[AUTHORS' ACCEPTED AND PRE-PROOFED VERSION]

# THE HAPPY-FISH-LITTLE-POND EFFECT ON ENJOYMENT: GENERALIZABILITY ACROSS MULTIPLE DOMAINS AND COUNTRIES

Geetanjali Basarkod[1], Herbert W. Marsh[1], Jiesi Guo[1], Philip D. Parker[1], Theresa Dicke[1], & Reinhard Pekrun[1, 2, 3]

[1] Institute for Positive Psychology and Education, Australian Catholic University

[2] Department of Psychology, University of Essex

[3] Department of Psychology, Ludwig-Maximilian University Munich

**Corresponding Author:** Geetanjali Basarkod, Research Fellow, Institute for Positive Psychology and Education, Australian Catholic University, L9, 33 Berry Street, North Sydney, 2060, Australia; geetanjali.basarkod@acu.edu.au.

Accepted Version. Date of Acceptance: January 8, 2023

**Abstract**

Academic enjoyment is an important educational construct given that it benefits students' engagement, persistence, wellbeing, and mental health. In this study, we examine two factors that determine this crucial emotion, namely student- and class-level achievement. Past research has been restricted to single-country or single-domain examinations of secondary school students, limiting generalizability of findings. To bridge this gap, we utilize Trends in International Mathematics and Science Study and Progress in International Reading Literacy Study (TIMSS-PIRLS) 2011 data (*N*=180,084 4th-grade students, 37 countries). Our results provide robust evidence that student-level achievement positively predicts enjoyment in math, science, and reading, while the effects of class-level achievement is negative—the Happy-Fish-Little-Pond Effect. These results showed relative universality across the domains and countries examined.

*Keywords:* Achievement emotions, academic enjoyment, happy-fish-little-pond effect, multiple domains, cross-national generalizability

# THE HAPPY FISH LITTLE POND EFFECT ON ENJOYMENT: GENERALIZABILITY ACROSS MULTIPLE DOMAINS AND COUNTRIES

Achievement emotions are emotions that are related to achievement activities as well as their success and failure outcomes (Pekrun, 2006). These emotions are closely associated with one's thoughts, behaviors, expressions, arousal, and general day-to-day functioning (Fredrickson, 2001; Pekrun et al., 2017). They are a central component of identity, well-being, and health. Furthermore, achievement emotions have been shown to predict a range of important educational outcomes, such as students' persistence (Simon et al., 2015) and drop-out intentions (Respondek et al., 2017). They also predict students' twenty-first century skills such as communication, collaboration, critical thinking, and creativity (Camacho-Morles et al., 2021). Given that achievement emotions benefit students in numerous ways, it is important to examine the factors that promote these emotions. The goal of the present study is, therefore, to examine constructs that predict achievement emotions.

Past research has shown students' academic achievement to be a principal predictor of achievement emotions (for a meta-analysis, see Camacho-Morles et al., 2021). Achievement is a positive predictor of positive emotions such as enjoyment and a negative predictor of negative emotions such as test anxiety. In addition to the link between student achievement and emotions, however, a recent study also demonstrated the role of context in driving emotions (Pekrun et al., 2019). Specifically, Pekrun et al. showed that while student achievement positively impacts positive emotions and negatively impacts negative emotions, class-level achievement (i.e., the average achievement levels of one's peers) shows the opposite effect. We focus on these two constructs as potential predictors of emotion.

Traditionally, the vast majority of research on the link between achievement and

emotions focused on negative emotions, and specifically on test anxiety (Barroso et al., 2021; von der Embse et al., 2018). However, in recent years, research on positive emotions has been steadily increasing (e.g., Pekrun et al., 2017). These studies have tended to focus on emotions within a specific subject domain and single-country samples (Camacho-Morles et al. 2021). Thus, there is a relative lack of research using multiple domains and countries to examine positive emotions. In the current study, instead of studying a broad array of positive emotions, we take a deep-dive into the examination of just one emotion—enjoyment. We explore enjoyment across three subject domains (math, science, and reading) as well as across 37 countries, using the combined 2011 Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) dataset. We examine student- and class-level achievement as predictors of this emotion. In the next section, we describe the importance of enjoyment in education, as well as its links with student- and class-level achievement.

