International Journal of Educational Research 104 (2020) 101692

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Big-fish-little-pond effect on achievement emotions in relation to mathematics performance and gender

Marja Eliisa Holm [a,*], Johan Korhonen [b], Anu Laine [a], Piia Maria Björn [c], Markku Sakari Hannula [a]

[a] Faculty of Educational Sciences, University of Helsinki, P.O. Box 9, FI-00014, Helsinki, Finland
[b] Faculty of Education and Welfare Studies, Åbo Akademi University, P.O. Box 311, FI-65101, Vaasa, Finland
[c] Department of Education, University of Turku, FI-20014, Turku, Finland

ARTICLE INFO

Keywords:
Achievement emotions
Adolescents
Big-fish-little-pond effect
Gender
Mathematics performance

ABSTRACT

This study investigated the big-fish-little-pond effect (BFLPE) on mathematics-related achievement emotions (enjoyment, pride, anger, anxiety, shame, hopelessness, and boredom) among adolescents ($N = 1322$) using multilevel modeling, controlling for the effects of gender and classroom size. The results indicated that only pride was influenced by the BFLPE. Hence, adolescents reported less pride in mathematically higher-performing classrooms (higher class-average). The cross-level interaction effects indicated that the BFLPE varies across mathematics performance levels and gender. In mathematically higher-performing classrooms, adolescents with lower mathematics performance reported less pride and more shame, whereas adolescents with higher mathematics performance reported less enjoyment and more boredom. Additionally, males reported more shame in higher-performing classrooms. We discuss the practical implications of supporting achievement emotions in higher-performing classrooms.

1. Introduction

Experiencing pleasant achievement emotions is an important educational goal for adolescents, as such emotions affect every aspect of learning and are strongly related to students' wellbeing (Pekrun, 2017). Although being a member of a higher-performing classroom might benefit adolescents' academic learning, the big-fish-little-pond effect (BFLPE) indicates that adolescents in mathematically higher-performing classrooms might have a low mathematics self-concept (Marsh, Parker, & Pekrun, 2019) and experience unpleasant mathematics-related achievement emotions (Pekrun, Murayama, Marsh, Goetz, & Frenzel, 2019). This negative classroom-level effect is evident although higher academic performance is related to a higher self-concept or pleasant achievement emotions at the student level. The BFLPE may arise when individuals compare their academic performance with those of their peers in a higher-performing classroom. This comparison might lead them to feel negative about their success, so they may experience more unpleasant achievement emotions (Pekrun et al., 2019) and have a lower self-concept than adolescents with the same performance level in lower-performing classrooms (Dijkstra, Kuyper, van der Werf, Buunk, & van der Zee, 2008; Marsh & Hau, 2003; Marsh et al., 2008; Seaton, Marsh, & Craven, 2009).

Several empirical studies have confirmed that the BFLPE influences students' academic self-concept (Fang et al., 2018), especially

* Corresponding author.
E-mail addresses: marja.hytonen@helsinki.fi (M.E. Holm), johan.korhonen@abo.fi (J. Korhonen), anu.laine@helsinki.fi (A. Laine), piia.bjorn@utu.fi (P.M. Björn), markku.hannula@helsinki.fi (M.S. Hannula).

https://doi.org/10.1016/j.ijer.2020.101692
Received 25 April 2020; Received in revised form 22 June 2020; Accepted 5 October 2020
Available online 23 October 2020

in mathematics (Marsh et al., 2019; Salchegger, 2016; Seaton et al., 2009; Trautwein, Lüdtke, Marsh, & Nagy, 2009). Across countries and age groups, evidence exists that students' placement in a higher-performing mathematics classroom or school is associated with lower mathematics self-concepts (Marsh, Parker, Guo, Pekrun, & Basarkod, 2020, 2019; Salchegger, 2016; Seaton et al., 2009; Trautwein et al., 2009). As there is a link between self-concept and mathematics-related achievement emotions (Van der Beek, Van der Ven, Kroesbergen, & Leseman, 2017), the BFLPE may relate to these emotions. However, only a few studies have investigated the BFLPE on mathematics-related achievement emotions and found the BFLPE on enjoyment, anger, pride, anxiety, shame, and hopelessness among German students (Frenzel, Pekrun, & Goetz, 2007b; Pekrun et al., 2019). As previous studies have indicated that specific achievement emotions—pride and shame—are related mainly to unfavorable social upward comparison (e.g., comparing higher performers; Smith, 2000; Webster, Duvall, Gaines, & Smith, 2003), investigating whether higher classroom performance affects specific mathematics-related achievement emotions in other educational settings is important.

Another important issue is whether the BFLPE varies across students' performance levels or genders (Plieninger & Dickhäuser, 2015; Trautwein et al., 2009). Although researchers have found that females and mathematically lower-performing adolescents reported unpleasant mathematics-related achievement emotions (Frenzel, Pekrun, & Goetz, 2007a; Holm, Hannula, & Björn, 2017; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017), whether higher classroom performance (i.e., BFLPE) contributes to these unpleasant emotions is unclear. However, studies have suggested that unfavorable social comparisons (e.g., competition and contrasting) to classmates might be typical for males or mathematically lower-performing adolescents (Wehrens et al., 2010). According to the BFLPE, social comparison in higher-performing classrooms might increase adolescents' unpleasant mathematics-related achievement emotions (Pekrun et al., 2019). Thus, examining whether the BFLPE on mathematics-related achievement emotions is moderated by individuals' gender and mathematics performance is important.

Therefore, this study investigated whether the class-average mathematics performance has negative contextual effects on adolescents' mathematics-related achievement emotions and whether adolescents with certain performance levels (lower or higher) or gender are more affected. In line with previous studies (Marsh et al., 2019), we referred to this contextual effect as the BFLPE. This effect means that classroom-level performance has a negative association with individuals' outcomes, such as emotions, beyond that explained by individual characteristics (Marsh et al., 2009).

1.1. *Achievement emotions*

Several theories explain emotions and their association with learning (Hascher, 2010; Tracy & Robins, 2004; Weiner, 2014). We use the control-value theory (Pekrun, 2006), as it describes various achievement emotions and how they relate to academic outcomes. This theory categorizes achievement emotions into activity-related emotions, which are experienced in relation to academic activities, and outcome-related emotions, which are experienced before or after academic outcomes (e.g., success or failure; Pekrun, 2006). The former include enjoyment during learning and boredom during homework, whereas the latter include pride after a successful exam and anxiety before a mathematics test. Achievement emotions may typically be experienced in certain situations (i.e., traits) or momentarily experienced in certain situations at certain times (i.e., states). According to the control-value theory, individuals' appraisals of control (e.g., competence) and value (e.g., importance) affect their achievement emotions. As classroom characteristics influence students' appraisals of control and value and, thus, their emotions (Pekrun, 2006), these characteristics must be considered when investigating achievement emotions. The BFLPE may be a classroom characteristic related to students' achievement emotions.

