018
			Hopelessness			
Configural	.983	.048	.022	90173.241	90288.288	
Metric	.982	.045	.022	90168.240	90270.503	.017
Scalar	.982	.043	.022	90162.952	90252.432	.019
Residual	.982	.040	.023	90172.362	90245.864	.000
			Self-concept			
Configural	.988	.051	.018	80672.479	80794.030	
Metric	.987	.050	.023	80694.766	80803.523	.041
Scalar	.987	.048	.023	80704.807	80800.769	.021
Residual	.985	.047	.025	80752.293	80832.261	.039

Note. BIC_{adj} = sample-size adjusted BIC. RDR = root deterioration per restriction index (Browne & Du Toit, 1992). RDR is used to compare more restrictive nested models (e.g., metric invariance) with less restrictive models (e.g., configural invariance).

Table S4 *Multi-Level Modeling in Study 1: Unstandardized Coefficients*

	Enjoyment		Anger		Anxiety		Hopelessness	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects								
Achievement	.334	.048	−.263	.046	−.449	.039	−.373	.044
Gender	.265	.061	−.114	.054	−.206	.050	−.337	.049
Class-level effects								
Achievement	−.151	.071	−.079	.056	−.126	.063	−.063	.053
Grade level	−.049	.039	.092	.032	.057	.035	.084	.030
Compositional effects								
Achievement	−.485	.082	.184	.072	.323	.073	.310	.067

Note. Coefficients are unstandardized path coefficients for latent-manifest models. *SE* = standard error. *p* < .05, 01, and .001 for coefficients higher than 1.96, 2.58, and 3.29 *SE* , respectively.

Table S5 *Multi-Level Modeling in Study 2 (Model 1): Unstandardized Coefficients*

	Enjoyment model				Pride model				Anger model			
	Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects												
Achievement t1	.132	.024	.582	.020	.108	.025	.586	.020	−.154	.020	.572	.021
Emotion t1	.524	.027	.046	.017	.482	.024	.034	.018	.448	.029	−.083	.019
Class-level effects												
Achievement t1	−.031	.045	.994	.043	−.043	.036	.987	.042	.024	.044	.969	.039
Emotion t1	.770	.090	.060	.120	.688	.150	−.019	.270	.701	.136	−.229	.157
Compositional effects												
Achievement t1	−.163	.050	.412	.045	−.151	.046	.401	.045	.178	.051	.397	.043
	Anxiety model				Shame model				Hopelessness model			
Student-level effects												
Achievement t1	−.161	.025	.565	.021	−.132	.025	.568	.021	−.167	.029	.567	.021
Emotion t1	.512	.028	−.090	.019	.480	.031	−.076	.016	.447	.032	−.080	.017
Class-level effects												
Achievement t1	.036	.046	.947	.050	−.047	.041	.938	.044	.042	.045	.966	.046
Emotion t1	.901	.216	−.323	.247	.703	.175	−.367	.164	.668	.192	−.214	.203
Compositional effects												
Achievement t1	.198	.055	.382	.055	.084	.051	.370	.049	.209	.058	.399	.050

Note. Coefficients are unstandardized path coefficients for latent-manifest models. Emo = emotion; ach = achievement. *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S6 *Multi-Level Modeling in Study 2 (Model 2): Unstandardized Coefficients*