## Academic Enjoyment

### *Importance of Academic Enjoyment*

Academic enjoyment (henceforth "enjoyment") represents a key positive emotion that activates cognitive resources for learning, leads to adaptive school outcomes like student engagement (Pekrun, 2006), and sustains prolonged goal-oriented behaviors that optimize academic achievement (Camacho-Morles et al. 2021; Fredrickson, 2001; Pekrun et al., 2017). Positive activating emotions like enjoyment can help preserve cognitive resources, facilitate focus of one's attention on the learning task, support interest and intrinsic motivation, while also supporting use of deep learning strategies and promoting students' self-regulation of learning (Pekrun et al., 2017). Enjoyment has also been shown to be essential for flow experiences that

foster engagement and creative problem solving (Csikszentmihalyi, 2000).

### *Predictors of Academic Enjoyment*

To explain the predictors of enjoyment, we draw on control-value theory (CVT; Pekrun, 2006, 2018) as a conceptual framework. According to CVT, emotions are a consequence of cognitive appraisals of an individual's control over—as well as the subjective value they place on—achievement activities and outcomes. Control appraisals are the individual's perceptions of their ability to successfully perform actions and attain outcomes, while value appraisals are the individual's perceptions of how important those actions and outcomes are. Thus, successfully performing actions and attaining outcomes that are of value to the individual predicts increased positive emotions (e.g., enjoyment, hope) and reduced negative emotions (e.g., anxiety, hopelessness).

**Student-Level Achievement.** Enjoyment is said to be instigated if the achievement activity and the materials to which it relates are positively valued, and if the activity is perceived to be sufficiently controllable. In other words, a student is more likely to enjoy learning a particular subject when they judge themselves to be competent enough to master that subject, provided they find it valuable and interesting. Thus, high levels of control and value appraisals are proximal antecedents of enjoyment (Pekrun, 2017). Here, we focus more on control appraisals as they are more strongly linked with student achievement (Forsblom et al., 2022; Peixoto et al., 2017), which is a key predictor of concern in this study.

A student's perception of control in a given subject is influenced by their success or failure in that subject. If the student has a history of succeeding in math (e.g., has achieved high math test scores), they will have higher perceptions of control in math, which in turn will lead

them to experience positive emotions such as enjoyment (Forsblom et al., 2022). Therefore, a student's level of achievement will predict their level of enjoyment in that subject.

A recent meta-analysis on the link between student achievement and activity-related emotions, which included 57 independent samples for enjoyment, showed that the overall mean true-score correlation between achievement and enjoyment was .27 (95% CI = .23 to .30; Camacho-Morles et al., 2021). While the majority of studies showed a positive association between the two constructs, four studies reported negative correlations. On inspection of the original studies reporting these correlations, all four correlations were small and non-significant ($r$ = -.02 to -.17; Behrens et al., 2019; Chevrier et al., 2019; Muis et al. 2015 [sample 2]; Ranellucci et al., 2015). For instance, Behrens et al. (2019) found a non-significant negative correlation ($r$ = -.11) between undergraduate medical students' performance on a complex ward-based simulation and their enjoyment of this simulation. This lack of a significant correlation was likely due to ceiling effects and reduced variance for achievement (i.e., students had high levels of achievement: mean of 4.26 out of 5 [$SD$=0.43]).

While meta-analyses are important tools to make inferences about the generalizability of an effect, we note that such studies are limited in relation to the primary studies that they draw upon. First, and relevant to our current study, the majority of studies included in the above meta-analysis were conducted in the math domain, heavily weighting the findings towards math. There were also not enough primary studies that tested the association between student achievement and enjoyment in the domain of reading, to enable an examination of the size of this effect in the reading domain. Second, it relied heavily upon studies conducted in Australia, Canada, Germany, the UK, and the USA. This is an issue because much of the research in psychology tends to utilize samples from Western, educated, industrialized, rich, and democratic (WEIRD)

countries. Restricting research to such samples can limit the extent to which we can generalize findings to students from developing countries and varied cultures (see Marsh et al., 2020, for a detailed discussion on this issue). Thus, a complementary strategy is to utilize large-scale cross-national samples which comprise a broad range of developed as well as developing countries. Doing so would provide robust evidence regarding the generalizability of results and can supplement the knowledge gained from traditional meta-analyses.