1.2. *Relation of the BFLPE to achievement emotions*

Researchers have applied the BFLPE to examine achievement emotions (Pekrun et al., 2019), hypothesizing that studying in a higher-performing classroom makes it more difficult for a student to succeed relative to classmates and thus increases the likelihood of failure. Conversely, studying in a lower-performing classroom makes it easier for a student to succeed and reduces the likelihood of failure relative to others. Because success and failure drive achievement emotions, higher class-average performance is thought to reduce positive achievement emotions and increase negative achievement emotions (Pekrun et al., 2019).

Only a few studies have investigated the BFLPE in relation to achievement emotions (Frenzel et al., 2007b; Goetz, Preckel, Zeidner, & Schleyer, 2008; Pekrun et al., 2019). Two such studies have examined the BFLPE on mathematics-related achievement emotions and focused on German students (Frenzel et al., 2007b; Pekrun et al., 2019). Frenzel et al. (2007b) found that German students (grades 5–10) in higher-performing mathematics classrooms experienced less enjoyment and more anger and anxiety but not boredom. Pekrun et al. (2019) also found that the BFLPE influences enjoyment, pride, anger, anxiety, shame, and hopelessness among German adolescents (grades 7–9). However, some researchers have suggested that social comparisons are associated with specific achievement emotions, including anxiety, pride, and shame (Baudoin & Galand, 2017; Smith, 2000; Webster et al., 2003). Particularly, Baudoin and Galand's (2017) findings suggested that teachers' emphasis of social comparison and competition in classrooms caused adolescents to experience anger, shame, and anxiety. Researchers indicated that students' pride decreased (Webster et al., 2003) and their shame increased (Smith, 2000) as a result of unfavorable upward social comparison. The control-value theory also states that outcome achievement emotions, such as pride, anxiety, and shame, are related to individuals' perception of success and failure (Pekrun, 2006) and might therefore be particularly affected by higher class-average performance. Together, these studies suggested that investigating whether higher classroom performance affects specific mathematics-related achievement emotions in different educational contexts is important.

1.3. Individual performance as a moderator of the BFLPE

Research has shown that students' low mathematics performance is related strongly to several unpleasant mathematics-related achievement emotions (Frenzel et al., 2007b; Holm et al., 2017; Pekrun et al., 2017); thus, examining whether individuals' mathematics performance moderates the BFLPE on some achievement emotions is essential. However, studies have investigated only whether the BFLPE on self-concept differs among adolescents with different mathematics performance levels, yielding conflicting results (Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007, 2019; Seaton et al., 2009; Trautwein et al., 2009). Some studies have shown that the mathematics-related self-concepts of lower-performing adolescents are slightly more affected by the BFLPE than those of higher-performing adolescents (Trautwein et al., 2009). Other studies have found that the BFLPE was a little stronger for mathematically higher-performing adolescents than for lower-performing adolescents (Marsh et al., 2019; Seaton et al., 2009).

Although to our knowledge, no previous studies have investigated individual performance as a moderator for the BFLPE in the context of mathematics-related achievement emotions, some researchers have suggested that this relationship might exist. Previous studies have indicated that social comparison might increase adolescents' anger, anxiety, shame, and hopelessness and decrease their pride and enjoyment in mathematically higher-performing classrooms (Pekrun et al., 2019). However, unpleasant outcome emotions, such as a low level of pride, shame, anxiety, and hopelessness, are related to failure according to the control-value theory (Pekrun, 2006). Thus, such emotions might be typical for lower-performing adolescents in higher-performing classrooms. Some findings, indeed, have revealed that adolescents who perceived they performed worse than their classmates reported the lowest level of pride (Webster et al., 2003). In this regard, investigating whether the BFLPE affected mainly the outcome emotions of mathematically lower-performing adolescents is important.

1.4. Gender as a moderator of the BFLPE

Mathematics is stereotypically viewed as a male subject (Spencer, Steele, & Quinn, 1999), and females tend to experience more unpleasant mathematics-related achievement emotions than males do (Frenzel et al., 2007a; Holm et al., 2017; Pekrun et al., 2019). Therefore, understanding whether the BFLPE is moderated by gender is important. Although studies have shown that the BFLPE on the mathematics self-concept is slightly larger for female adolescents than for males in German samples (Marsh et al., 2007; Plieninger & Dickhäuser, 2015), whether this association exists in several mathematics-related achievement emotions remains unclear.

Some mechanisms suggest that the BFLPE for mathematics-related achievement emotions differed across genders. Specifically, previous findings indicated that females engage in social comparisons more often than males do (Guimond et al., 2007), so females might be more vulnerable to the negative effect of class-average performance (Plieninger & Dickhäuser, 2015). Although this vulnerability might cause females to report more unpleasant mathematics-related achievement emotions in higher-performing classrooms, previous studies have shown that females have more empathic responses to social comparison compared with males (Buunk, Kuyper, & van der Zee, 2005; Mestre, Samper, Frias, & Tur, 2009; Wehrens et al., 2010). Females might perceive that it is good for other students to succeed (Buunk et al., 2005; Wehrens et al., 2010). In turn, males often compared themselves to others, reflecting a stronger feeling of competition (e.g., I must be the best; Buunk et al., 2005; Wehrens et al., 2010). Previous findings suggested that contrastive and competitive comparison might relate to adolescents' anxiety, shame, and anger (Baudoin & Galand, 2017). Perhaps in higher-performing classrooms, males' contrasting comparison and competition might cause them to report more negative mathematics-related achievement emotions, such as shame, anxiety, and anger. The stereotype that mathematics is a male subject (Spencer et al., 1999) might also cause males in higher-performing classrooms to believe they are not fulfilling the stereotype and expectations. The control-value theory states that expectations exceeding students' capabilities are related to several unpleasant achievement emotions (Pekrun, 2006). Students who perform lower than expected can experience shame (Turner & Schallert, 2001). Hence, investigating whether higher classroom performance affects specific mathematics-related achievement emotions of males, such as anxiety, shame, and anger, is important.

1.5. The present study

Most empirical studies focused on the BFLPE in relation to self-concepts (meta-analysis; Fang et al., 2018) and investigated whether the BFLPE on adolescents' mathematics self-concepts varies across performance levels (Marsh et al., 2019; Seaton et al., 2009; Trautwein et al., 2009) and genders (Marsh et al., 2007; Plieninger & Dickhäuser, 2015). These associations are scarcely studied in relation to mathematics-related achievement emotions, although recent research suggested that the BFLPE should be considered in relation to multiple mathematics-related achievement emotions (Pekrun et al., 2019). To fill these gaps, the present study investigated the BFLPE on mathematics-related achievement emotions among a representative sample of Finnish adolescents and whether the BFLPE varies in strength across adolescents with different performance levels and genders.