	Enjoyment model						Pride model						Anger model					
	Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects																		
Achievement t1	.107	.026	.240	.027	.537	.021	.081	.026	.241	.027	.535	.021	−.130	.022	.228	.027	.537	.020
Self-concept t1	.079	. 030	.447	.036	.143	.020	.072	.032	.461	.034	.147	.021	−.066	.026	.463	.028	.101	.022
Emotion t1	.478	.031	.088	.032	−.036	.023	.445	.027	.072	.028	−.047	.020	.423	.031	−.102	.026	−.043	.023
Class-level effects																		
Achievement t1	−.109	.049	−.143	.063	.959	.046	−.099	.049	−.148	.075	.924	.063	.034	.046	−.097	.055	.966	.040
Self-concept t1	.615	.272	1.239	.293	.272	.234	.558	.329	1.415	.490	.629	.392	−.312	.218	1.062	.245	.074	.234
Emotion t1	.453	.181	−.229	.173	−.077	.132	.111	.395	−.610	.584	−.645	.557	.547	.179	.106	.187	−.194	.175
Compositional effects																		
Achievement t1	−.216	.058	−.383	.074.	.422	.049	−.180	.058	−.388	.083	.389	.069	.164	.054	−.325	.065	.430	.044
	Anxiety model						Shame model						Hopelessness model					
Student-level effects																		
Achievement t1	−.140	.025	.220	.028	.533	.020	−.110	.025	.220	.027	.530	.021	−.134	.029	.223	.028	.534	.021
Self-concept t1	−.069	.027	.458	.026	.099	.022	−.059	.029	.479	.028	.103	.019	−.101	.030	.464	.027	.101	.021
Emotion t1	.482	.030	−.118	.025	−.047	.023	.464	.032	−.082	.023	−.049	.017	.407	.033	−.094	.026	−.040	.019
Class-level effects																		
Achievement t1	.035	.046	−.112	.063	.947	.051	−.051	.043	−.105	.060	.933	.045	.043	.046	−.101	.056	.965	.046
Self-concept t1	−.241	.190	.948	.207	.083	.237	.113	.189	.988	.191	.112	.215	−.110	.218	1.014	.234	.112	.286
Emotion t1	.746	.236	−.116	.266	−.275	.281	.736	.169	−.034	.194	−.340	.173	.601	.229	.022	.227	−.143	.295
Compositional effects																		
Achievement t1	.175	.055	−.332	.074	.414	.055	.058	.054	−.326	.071	.403	.050	.177	.058	−.324	.068	.430	.050

Note. Coefficients are unstandardized path coefficients for latent-manifest models. Emo = emotion; self = self-concept; ach = achievement. *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S7 *Multi-Level Modeling in Study 3 (Model 1): Unstandardized Coefficients*

	Enjoyment model				Anger model				Anxiety model				Hopelessness model			
	Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2		Emo t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	SE	Path	*SE*
Student-level effects																
Achievement t1	.159	.018	.884	.012	−.202	.019	.891	.012	−.208	.017	.892	.013	−.221	.017	.894	.013
Emotion t1	.662	.013	.045	.007	.555	.017	−.027	.010	.600	.015	−.021	.010	.572	.016	−.017	.011
Class-level effects																
Achievement t1	. 034	.023	.898	.029	−.085	.031	.904	.033	−.051	.021	.925	.034	−.061	.032	.911	.035
Emotion t1	.615	.063	.012	.074	.679	.120	.041	.119	.615	.106	.284	.171	.627	.128	.125	.169
Compositional effects																
Achievement t1	−.125	.029	.014	.032	.117	.037	.013	.036	.157	.027	.033	.037	.160	.036	.035	.080

Note. Coefficients are unstandardized path coefficients for latent-manifest models. Emo = emotion; ach = achievement. *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S8 *Multi-Level Modeling in Study 3 (Model 2): Unstandardized Coefficients*

	Enjoyment model						Anger model					
	Emo t2		Self t2		Ach t2		Emo t2		Self t2		Ach t2	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
Student-level effects												
Achievement t1	.079	.020	.195	.018	.865	.016	−.129	.021	.181	.018	.864	.016
Self-concept t1	.175	.023	.612	.022	.042	.015	−.159	.021	.661	.018	.058	.012
Emotion t1	.559	.020	.127	.020	.020	.009	.474	.019	−.072	.016	.003	.009
Class-level effects												
Achievement t1	.034	.024	.005	.020	.902	.031	−.099	.038	.027	.026	.919	.034
Self-concept t1	−.011	.177	.693	.142	−.091	.186	−.290	.324	1.043	.245	.285	.277
Emotion t1	.624	.117	.056	.087	.056	.087	.457	.317	.284	.217	.284	.217
Compositional effects												
Achievement t1	−.045	.030	−.190	.026	.036	.035	.030	.045	−.155	.031	.055	.038
	Anxiety model						Hopelessness model					
Student-level effects												
Achievement t1	−.143	.020	.176	.018	.865	.016	−.150	.019	.178	.018	.866	.016
Self-concept t1	−.148	.019	.661	.017	.061	.013	−.159	.020	.666	.018	.064	.012
Emotion t1	.525	.018	−.075	.014	.010	.011	.494	.018	−.060	.015	.015	.011
Class-level effects												
Achievement t1	−.073	.026	.036	.026	.928	.033	−.108	.040	.035	.030	.927	.036
Self-concept t1	−.476	.213	1.108	.186	.313	.203	−.974	.373	1.148	.281	.365	.304
Emotion t1	.158	.227	.473	.207	.473	.207	−.297	.427	.466	.321	.466	.321
Compositional effects												
Achievement t1	.069	.032	−.140	.031	.063	.037	.043	.044	−.143	.034	.062	.040