He et al. (2019) conducted one of the few cross-national examinations of the association between achievement and enjoyment. They examined this association in the domain of science, using data from 21 developed countries that participated in both TIMSS 2015 and Programme for International Student Assessment (PISA) 2015. They found that while the association between science achievement and science enjoyment was positive in all 21 developed countries considered, and in both datasets, the size of the correlations varied between .10 and .40 (these numbers are approximate as the results were only plotted on a graph). However, in addition to only considering one subject and no developing countries, we note that the enjoyment measures used by He et al. (2019) from both the TIMSS and the PISA datasets included a scale comprised of items which were not all true measures of enjoyment. For instance, the TIMSS scale of "students like learning science" was used to measure enjoyment, which contained items such as "I wish I did not have to study Science", and "Science is boring". These items are not pure enjoyment items, and boredom is an emotion (or the lack thereof) that is different from enjoyment (Pekrun et al., 2010). Thus, while there is some past cross-national research on this topic, further research is required with clear measures of enjoyment across multiple domains.

**Group-Level Achievement.** The role of context in the development of emotions has also been shown to be important. Parents and teachers often believe in the benefits of educating

children in selective schools. It is assumed that students attending such schools will have a brighter and more successful future. However, evidence tends to point to the contrary; the average ability level of a class or school has been shown to negatively influence individual students' motivational outcomes, over and above the positive effects of individual-student achievement on these outcomes (Marsh et al., 2020; Pekrun et al., 2019).

Put simply, this effect implies that given the same level of individual achievement, students tend to have lower levels of motivational outcomes in high-ability schools than in low-ability schools (Parker et al., 2021). Based on CVT, Pekrun et al. hypothesized that, in a class of high achievers, a student's opportunities to be successful relative to others may be relatively low, all else being equal. This is in line with the social comparison theory underlying the negative effect of group-average achievement on self-concept (an individual's perceptions of their own ability)—the Big-Fish-Little-Pond Effect (Marsh & Parker, 1984). This effect posits that students evaluate their ability not only in relation to their own achievement levels, but also in relation to those around them. Such social comparisons help individuals to obtain plausible information for decision-making in daily life (e.g., Gruder, 1971). Thus, if a student is surrounded by peers who achieved top marks in a standardized test, that student is less likely to perceive their own performance as successful. In contrast, if a student is surrounded by peers who achieved low marks, the student will perceive their own score to be a success.

Past research has demonstrated that controlling for social comparison processes does indeed minimize the negative effect of group-average achievement on self-concept. For example, Huguet et al. (2009) found that the negative effect of class-average achievement on individual self-concept was eliminated after controlling for students' comparisons with their classmates. Similarly, Marsh et al., (2014) showed that accounting for an explicit measure of students'

comparisons with their peers reduced the negative effect of group-average achievement on reading and math self-concepts. Thus, past research shows that social comparison processes do underlie the negative effect of group-average achievement on individual student outcomes such as self-concept.

Based on social comparison processes and CVT, higher class-average achievement is likely to lead to students feeling less in control of the subject, thereby lowering positive emotions and increasing negative emotions (Pekrun et al., 2019). Indeed, Pekrun et al. found that while an individual's performance in their math class was predictive of increased positive emotions about math, the average achievement of that individuals' peers reduced the individual's positive emotions about math. They labeled the negative effect of group-average achievement on students' positive emotions (and positive effect on negative emotions) as the Happy Fish-Little-Pond Effect (HFLPE), after the well-known Big-Fish-Little-Pond Effect for self-concept (Marsh & Parker, 1984).

To date, the Pekrun et al. (2019) study is the only one to have examined the HFLPE on student enjoyment. While elucidating another predictor of students' academic emotions (i.e., class-average achievement), the Pekrun et al. (2019) study was limited in generalizability across subject domains and country of assessment. The samples considered were from secondary school math classes in Germany. Thus, it remains to be seen whether the HFLPE on academic emotions extends to subjects other than mathematics, to samples from countries other than Germany, and to primary school students. In recent years, the psychology literature is undergoing a replication crisis; many of the most salient findings in the literature have not been confirmed in subsequent studies and the generalizability of these results is, therefore, questionable. The current study aims to ascertain whether the HFLPE stands up to scrutiny in relation to replication across domains

and generalizability across countries.