We assessed subject-specific mathematics-related achievement emotions, not general achievement emotions, because it has been shown that achievement emotions are organized in subject-specific ways (Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007). As a school subject, mathematics plays a crucial role in students' future outcomes, and students struggling with mathematics set lower educational goals (Widlund, Tuominen, Tapola, & Korhonen, 2020) and are at risk of dropping out of school (Hakkarainen, Holopainen, & Savolainen, 2015). We examined trait emotions rather than state emotions because stable trait emotions relate to adolescents' emotional wellbeing (Pekrun, 2017). Test scores, rather than grades, were used as an indicator of mathematics performance, as classrooms may be graded on a curve (Köller, Zeinz, & Trautwein, 2008), so grades could not be used to assess differences in class-average achievement. In the analysis, classroom size (which varied in this study) was controlled for at the classroom level, and gender was controlled for at the individual and classroom levels. These controls were done because a larger class size is related to negative achievement emotions, such as anxiety (Khajavy, MacIntyre, & Barabadi, 2018), and gender is related to mathematics-related achievement emotions at the individual and classroom levels (Frenzel et al., 2007a; Holm et al., 2017; Pekrun et al., 2019).

The first research question is as follows: Are mathematics-related enjoyment, pride, anger, anxiety, shame, hopelessness, and boredom influenced by the BFLPE? Previous findings indicated that upward social comparison reduced students' pride and increased their shame (Smith, 2000; Webster et al., 2003); social comparison and competition in classrooms caused adolescents to experience anger, shame, and anxiety (Baudoin & Galand, 2017); and the BFLPE influenced mathematics-related enjoyment, pride, anger, anxiety, shame, and hopelessness (Frenzel et al., 2007b; Pekrun et al., 2019). Therefore, we hypothesized that the BFLPE influences mathematics-related enjoyment, pride, anxiety, anger, shame, and hopelessness among adolescents. As Frenzel et al. (2007b) did not find the BFLPE on boredom, we expected that there is no BFLPE on boredom. Specifically, we expected that adolescents in mathematically higher-performing classrooms would experience less mathematics-related enjoyment and pride and more anxiety, anger, shame, and hopelessness compared with adolescents with the same performance level in lower-performing classrooms.

The second research question is as follows: Does the BFLPE on the mathematics-related achievement emotions of adolescents vary across mathematics performance levels? The empirical findings on these interaction effects are limited. However, based on the literature on social comparison and theory of achievement emotions (Pekrun, 2006; Webster et al., 2003), we hypothesized that the BFLPE on outcome achievement emotions, such as pride, shame, hopelessness, and anxiety, would be stronger for mathematically lower-performing adolescents than for higher-performing ones.

The third research question is as follows: Does the BFLPE have a greater impact on mathematics-related achievement emotions for females or males? Empirical findings on these interaction effects are limited. However, based on the literature on social comparison and gender (Baudoin & Galand, 2017; Buunk et al., 2005; Spencer et al., 1999; Turner & Schallert, 2001; Wehrens et al., 2010), we hypothesized that the BFLPE on mathematics-related shame, anxiety, and anger would be higher for males than for females.

2. Method

2.1. Participants

The sample comprised eighth-grade participants (14–15 years old) from compulsory schools in all five provinces of Finland. In Finland, students attend six years of primary school, usually beginning in the year they turn seven, followed by three years of lower secondary school. We used cluster, stratified, and systematic sampling methods (Lehtonen & Pahkinen, 2004) to obtain a geographically representative sample. We drew the school sample from a statistical list of Finnish compulsory schools (i.e., cluster sampling). Before drawing up the sample, we sorted the schools by province and by municipality or city in each province (i.e., implicitly stratified sampling). Thus, the sample was selected according to the proportion of schools in each province. We drew each qth school (sampling interval: $q = N / n$; N = all schools, n = selected schools) from this stratified list (i.e., systematic sampling). We recruited all general education classrooms with eighth-grade students from the sample schools.

As Table 1 shows, the final sample included 1322 students from 77 general education classrooms ($M = 17.17$, $SD = 3.70$, minimum = 10, maximum = 26) in 27 schools. Table 1 also presents the mean age of the students ($M = 14.44$ and $SD = .53$), the number of females ($n = 688$) and males ($n = 634$), the students' average mathematics performance ($M = 23.02$ and $SD = 7.24$), and the class-average mathematics performance ($M = 22.85$ and $SD = 3.47$).

2.2. Measures

2.2.1. Mathematics-related achievement emotions

We used the Achievement Emotions Questionnaire–Mathematics (AEQ-M; Pekrun, Goetz, & Frenzel, 2005) to assess mathematics-related achievement emotions. This self-reported instrument contains 60 items that assess trait emotions—enjoyment, pride, anger, anxiety, shame, hopelessness, and boredom—related to mathematics classrooms, learning, and testing. The respondents were asked to express their emotions on a 5-point Likert scale. The original English version of the AEQ-M is available in the user's manual (Pekrun et al., 2005). This version was translated into Finnish by a bilingual expert and then pilot-tested in a Finnish school. Thirty students from two classrooms and 10 students receiving special education filled out the pilot questionnaire and offered feedback to their teachers and the researcher. They provided no negative feedback regarding the language and structure of the AEQ-M and reported that they understood it well. In this study, we used 52 AEQ-M items measuring enjoyment (10 items; e.g., enjoyment in the

Table 1
Demographics of the Study Groups.

Variable	Sample
Students, *n*	1322
Age, *M(SD)*	14.44(.53)
Boys, *n*	634
Mathematics performance, *M(SD)*	23.02(7.24)
Classrooms, *n*	77
Average class size, *M*(*SD*)	17.17(3.70)
Average mathematics performance in classes, *M*(*SD*)	22.85(3.47)

mathematics classroom), pride (five items; e.g., pride after a mathematics test), anger (nine items; e.g., anger because of mathematics homework), anxiety (nine items; e.g., so anxious that could not take the mathematics test), shame (seven items; e.g., shame after a mathematics test), hopelessness (six items; e.g., hopelessness during a mathematics test), and boredom (six items; e.g., boredom during mathematics homework).

We eliminated eight of the 60 items from the original questionnaire because confirmatory factor analyses showed poor model fits for anxiety, shame, and pride (comparative fit index [CFI] $< .88$; root mean square error of approximation [RMSEA] $> .10$; see Section 2.4.3). High modification indices (Byrne, 2012) suggested high correlations between similarly worded emotion item residuals related to the same situations (classroom, learning, and testing). These correlations were in line with AEQ-M models, including correlations between all item residuals representing the same situations (Pekrun et al., 2019). To avoid excessive statistical model complexity, we excluded the other correlated item. After removal of the eight items, the model fits for all emotions were acceptable (see Appendix A). Cronbach's α was calculated as a measure of the internal consistency reliability of the emotion scale. We also used composite reliability (CR) to determine the internal consistency reliability of the corresponding emotion factors (see Section 2.4.3). As Appendix A shows, the reliability of enjoyment (α = .89 and CR = .90), pride (α = .80 and CR = .81), anger (α = .90 and CR = .90), anxiety (α = .87 and CR = .87), shame (α = .80 and CR = .81), hopelessness (α = .89 and CR = .89), and boredom (α = .87 and CR = .87) ranged from good ($.8 \leq x < .9$) to excellent ($x \geq .9$). Notably, the results did not differ from those obtained by analyzing the original AEQ-M.