Note. Coefficients are unstandardized path coefficients for latent-manifest models. Emo = emotion; self = self-concept; ach = achievement. *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S9. *Compositional Effects of Achievement: Estimates from Doubly-Latent Models*

Outcome variable	Enjoyment models		Pride models		Anger models		Anxiety models		Shame models		Hopelessness models	
	1	2	1	2	1	2	1	2	1	2	1	2
	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*	CE *SE*
Study 1												
Emotion	−.527 .085				.198 .077		.336 .081				.334 .071	
	−1.032 .167				.416 .163		.685 .164				.644 .137	
Study 2												
Emotion	−.166 .056	−.206 .060	−.161 .051	−.179 .058	.186 .056	.161 .058	.201 .057	.172 .057	.083 .053	.059 .053	.218 .059	.184 .059
	−.235 .083	−.304 .095	−.232 .076	−.257 .086	.285 .087	.247 .092	.308 .087	. 264 .088	.118 .076	.088 .079	.308 .085.	.259 .085
Self−concept		−.384 .075		−.386 .083		−.327 .069		−.335 .076		−.329 .073		−.326 .071
		−.532 .111		−.535 .122		−.467 .106		−.463 .111		−.455 .107		−.451 .104
Study 3												
Emotion	−.126 .030	−.045 .031			.115 .037	.005 .059	.161 .028	.071 .033			.162 .036	.027 .050
	−.201 .048	−.072 .048			.193..061	.009 .099	.266 .044	.117 .054			.255 .054	.043 .079
Self−concept		−.197 .027				−.161 .041		−.146 .032				−.151 .038
		−.304 .040				−.249 .062		−.227 .049				−.234 .057

Note. Coefficients CE are compositional effects of achievement on emotion and self-concept. Upper coefficients are unstandardized coefficients, lower coefficients are standardized effect sizes (ES 2; Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S10. *Relations of Gender with Time 1 Variables in Studies 2 and 3*

	Emo t1		Self t1		Ach t1		Emo t1		Self t1		Ach t1		Emo t1		Self t1		Ach t1	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
									Study 2 – Model 1									
			Enjoyment model						Pride model						Anger model			
Gender	.142	.024			.173	.029	.161	.025			.173	.029	−.008	.030			.173	.029
SES	.020	.025			−.060	.025	.037	.024			−.060	.025	.033	.026			−.060	.025
			Anxiety Model						Shame model						Hopelessness model			
Gender	−.098	.028			.173	.029	−.075	.029			.173	.029	−.136	.028			.173	.029
SES	.047	.023			−.060	.025	.034	.026			−.060	.025	.036	.024			−.060	.025
									Study 2 – Model 2									
			Enjoyment model						Pride model						Anger model			
Gender	.141	.024	.239	.025	.173	.029	.162	.024	.239	.026	.172	.029	−.007	.030	.240	.026	.173	.029
SES	.019	.025	−.018	.021	−.060	.025	.039	.024	−.017	.021	−.060	.025	.033	.026	−.015	.021	−.060	.025
			Anxiety Model						Shame model						Hopelessness model			
Gender	−.098	.028	.240	.025	.173	.029	−.075	.029	.239	.025	.173	.029	−.136	.028	.240	.026	.173	.029
SES	.048	.022	−.016	.022	−.060	.025	.034	.026	−.017	.021	−.060	.025	.037	.024	−.015	.021	−.060	.025
									Study 3 – Model 1									
			Enjoyment model												Anger model			
Gender	.147	.020			.255	.018							−.116	.019			.255	.018
SES	−.030	.048			−.025	.013							.076	.044			−.025	.013
			Anxiety model												Hopelessness model			
Gender	−.213	.016			.255	.018							−.203	.016			.255	.018
SES	.037	.027			−.025	.013							.019	.033			−.025	.013
									Study 3 – Model 2									
			Enjoyment model												Anger model			
Gender	.147	.020	.242	.017	.255	.018							−.114	.019			.255	.018
SES	−.030	.048	−.056	.061	−.025	.013							.078	.043			−.025	.013
			Anxiety model												Hopelessness model			
Gender	−.212	.016	.243	.016	.255	.018							−.201	.016	.244	.016	.255	.018
SES	.038	.027	.042	.064	−.025	.013							.023	.032	.027	.064	−.025	.013