If evidence for a given theory—the HFLPE in this case—is found in multiple countries and cultural contexts, it can be thought of as a universal principle with pan-human validity (Segall et al., 1998; also see Marsh et al., 2020). CVT assumes that general functional mechanisms of human emotions are universal, but that the specific contents of emotions as well as the specific values of process parameters (such as the intensity of the emotions) may be specific to different individuals or groups of individuals (Pekrun, 2006, 2018). This "relative universality" implies that the functional relations between achievement (both student- and class-level) and enjoyment are expected to be universal, but that levels of enjoyment may differ across domains and cultural contexts (see Pekrun, 2018, for a summary of supporting evidence). However, no study has yet been conducted regarding the relative universality of the effect of student- and group-level achievement on positive emotions such as enjoyment.

**Gender, Age, and Socio-Economic Status.** We provide a brief overview of the relationship of gender, age, and socio-economic status (SES) with enjoyment, as these variables are commonly controlled for in studies examining compositional effects (Dicke et al., 2018)—and in research on enjoyment. For instance, in the only known past study of the HFLPE mentioned above, Pekrun et al. (2019) controlled for gender and SES (they also controlled for grade level as the sample consisted of students from Grades 5-10). They showed that boys reported higher levels of math enjoyment as compared to girls, but that there were no significant effects of SES on math enjoyment. Similar results showing lower math enjoyment in girls were found by Frenzel et al. (2007) and Goetz et al. (2008). Gunderson et al. (2017) compared the math enjoyment levels of groups of students in $1^{st}/2^{nd}$ grade, $5^{th}/6^{th}$ grade, $10^{th}/11^{th}$ grade, and college. Their results demonstrated that math enjoyment significantly differed between age

groups, with older students showing lower levels of math enjoyment compared to younger students.

In a study examining demographic characteristics of students who displayed high levels of reading enjoyment, Rogiers et al. (2020) showed that high-school aged Flemish girls had higher levels of reading enjoyment as compared to boys. Such an effect of girls reporting higher reading enjoyment has been shown in other studies (e.g., De Naeghel et al., 2014; Goetz, 2004; Goetz et al., 2008; Smith et al., 2012). Rogiers et al. (2020) further demonstrated that students from higher SES backgrounds tended to report higher levels of reading enjoyment, while Smith et al. 2012) demonstrated that reading enjoyment declined between the ages of 8 and 12.

In terms of science enjoyment, Akpınar et al. (2009) showed that there were no gender differences for primary-aged school children in Turkey, while Addabbo et al. (2016) showed boys to have higher levels of science enjoyment compared to girls in a sample of Italian 15-year-old students. Wan and Lee (2017) showed similar results as Addabbo et al. (2016), where boys showed greater levels of science enjoyment in a sample of ninth graders in Hong Kong. In relation to SES, Ainley and Ainley (2011) provided correlations between SES and science enjoyment for four PISA countries (Colombia, USA, Estonia, and Sweden). Their results demonstrated a positive correlation for all countries except for Colombia (where a very small negative correlation was noticed). Murphy and Beggs (2003) reported that students' levels of science enjoyment tended to decrease with age, with students aged 8-9 reporting higher levels of enjoyment than those aged 10-11.

Thus, while there are some inconsistencies, past research tends to suggest that boys have higher levels of math enjoyment, lower levels of reading enjoyment, and similar or higher levels

of science enjoyment compared to girls. Students who come from higher socio-economic backgrounds tend to report similar levels of math enjoyment, but higher levels of reading and science enjoyment compared to students from lower SES backgrounds. Older students tend to report lower levels of enjoyment in all three domains. Given that some differences in enjoyment are noticed in relation to these three demographic variables, it is important to account for them in models that predict enjoyment.

## The Present Study

The present study examines the effect of student- and class-level achievement on student enjoyment, while controlling for gender, age, and SES. We used the TIMSS-PIRLS 2011 cross-national data from primary-school aged (Grade 4) students in 37 different countries, responding about their mathematics, science, and reading classes. Importantly, the same group of fourth-grade students responded to both assessments. This provides an unprecedented opportunity for researchers to compare fourth graders across three fundamental curricular areas (mathematics, science, and reading) and 37 countries. In addition, the 37 included countries were a mixture of developed and developing countries, providing a better basis of testing generalizability. We use these data in the current study.

The present study provides a novel extension of past research by (i) examining the generalizability of the effects of student- and class-level achievement on student-level enjoyment across math, science, and reading domains; (ii) exploring the cross-national generalizability of these effects across 37 countries; and (iii) studying these effects for a younger sample than has been done previously as past research on the link between achievement and emotions is largely limited to secondary school students. The study aims to answer the following two research questions.