2.2.2. Mathematics performance

To assess mathematics performance, we used a Finnish standardized test for mathematical skills for grades 7–9 (the KTLT; Räsänen & Leino, 2005). The KTLT assesses core mathematics skills, including arithmetic (e.g., addition, subtraction, and multiplication), word problems (e.g., solving for the tax rate), algebra (e.g., equation solving), geometry (e.g., solving for the surface area), and unit conversion (e.g., rounding a large number to the nearest hundred). This test is widely used in Finland and has been shown to have good internal reliability (Holm et al., 2018; Korhonen, Tapola, Linnanmäki, & Aunio, 2016; Pesu, Aunola, Viljaranta, & Nurmi, 2016) and good criterion validity with other measures of mathematical skills (Räsänen & Leino, 2005). The KTLT is a paper-and-pencil test consisting of 40 items, with one point given for a correct answer and zero for an incorrect one (for a maximum of 40 points). Of the four versions (A, B, C, and D), we chose version B, as it has the highest reported internal reliability (Cronbach's α = .90; Räsänen & Leino, 2005). In this study, the reliability of the KTLT was good (α = .89).

2.3. Procedures

The assessment was completed in spring 2010. Throughout the study, we followed the Finnish ethical principles of research in the humanities and social and behavioral sciences (National Advisory Board on Research Ethics, 2009). Permission for the study was obtained from municipal education departments and headteachers, and informed consent was obtained from the students' parents. The study materials were sent to the schools via regular mail, and detailed instructions for implementing the research were provided to the teachers. The students were given 30 min to complete the AEQ-M at the end of a mathematics lesson. The teachers read the instructions for the AEQ-M. Before the AEQ-M was administered, the students were assured that their responses would be confidential and that there were no right or wrong answers, and they were asked to express their personal opinions. During another lesson, the students were given 40 min to complete the KTLT. After all the measures were carried out, the teachers collected the materials and returned them to the researchers via regular mail.

2.4. Data analysis

2.4.1. Missing values

We imputed the missing data for the AEQ-M (0.5 %) in SPSS 25 using an expectation–maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). We used the EM algorithm rather than other algorithms, such as full information maximum likelihood (FIML), because the EM algorithm also provides unbiased estimates, produces quite consistent results with other algorithms (e.g., FIML; Enders & Peugh, 2004), and is recommended if the percentage of missing data is less than 5 % (Hazzi & Maldaon, 2015). Other emotion items served as auxiliary variables to impute the missing value of the scale items. We excluded 55 students from the analysis, as they did not complete the KTLT.

2.4.2. Multilevel modeling

Using multilevel modeling (MLM) with Mplus (version 7; Muthén & Muthén, 1998–2013), we investigated the BFLPE on mathematics-related achievement emotions. We controlled for gender at the individual and classroom levels and for classroom size at the classroom level. All models were estimated with the robust maximum likelihood estimator (MLR), as it is robust to the non-normality of the observed variables. We used separate models for each emotion because the number of parameters in the seven-factor model was higher than the number of classrooms (i.e., the number of clusters). In MLM, having more clusters than parameters is advisable (Muthén, 2008). All models were doubly latent models in which latent individual-level variables, mathematics performance[1] (latent predictor), and emotion (latent dependent variable) were modeled as latent constructs at the classroom level (i.e., latent

[1] Mathematics performance is a one-factor model in which standardized arithmetic, word problems, algebra, geometry, and unit conversion scales are indicators (CFI $> .98$; RMSEA $< .05$).

aggregation; Marsh et al., 2009). The classroom-level factor loadings were constrained to be equal to the individual-level factor loadings. We used the doubly latent approach because it accounted for sampling and measurement errors (Marsh et al., 2009). The model also included a manifest predictor—dummy-coded gender—which was aggregated at the classroom level (i.e., manifest aggregation; Marsh et al., 2009).

In the preliminary analysis, we separately tested fully unconditional models of each emotion, defining latent constructs at the individual and classroom levels. In order to determine whether MLM was required, we determined the intraclass correlations (ICCs; Garson, 2013) to obtain an estimate of the proportion of variance in each emotion between classrooms. The ICC for mathematics performance was also determined. Analyses at the classroom level were warranted if the ICC was .05 or higher (LeBreton & Senter, 2008).

In the main analysis, we investigated the BFLPE on achievement emotions, controlling for gender at the individual and classroom levels and classroom size at the classroom level. We utilized models in which latent mathematics performance and dummy-coded gender (males = 1) predicted emotions at the individual level and latent aggregated mathematics performance, gender proportion, and classroom size predicted emotions at the classroom level. In MLM, centering individual-level predictor variables is crucial (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009). In the doubly latent model, the indicators of latent mathematics performance are grand-mean centered, causing independent effects at the classroom level (Marsh et al., 2009). The individual-level dummy predictor, gender, is group-mean centered (i.e., centering within cluster; Enders & Tofighi, 2007), also causing independent effects at the classroom level. A parameter, the contextual effect (BFLPE), was constructed using the model constraint command in Mplus (Marsh et al., 2009). The contextual effect is present if the classroom-level effect of the latent aggregated mathematics performance is significantly different from the corresponding individual-level effect of the students' mathematics performance. A significant contextual effect indicates that the average mathematics test score at the classroom level is related to students' emotions at the individual level, beyond what can be explained by individual-level effects. The effect sizes for contextual effects were estimated using effect size 2 (ES2[2]) defined for continuous level 2 predictors in MLMs (Marsh et al., 2009). ES2 is comparable with Cohen's *d*.

Finally, we probed for possible interaction effects. Using the random-slope model (Marsh et al., 2009), we specified two cross-level interaction effects: (1) the interaction effects between the class-average mathematics performance and individual mathematics performance and (2) between the class-average mathematics performance and gender. A significant interaction effect indicates that students with different mathematics performance levels or gender are affected differently by the BFLPE.

2.4.3. *Fit indices*

We used the chi-square (χ^2), CFI, Tucker–Lewis index (TLI), and RMSEA as model fit indicators. CFI and TLI values greater than .90 and RMSEA values less than .08 show acceptable fit with the data (Marsh, Hau, & Wen, 2004). Composite reliability (Geldhof, Preacher, & Zyphur, 2014) was calculated as a measure of the internal consistency reliability of the emotion factors.

3. Results

3.1. *Unconditional multilevel models*

In the preliminary analysis, the separate unconditional multilevel models of emotions showed a good model fit, and the composite reliability of the factors was good at both the individual and classroom levels (see Table 2). The between variance at the classroom level was significant for all seven emotions. As Table 2 shows, the ICC is acceptable ($\geq$ .05, which indicates a small-to-medium effect), warranting analysis at the classroom level (LeBreton & Senter, 2008). The between variance at the classroom level was significant for mathematics performance. The ICC for mathematics achievement was .19, which indicates a medium effect ($\geq$ .10, LeBreton & Senter, 2008).