Note. Coefficients are standardized path coefficients. Emo = emotion; self = self-concept; ach = achievement. *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S11. *Supplemental Analyses: Multi-Level Models for Grades 7, 8, and 9 in Study 2*

	Enjoyment		Pride		Anger		Anxiety		Shame		Hopelessness	
Predictor	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*	Path	*SE*
							Grade 7					
Student level												
Achievement	.248	.019	.201	.021	−.289	.022	−.320	.020	−.248	.022	−.297	.020
Gender	.086	.027	.099	.025	−.017	.025	−.035	.022	.030	.020	−.097	.023
SES	−.059	.020	−.061	.021	−.029	.019	−.006	.020	−.036	.021	−.002	.019
Class level												
Achievement	−.114	.127	−.158	.162	−.339	.107	−.623	.105	−.742	.084	−.451	.128
Compositional effect												
Achievement	−.711	.096	−.587	.089	.573	.095	.521	.086	.268	.093	.572	.094
							Grade 8					
Student level												
Achievement	.297	.020	.251	.020	−.291	.020	−.324	.018	−.249	.018	−.304	.017
Gender	.095	.025	.068	.022	−.004	.023	−.013	.021	.053	.022	−.073	.019
SES	−.022	.024	−.033	.023	−.021	.020	−.019	.022	−.034	.024	−.015	.023
Class level												
Achievement	−.004	.127	.028	.155	−.488	.118	−.803	.069	−.788	.075	−.702	.097
Compositional effect												
Achievement	−.865	.097	−.719	.095	.668	.080	.571	.076	.296	.085	.596	.075
							Grade 9					
Student level												
Achievement	.306	.019	.250	.021	−.302	.018	−.291	.019	−.211	.019	−.292	.018
Gender	.050	.026	.081	.023	.010	.022	−.034	.020	.014	.021	−.104	.021
SES	−.006	.025	−.036	.024	.015	.025	−.015	.025	−.005	.025	−.000	.024
Class level												
Achievement	−.087	.133	−.070	.148	−.556	.111	−.582	.105	−.638	.097	−.467	.125
Compositional effect												
Achievement	−.951	.098	−.770	.097	.624	.091	.568	.092	.263	.093	.659	.091

Note. Coefficients are standardized path coefficients for latent-manifest models. For compositional effects, coefficients are standardized effect sizes (ES 2, Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.

Table S12. *Supplemental Analyses: Compositional Effects of Achievement on Emotion in Multi-Level Models Controlling for Teacher Behavior in Study 2*

	Enjoyment		Pride		Anger		Anxiety		Shame		Hopelessness	
Model	ES2	*SE*	ES2	*SE*	ES2	*SE*	ES2	*SE*	ES2	*SE*	ES2	*SE*
Longitudinal												
Grade 5-6	−.280	.106	−.240	.091	.327	.097	.297	.086	.123	.087	.333	.090
Cross−sectional												
Grade 7	−.549	.079	−.449	.083	.428	.084	.434	.084	.201	.091	.512	.095
Grade 8	−.670	.078	−.518	.085	.529	.068	.470	.072	.238	.085	.495	.069
Grade 9	−.713	.091	−.563	.088	.450	.080	.465	.087	.198	.094	.540	.086

Note. Coefficients ES2 are standardized effect sizes for compositional effects (Marsh et al., 2009). *SE* = standard error. $p < .05$, 01, and .001 for coefficients higher than1.96, 2.58, and 3.29 *SE* , respectively.