**RQ1.** Does student-level achievement positively predict student-level enjoyment across subject domains and countries?

Based on prior research, we expect student-level achievement to positively predict student-level enjoyment across subject domains (**Hypothesis 1**) and countries (**Hypothesis 2**). We leave as an exploratory question if the size of the effects will differ across domains and countries.

**RQ2.** Does class-level achievement negatively predict student-level enjoyment across subject domains and countries?

We expect class-level achievement to negatively predict student-level enjoyment across subject domains (**Hypothesis 3**) and countries (**Hypothesis 4**). While we expect the direction of results to be consistent across domains and countries, we again leave as an exploratory question whether the size of the effects will vary from domain-to-domain and from country-to-country.

## Methods

### Data and Participants

TIMSS is an international assessment of mathematics and science competence in nationally representative samples of fourth- and eighth-grade students. PIRLS is an international assessment of reading comprehension in nationally representative samples of fourth-grade students. In 2011, 37 countries administered the TIMSS and PIRLS assessments to the same samples of fourth-grade students. Though the two surveys were administered independently, the combined dataset includes adjustments to the achievement tests and contextual questionnaires to make them appropriate for use in the combined dataset. These adjustments include (i) rescaling of item parameters for the achievement scales such that the correlation structure between the

three domains of achievement (i.e., reading, math, and science) is preserved; (ii) rescaling of the contextual questionnaires; and (iii) calculation of appropriate sample weights based only on the data of students who participated in both the TIMSS 2011 and PIRLS 2011 surveys.

Both datasets employed a stratified two-stage cluster sample design; in the first stage, schools were randomly selected from the population of schools based on stratification, and in the second stage, classes were sampled at the target level (fourth grade) within the selected schools. Thus, students participating in these datasets largely belonged to intact classes. The final sample size used in our analyses included 180,084 students (49.26% female; Age $M$ = 10.92, $SD$ = 0.82) from 8,372 classes and 6,196 schools. The average sample size in each country was 4,867 (range = 2,901 to 14,449; see Supplementary Materials Section 1, Table S1 for the sample size for each country). We utilize class as our Level 2 cluster variable and country as the Level 3 cluster variable (see Data Analysis section for more detail).

## Measures

### *Enjoyment*

Student enjoyment in each subject domain was measured by a single item, “I enjoy learning mathematics”, “I enjoy learning science”, and “I enjoy reading”. Students were asked to rate their agreement with these statements on a 4-point Likert-type scale from 1 (*Agree a lot*) to 4 (*Disagree a lot*). Items were reverse-scored such that higher scores indicated greater enjoyment. Past research has shown single items are sufficiently reliable to measure emotions, and can often be as reliable and valid as their multi-item counterparts (Allen et al., 2022). This has also been shown specifically for emotions (Gogol et al., 2014). In addition, in our case a particularly relevant benefit for the use of single-item measures is that the use of multiple items can contaminate the construct being studied (Bergkvist & Rossiter, 2007). For example, and as

mentioned in the introduction, some past research on enjoyment has included items that assess boredom—which is another emotion (e.g., He et al., 2019). Thus, our single-item measure of enjoyment is clean and reliable.

### *Achievement*

*Student achievement* was measured through a combination of multiple-choice and free-text responses in each domain. An example item in the math domain was, "Al wanted to find how much his cat weighed. He weighed himself and noted that the scale read 57kg. He then stepped on the scale holding his cat and found that it read 62kg. What was the weight of the cat in kilograms?" An example item in the science domain was, "When you blow into water using a straw, bubbles are formed and rise to the top. Why do the bubbles rise in water?". For reading, students were given a variety of texts which they had to read and answer questions about. For instance, they had to read a story called "An Unbelievable Night" which followed the events that occurred when the 10-year-old protagonist Anina woke in the middle of the night to go to the bathroom. The questions asked about this passage included, "Name one thing Anina had difficulty explaining to her parents."

Students were administered one of 13 PIRLS 2011 booklets, each with two reading passages and their accompanying items, as well as one of 14 TIMSS 2011 assessment booklets, each with a series of mathematics and science items. This was done to keep student burden to a minimum. Given that each student only responded to a subset of items for each subject, five plausible values were estimated for each student for each subject, via models using item response theory. These plausible values were used to infer students' abilities based on their performance on the items (see Data Analysis section for information on how these plausible values were handled).

*Class-average achievement* was calculated as the average achievement of students in a class (after student-level achievement was grand-mean centered). This average was calculated separately for each plausible value within each subject, and then handled the same way as student-level achievement (see Data Analysis section).