3.2. *The BFLPE on mathematics-related achievement emotions*

In the main analysis, we examined the BFLPE on mathematics-related achievement emotions using MLM, controlling for the effects of gender at the individual and classroom levels and classroom size at the classroom level. Table 3 shows the correlations between the latent variables used in MLM. Table 4 presents the unstandardized coefficients (*B*) and standard error (*SE*) of this analysis.

At the individual level, mathematics performance was significantly linked to enjoyment (B = .51) and pride (B = .83) and negatively linked to anger ($B = -.36$), anxiety ($B = -.45$), shame ($B = -.23$), hopelessness ($B = -.49$), and boredom ($B = -.23$). At the classroom level, the class-average mathematics performance was significantly related to class-average anger ($B = -.28$) and hopelessness ($B = -.29$). Thus, on average, less mathematics-related hopelessness and anger were reported in classrooms with higher class-average mathematics performance. The contextual effects on emotions were present if the effects of class-average mathematics performance at the classroom-level were significantly different from the corresponding individual-level effects. As shown in Table 4, this is evident for pride, as the class-average mathematics performance at the classroom-level had a significant negative contextual effect on students' pride ($B = -.58$, $SE = .14$, $p < .001$; ES2 = $-.36$). The effect size (ES2) was moderate. This BFLPE indicated that individuals in higher-performing classrooms experienced less pride than individuals with the same performance level in lower-performing

[2] ES2 = (2 x B x SD_{pre})/SD_{out}, where B is the unstandardized coefficient for the contextual effect, SD_{pre} is the SD of the predictor, and SD_{out} is the SD of the individual-level outcome.

Table 2
Unconditional Models of Achievement Emotions.

Emotions	χ^2	*df*	TLI	CFI	RMSEA	ICC	CR *w/b*
Enjoyment	397.27***	79	.95	.94	.06	.12	.89/.98
Pride	104.59***	14	.96	.94	.07	.05	.80/.94
Anger	440.96***	62	.93	.92	.07	.07	.89/.97
Anxiety	391.67***	62	.93	.92	.06	.05	.87/.95
Shame	123.53***	34	.96	.96	.05	.05	.80/.94
Hopelessness	177.05***	23	.96	.95	.07	.05	.89/.97
Boredom	58.58***	23	.99	.98	.03	.05	.87/.91

Note. TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; ICC = intraclass coefficient; CR = composite reliability; *w* = within; *b* = between.
*** $p < .001$.

Table 3
Correlations Between Individual and Classroom-Level Predictors and Mathematics-Related Achievement Emotions.

	Enjoyment	Pride	Anger	Anxiety	Shame	Hopelessness	Boredom
Individual							
Math	.41	.54	−.29	−.34	−.32	−.39	−.19
Gender	.04	.10	.07	.02	.04	−.08	−.01
Classroom							
Math_A	.29	.38	−.42	−.46	−.37	−.51	−.30
Gender_P	.32	.28	.04	.18	.33	.18	−.21
Class size	.16	.09	−.13	.02	.26	−.00	−.20

Note. P = classroom-level proportion; A = classroom average. The correlations are based on the latent variables.

Table 4
The Big-Fish-Little-Pond Effect (BFLPE) on Mathematics-Related Achievement Emotions.

	Enjoyment *B*(*SE*)	Pride *B*(*SE*)	Anger *B*(*SE*)	Anxiety *B*(*SE*)	Shame *B*(*SE*)	Hopelessness *B*(*SE*)	Boredom *B*(*SE*)
Individual							
MathI	.51(.05)***	.83(.07)***	−.36(.04)***	−.45(.04)***	−.23(.03)***	−.49(.05)***	−.23(.04)***
Sex	.07(.06)	.20(.07)**	.11(.06)	.03(.05)	.03(.04)	−.13(.05)**	−.02(.06)
Classroom							
MathA	.26(.15)	.25(.14)	−.28(.13)*	−.26(.14)	−.10(.06)	−.29(.12)*	−.14(.12)
SexP	.60(.26)*	.40(.27)	.08(.19)	.20(.21)	.19(.12)	.23(.21)	−.20(.20)
Class size	.01(.01)	.00(.01)	−.01(.01)	.00(.01)	.01(.01)	−.00(.01)	−.01(.01)
Contextual[a]							
MathA	−.26(.16)	−.58(.14)***	.08(.13)	.19(.15)	.12(.07)	.20(.13)	.09(.13)

Note. MathI = students mathematics performance; MathA = classroom-average mathematics performance; SexP = gender proportions in the classrooms; The coefficients (*B*) are unstandardized.
[a] Contextual effect (BFLPE) is the difference between the unstandardized classroom-level regression coefficient and the individual-level regression coefficient. The effect sizes (ES2) for contextual effects on the emotions were −.20 (enjoyment), −.36 (pride), .06 (anger), .14 (anxiety), .17 (shame), .16 (hopelessness), and .07 (boredom).
* $p < .05$.
** $p < .01$.
*** $p < .001$.

classrooms. Although the effect sizes also indicated substantial BFLPEs on enjoyment (ES2 = −.20), anxiety (ES2 = .14), shame (ES2 = .17), and hopelessness (ES2 = .16), these effects did not reach significance (Table 4). The effect sizes regarding boredom and anger were negligible (ES2 < .10), indicating no BFLPEs on anger and boredom.

Next, cross-level interaction effects were specified and tested to determine their statistical significance in the random-slope models. Notably, all significant effects presented in Table 4 remained significant when we added the two cross-level interactions (class-average math × individual math [1]; class-average math × individual gender [2]) to the model. We found statistically significant regression coefficients for the interaction effect between class-average mathematics performance and individual mathematics performance on enjoyment ($B = -.07$, $SE = .02$, $p < .001$), pride ($B = .26$, $SE = .06$, $p < .001$), shame ($B = -.13$, $SE = .01$, $p < .001$), and boredom ($B = .20$, $SE = .01$, $p < .001$). This finding indicated that the BFLPE had a greater impact on the pride and shame of individuals with lower mathematics performance compared with their higher-performing peers. Hence, adolescents with lower mathematics performance reported less pride and more shame in mathematically higher-performing classrooms than adolescents with the same performance level in lower-performing classrooms. In turn, the significant cross-level interactions indicated that the BFLPE had a greater impact on the enjoyment and boredom of individuals with higher mathematics performance than those of their lower-performing peers. Hence, adolescents with higher mathematics performance reported less enjoyment and more boredom in mathematically higher-performing classrooms than adolescents with the same performance level in lower-performing classrooms.

We also found a statistically significant regression coefficient for the interaction effect between class-average mathematics performance and gender on shame ($B = .27$, $SE = .10$, $p < .01$). Hence, the BFLPE had a greater impact on shame for males than for females. Males reported more shame in classrooms with higher average mathematics performance than males in classrooms with lower mathematics performance.

4. Discussion

We investigated the BFLPE on mathematics-related achievement emotions and its interactions with mathematics performance and gender. First, on the whole sample level, we found that the BFLPE influenced only pride. Adolescents in mathematically higher-performing classrooms reported less pride compared with adolescents with the same performance level in lower-performing classrooms. Second, the significant interaction effects indicated that the BFLPE influenced shame and pride among mathematically lower-performing adolescents and enjoyment and boredom among higher-performing adolescents. Thus, in mathematically higher-performing classrooms, lower-performing adolescents reported less pride and more shame, whereas higher-performing adolescents reported less enjoyment and more boredom. Third, the significant interaction effects indicated that the BFLPE influenced males' shame. Thus, males reported more shame in mathematically higher-performing classrooms than in lower-performing classrooms.

4.1. The BFLPE related to pride

The results did not fully support our hypothesis that adolescents in higher-performing classrooms report less enjoyment and pride and more anger, anxiety, shame, and hopelessness. We found that the BFLPE influenced only pride. Our results extend the findings that adolescents' pride might decrease as a result of unfavorable upward comparisons to others (Smith, 2000; Webster et al., 2003). These results altogether suggest that pride might be a crucial emotion in social comparison. Our results build on previous studies confirming the BFLPE on adolescents' mathematics self-concept in several countries, including Finland (Salchegger, 2016, Marsh et al., 2019, 2020; Seaton et al., 2009; Trautwein et al., 2009). In addition to self-concept, our results suggest that pride should be considered when the BFLPE is investigated. The control-value theory states that pride is associated with self-inflicted success (Pekrun, 2006). Perhaps students perceived their own success as inadequate in classrooms with higher-performing peers, so they did not experience pride.

Our results support the finding that students in higher-performing mathematics classrooms do not report boredom (Frenzel et al., 2007b). Boredom is often related to valueless academic activities (Pekrun, 2006), so this emotion might not play an important role in social comparison. We found that adolescents in higher-performing classrooms did not report more unpleasant emotions except less pride. However, our results also indicated substantial BFLPEs on enjoyment, anxiety, shame, and hopelessness in terms of the effect sizes, even if these effects did not reach significance. Such findings are in line with previous results that found the BFLPE on mathematics-related anxiety, enjoyment, shame, and hopelessness among German students (Frenzel et al., 2007b; Pekrun et al., 2019).

In our study, the effect sizes for other emotions except pride were not significant. A relatively small sample size at the classroom level and differences in the educational contexts might partly explain this disparity. In the present sample, performance-based tracking (i.e., division of students into different school types by performance) was not used; previous studies focused on students from schools with different performance-based tracking systems (begun after the age of 10; Frenzel et al., 2007b; Pekrun et al., 2019). Previous studies have shown that the BFLPE on self-concept is much smaller in countries without school-level tracking, such as Finland (Salchegger, 2016). A tracking system may limit students' awareness of their actual potential (e.g., higher-performing students perceive their performance as lower) and increase competition and comparison, which has been suggested to exacerbate the BFLPE (Salchegger, 2016). However, we found a significant BFLPE on pride even when a tracking system was not used. Future research should explore the connections between the BFLPE and achievement emotions more extensively in different cultural contexts.

4.2. Mathematics performance levels moderate different emotions

Our study was the first to show that the BFLPE on pride and shame is stronger for mathematically lower-performing adolescents than for higher-performing adolescents, and the BFLPE on enjoyment and boredom is stronger for mathematically higher-performing adolescents than for lower-performing adolescents. These results partly support our hypothesis that the BFLPE on outcome-related emotions, such as pride, anger, shame, and anxiety, is stronger for lower-performing adolescents than for higher-performing ones. These results extend previous findings that students who perceived they performed worse than others experienced less pride (Webster et al., 2003). We found that mathematically lower-performing adolescents reported less pride and more shame in higher-performing classrooms than adolescents with the same performance level in lower-performing classrooms. However, we found that mathematically higher-performing adolescents reported more unpleasant activity-related achievement emotions—more boredom and less enjoyment—in higher-performing classrooms than adolescents with the same performance level in lower-performing classrooms. Previous studies found the BFLPE on mathematics-related enjoyment among German adolescents (Frenzel et al., 2007b; Pekrun et al., 2019). We found that this was evident only for mathematically higher-performing adolescents. Previous studies did not find the BFLPE on mathematics-related boredom among German students (Frenzel et al., 2007b). We found the BFLPE on boredom only among mathematically higher-performing adolescents.

We found that the BFLPE is related to outcome emotions (shame and pride) among lower-performing adolescents and activity-related emotions (boredom and enjoyment) among higher-performing adolescents. In this case, the results support the control-value theory, which categorizes achievement emotions into outcome- and activity-related emotions (Pekrun, 2006). Our results suggest that this theoretical distinction should be considered when the BFLPE is investigated. Some studies have found that the BFLPE influences the self-concept of lower-performing adolescents; others have found that this effect is more obvious in higher-performing adolescents (Marsh et al., 2019, Seaton et al., 2009; Trautwein et al., 2009). The present results also indicated that the BFLPE can be related to unpleasant emotions among lower- and higher-performing adolescents, but the results depend on whether the achievement emotions are outcome or activity related.

Why do such differences occur in relation to activity- and outcome-related achievement emotions? Specifically, lower-performing adolescents might report outcome-related emotions, more shame, and less pride in higher-performing classrooms because of upward social comparison (Smith, 2000; Webster et al., 2003). The control-value theory states that students experience less pride and more shame when they perceive that they caused their own failure (Pekrun, 2006). Perhaps mainly lower-performing adolescents experience more that they will not succeed or exert control when they compare themselves to classmates in higher-performing classrooms and thus report more shame and less pride. The control-value theory states that boredom and a low level of enjoyment are related to valueless mathematics activities (Pekrun, 2006). Hence, social comparison to classmates does not necessarily explain the fact that higher-performing adolescents reported boredom and a low level of enjoyment in higher-performing classrooms. Perhaps learning and teaching methods can also explain these results. Previous studies, indeed, suggested that teachers in higher-performing classrooms progress quickly in mathematics teaching, giving quick demonstrations of method without explanation and without allowing students to discover the meaning of different methods (Boaler, Wiliam, & Brown, 2000). Teachers might expect that students understand without explanation, as they are top performers in the subject (Boaler et al., 2000). These mechanisms might cause mathematics learning to be valueless, so higher-performing adolescents might report more boredom and less enjoyment in higher-performing classrooms. Future studies should investigate whether different teaching and learning methods rather than social comparison explain the fact that higher-performing adolescents reported such activity-related emotions in higher-performing classrooms.

4.3. The BFLPE related to males' shame

Our study was the first to show that males reported more shame in classrooms with higher-performing classmates than with lower-performing classmates. Our results partly support the hypothesis that the BFLPE on mathematics-related anxiety, shame, and anger is higher for males than for females. We found this result for shame, but not for anger and anxiety. Previous findings showed that males might be more competitive and tend to focus more on how they differ from their classmates (Wehrens et al., 2010). Our results suggested that this contrast comparison might cause males to experience shame in higher-performing classrooms. The stereotype that mathematics is a male subject might also explain these results (Spencer et al., 1999). According to the control-value theory, shame is related to self-inflicted failure (Pekrun, 2006). Perhaps males in higher-performing classrooms believe that they are not fulfilling the stereotype and expectations of being good at mathematics, so they experience shame (Turner & Schallert, 2001). Although previous studies showed that females tend to experience more unpleasant emotions (Frenzel et al., 2007a; Holm et al., 2017; Pekrun et al., 2019), our results suggest that being with higher-achieving peers was not necessarily related to females' unpleasant achievement emotions.

4.4. Limitations and future implications

Some limitations of this research should be addressed in future studies. We used a self-report measure to assess achievement emotions. Although the AEQ-M showed good internal reliability in our study and in others (e.g., Frenzel et al., 2007b; Lazarides & Buchholzb, 2019) and might be the easiest type of instrument to implement in the classroom, future studies could use other methods, such as physiological measurements, to analyze achievement emotions. As this study focused on trait emotions rather than momentarily experienced state emotions, future studies should examine the BFLPE in relation to state emotions. Future research should include other variables, such as motivation and attitudes, in the model when investigating the BFLPE on achievement emotions (Dai & Rinn, 2008).

Although the present investigation was based on a representative sample, generalizability is always an issue. As this sample was restricted to Finnish adolescents in mathematics classes, future studies should focus on the primary or preschool level, other school subjects, and different cultures. Notably, in our Finnish data, the ICC for mathematics performance was relatively low, indicating that the variance in mathematics performance between classrooms was not very substantial. This low variance might be attributed to the specifics of the Finnish school system in which the variance in mathematics performance between schools is shown to be quite small (Schleicher, 2019). Although the conditions for detecting the BFLPE of mathematics performance were not optimal, this study still found such effects. Hence, future studies on achievement emotions should consider the BFLPE. Finally, as this study used a cross-sectional design, longitudinal and experimental studies are needed to clarify causalities.

4.5. Practical implications

Our results suggest that placing individuals in higher-performing classrooms may incur an emotional cost, but not all achievement emotions were harmed. Lower-performing adolescents' pride and shame, higher-performing adolescents' enjoyment and boredom, and males' shame are vital to consider in higher-performing classrooms. Policymakers and educators should develop educational strategies to reduce the emotional cost associated with higher-achieving classrooms.

First, our results suggested that teachers and educators should promote more pride and less shame among lower-performing adolescents in higher-performing classrooms. The BFLPE could be reinforced with highly competitive classroom environments that emphasize normative feedback and rank ordering of students (Marsh et al., 2007). Mathematically lower-performing students do not necessarily receive enough emotional support in general mathematics classrooms (Holm, Björn, Laine, Korhonen, & Hannula, 2020). Teachers could support, give individualized feedback, and emphasize the strengths of mathematically lower-performing students instead of comparing them with their higher-performing classmates (Goetz et al., 2008; Pekrun et al., 2019). Thus, lower-performing students might experience less that they performed worse than others in higher-performing classrooms, and their pride might increase and their shame might decrease (Smith, 2000; Webster et al., 2003). In addition, differentiated instruction strategies that match students' abilities may attenuate the BFLPE for lower-performing students (Roy, Guay, & Valois, 2015). Varying goals, materials, assignments, and support that match the learning needs of mathematically lower-performing students might lessen their comparison with higher-performing peers and increase their pride and decrease their shame.

Second, our results suggested that in higher-performing classrooms, teachers and educators should promote more enjoyment and less boredom among mathematically higher-performing adolescents. More meaningful and valuable instructions, feedback, and learning activities in higher-performing classrooms might decrease boredom and increase enjoyment among higher-performing adolescents (Pekrun, 2006). Third, the results suggested that teachers should prevent shame among male adolescents in higher-achieving classrooms. Teachers could provide support and give individualized feedback to males in higher-performing classrooms and avoid reinforcing the stereotype that males should perform well in mathematics (Spencer et al., 1999).

Funding

This research was supported by grants from the Finnish Cultural Foundation for the first author.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. The Goodness-of-Fit Statistics for Mathematics-Related Achievement Emotions Observed in Confirmatory Factor Analysis

Models	Items	χ^2	df	CFI	TLI	RMSEA	CR	α
Enjoyment	10	283.81***	35	.95	.94	.073	.90	.89
Pride	5	46.32***	5	.96	.92	.079	.81	.80
Anger	9	252.12***	27	.94	.92	.079	.90	.90
Anxiety	9	219.91***	27	.93	.91	.074	.87	.87
Shame	7	62.13***	14	.97	.95	.051	.81	.80
Hopelessness	6	80.17***	9	.96	.94	.077	.89	.89
Boredom	6	28.62***	9	.99	.99	.041	.87	.87

Note. df = degree of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CR = composite reliability; α = Cronbach's α.

***$p < .001$.

References

Baudoin, N., & Galand, B. (2017). Effects of classroom goal structures on student emotions at school. *International Journal of Educational Research, 86*, 13–22. https://doi.org/10.1016/j.ijer.2017.08.010.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping: Disaffection, polarisation and the construction of failure. *British Educational Research Journal, 26*, 631–648. https://doi.org/10.1080/713651583.

Buunk, B. P., Kuyper, H., & van der Zee, Y. G. (2005). Affective response to social comparison in the classroom. *Basic and Applied Social Psychology, 27*, 229–237. https://doi.org/10.1207/s15324834basp2703_4.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* New York, NY: Taylor & Francis Group.

Dai, D. Y., & Rinn, A. N. (2008). The big-fish–little-pond effect: What do we know and where do we go from here? *Educational Psychology Review, 20*, 283–371. https://doi.org/10.1007/s10648-008-9071-x.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*(1), 1–38.

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*, 828–879. https://doi.org/10.3102/0034654308321210.

Enders, C. K., & Peugh, J. L. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling, 11*, 1–19. https://doi.org/10.1207/S15328007SEM1101_1.

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

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontier psychology, 29*, 1–11. https://doi.org/10.3389/fpsyg.2018.01569.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007a). 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. https://doi.org/10.1007/BF03173468.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007b). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction, 17*, 478–493. https://doi.org/10.1016/j.learninstruc.2007.09.001.

Garson, G. D. (2013). *Hierarchical linear modeling: Guide and applications*. Thousand Oaks, California: Sage Publications, Inc.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*, 72–91. https://doi.org/10.1037/a0032138.

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://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. https://doi.org/10.1080/10615800701628827.

Guimond, S., Branscombe, N. R., Brunot, S., Buunk, A. P., Chatard, A., Désert, M., … Yzerbyt, V. (2007). Culture, sex, and the self: Variations and impact of social comparison processes. *Journal of Personality and Social Psychology, 92*, 1118–1134. https://doi.org/10.1037/0022-3514.92.6.1118.

Hakkarainen, A. M., Holopainen, L. K., & Savolainen, H. K. (2015). A five-year follow-up on the role of educational support in preventing dropout from upper secondary education in Finland. *Journal of Learning Disabilities, 48*, 408–421. https://doi.org/10.1177/0022219413507603.

Hascher, T. (2010). Learning and emotion: Perspectives for theory and research. *European Educational Research Journal, 9*, 13–28. https://doi.org/10.2304/eerj.2010.9.1.13.

Hazzi, O., & Maldaon, I. (2015). A pilot study: Vital methodological issues. *Business: Theory and Practice, 16*, 53–62. https://doi.org/10.3846/btp.2015.437.

Holm, M. E., Hannula, M. S., & Björn, P. M. (2017). Mathematics-related emotions among Finnish adolescents across different performance levels. *Educational Psychology, 37*, 205–218. https://doi.org/10.1080/01443410.2016.1152354.

Holm, M. E., Aunio, P., Björn, P. M., Klenberg, L., Korhonen, J., & Hannula, M. S. (2018). Behavioral executive functions among adolescents with mathematics difficulties. *Journal of Learning Disabilities, 51*, 578–588. https://doi.org/10.1177/0022219417720684.

Holm, M. E., Björn, P. M., Laine, A., Korhonen, J., & Hannula, M. S. (2020). Achievement emotions among adolescents receiving special education support in mathematics. *Learning and Individual Differences, 79*. https://doi.org/10.1016/j.lindif.2020.101851. Advance online publication.

Khajavy, G. H., MacIntyre, P. D., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate: Applying doubly latent multilevel analysis in second language acquisition research. *Studies in Second Language Acquisition, 40*, 605–624. https://doi.org/10.1017/S0272263117000304.

Köller, O., Zeinz, H., & Trautwein, U. (2008). Class-average achievement, marks, and academic self-concept in German primary schools. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self-processes, learning, and enabling human potential* (pp. 331–352). Charlotte, NC: Information Age.

Korhonen, J., Tapola, A., Linnanmäki, K., & Aunio, P. (2016). Gendered pathways to educational aspirations: The role of academic self-concept, school burnout, achievement and interest in mathematics and reading. *Learning and Instruction, 46*, 21–33. https://doi.org/10.1016/j.learninstruc.2016.08.006.

Lazarides, R., & Buchholzb, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction, 61*, 45–59. https://doi.org/10.1016/j.learninstruc.2019.01.001.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815–852. https://doi.org/10.1177/1094428106296642.

Lehtonen, R., & Pahkinen, E. (2004). *Practical methods for design and analysis of complex surveys* (2nd ed.). Chichester, England: John Wiley & Sons, Ltd.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*, 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001.

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

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on the hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341. https://doi.org/10.1207/s15328007sem1103_2.

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

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

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O., & 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. https://doi.org/10.1080/00273170903333665.

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist, 24*, 231–242. https://doi.org/10.1027/1016-9040/a000332.

Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human cross-cultural generalizability over 68 countries. *European Journal of Personality, 34*, 180–202. https://doi.org/10.1002/per.2232.

Mestre, M. V., Samper, P., Frias, M. D., & Tur, A. M. (2009). Are women more empathetic than men? A longitudinal study in adolescence. *The Spanish Journal of Psychology, 12*, 76–83. https://doi.org/10.1017/S1138741600001499.

Muthén, L. K. (2008). *Re: Mplus discussion* [Online forum comment]. October 4, Retrieved from http://www.statmodel.com/discussion/messages/12/3609.html?1223146887.

L.K. Muthén, B.O. Muthén. (1998–013). Mplus user's guide, (7th ed.). Los Angeles, CA: Authors.

National Advisory Board on Research Ethics. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. Helsinki, Finland: Authors. Retrieved from http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf.

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. https://doi.org/10.1007/s10648-006-9029-9.

Pekrun, R. (2017). Emotion and achievement during adolescence. *Child Development Perspectives, 11*, 215–221. https://doi.org/10.1111/cdep.12237.

Pekrun, R., Goetz, T., & Frenzel, A. C. (2005). *Academic emotions questionnaire – Mathematics (AEQ-M) – User's manual*. Munich, Germany: University of Munich.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*, 1653–1670. https://doi.org/10.1111/cdev.12704.

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology, 117*, 166–185. https://doi.org/10.1037/pspp0000230.

Pesu, L. A., Aunola, K., Viljaranta, J., & Nurmi, J. E. (2016). The development of adolescents' self-concept of ability through grades 7-9 and the role of parental beliefs. *Frontline Learning Research, 4*, 92–109. https://doi.org/10.14786/flr.v4i3.249.

Plieninger, H., & Dickhäuser, O. (2015). The female fish is more responsive: Gender moderates the BFLPE in the domain of science. *Educational Psychology, 35*, 213–227. https://doi.org/10.1080/01443410.2013.814197.

Räsänen, P., & Leino, L. (2005). *KTLT – Laskutaidon testi luokka-asteille 7–9 [KTLT – The test for mathematical skills for grades 7–9]*. Jyväskylä, Finland: Niilo Mäki Instituutti.

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

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

Schleicher, A. (2019). *PISA 2018 Insights and interpretations*. Paris, France: Organisation for Economic Co-operation and Development [OECD].

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

Smith, R. H. (2000). Assimilative and contrastive emotional reactions to upward and downward social comparisons. In J. Suls, & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 173–200). New York, NY: Plenum.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28. https://doi.org/10.1006/jesp.1998.1373.

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry, 15*, 103–125. https://doi.org/10.1207/s15327965pli1502_01.

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

Turner, J. E., & Schallert, D. L. (2001). Expectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology, 98*, 320–329. https://doi.org/10.1037/0022-0663.93.2.320.

Van der Beek, J., Van der Ven, S., Kroesbergen, E., & Leseman, P. (2017). Self-concept mediates the relation between achievement and emotions in mathematics. *British Journal of Educational Psychology, 87*, 478–495. https://doi.org/10.1111/bjep.12160.

Webster, J. M., Duvall, J., Gaines, L. M., & Smith, R. H. (2003). The roles of praise and social comparison information in the experience of pride. *Journal of Social Psychology, 143*, 209–232. https://doi.org/10.1080/00224540309598441.

Wehrens, M. J. P. W., Buunk, A. P., Lubbers, M. J., Dijkstra, P., Kuyper, H., & van der Werf, G. P. C. (2010). The relationship between affective response to social comparison and academic performance in high school. *Contemporary Educational Psychology, 35*, 203–214. https://doi.org/10.1016/j.cedpsych.2010.01.001.

Weiner, B. (2014). The attribution approach to emotion and motivation: History, hypotheses, home runs, headaches/heartaches. *Emotion Review, 6*, 353–361. https://doi.org/10.1177/1754073914534502.

Widlund, A., Tuominen, H., Tapola, A., & Korhonen, J. (2020). Gendered pathways from academic performance, motivational beliefs, and school burnout to adolescents' educational and occupational aspirations. *Learning and Instruction, 66*. https://doi.org/10.1016/j.learninstruc.2019.101299. Advance online publication.