### *Covariates*

In all our models, we controlled for three covariates commonly included in research on compositional effects (Dicke et al., 2018). *Socio-economic status* (SES) was assessed with the Home Resources for Learning Scale. This scale was provided in the home background datafile, which contained responses from the students' parents or guardians. The SES score is a composite score based on information about the educational and occupational background of parents, number of books at home, internet access, and whether the student has their own room at home (more information available from the TIMSS-PIRLS website at https://timss.bc.edu/methods/pdf/P11_R_Scales_HRL.pdf). *Sex* was coded as female = 1 and male = 2. *Age* was assessed using the month and year students were born in, converted to a continuous variable of years.

## Data Analysis

For the purpose of this study, we merged student and home data using the International Database Analyzer of the International Association for the Evaluation of Educational Achievement. Data were merged at the student level, with each row in the dataset corresponding to one student. In this study, we used data from all 37 countries, for every student who belonged to a class with at least 10 students (to calculate class-average achievement reliably; for a similar procedure, see Pekrun et al., 2019). All analyses were run separately for each subject. Analyses were conducted in R (Version 3.6.2; R Core Team, 2020), with code publicly available on the

Open Science Framework (https://osf.io/kb29h/).

The number and percentage of missing cases per variable used in this study are shown in the Supplementary Materials (Section 2, Table S2). Missingness was dealt with through multiple imputations using the package *Amelia II* (Honaker et al., 2011) with Class ID as the clustering variable. Five imputed datasets were created, retaining all variables and covariates mentioned above. We assigned one plausible value for math, science, and reading achievement at the student (Level 1) and class levels (Level 2) to each of the five imputed datasets. All analyses were conducted five times, using each imputed dataset once, and results were combined using Rubin's (1987) approach.

Three-level multilevel models were conducted using *lme4* (Bates et al., 2015)—one for each subject domain. In the multilevel models, all student variables were at the student level (Level 1), except class-average achievement which was a class-level (Level 2) variable. Random intercepts for class and country were included to account for the clustering of students within classes and classes within countries. Non-categorical Level 1 variables were standardized across the whole sample (i.e., grand-mean centered). Class-average achievement was calculated after the corresponding Level 1 achievement variables had been standardized. All models were weighted using the final survey weight for students, normalized for each country (i.e., the sum of the weights in each country was equal to the sample size in that country).

We interpret the size of the fixed effects from these multilevel models based on recommendations by Else-Quest et al. (2010); effect sizes of less than 0.10 can be considered trivial when sample sizes are large (as estimates of even 0.01 may be statistically significant) and estimates of more than 0.10 can be considered non-trivial. Using our three-level multilevel

models, we examine the cross-national generalizability of the effects of student- and class-level achievement on enjoyment in three ways: (i) the size of the random variance components; (ii) the relationship between the standard deviation of the random effects (i.e., square root of the random variance components) and their corresponding fixed effects; and (iii) the cross-level interactions between country-level moderators and the fixed effects.

There would be support for generalizability if (i) residual variance components are equal to or less than .01 (Marsh et al., 2020); (ii) the standard deviation of country-to-country variation (square root of the random variance component) is less than half of the corresponding fixed-effect estimate; Marsh, 2016); and (iii) statistically significant interaction effects are smaller than .05 (Marsh et al., 2020; also see Guo et al., 2021, 2022).

To graphically explore country-to-country variation in the estimates of student- and class-achievement on enjoyment, we created forest plots. We estimated two-level models (Level 1: student; Level 2: class) in each country independently (controlling for the covariates). We then extracted estimates and confidence intervals for Level 1 and Level 2 achievement predicting enjoyment, and graphed them in forest plots (see Figures 1 and 2).

## Results

### Descriptive Statistics and Correlations

We provide descriptive statistics for all variables included in this study in the Supplementary Materials (Section 2, Table S2). Table 1 presents zero-order correlations across the whole sample, using group-mean centering at the country level. We interpret the size of correlations based on Funder and Ozer's (2019) recommendations (i.e., *r*s of over .05, .10, .20, .30, and .40 are interpreted as very small, small, medium, large, and very large, respectively). The correlational analyses indicated positive correlations between student-level achievement and

enjoyment, for all three domains. We note, however, that while statistically significant (given the large sample size), the magnitude of these correlations in the domains of math and science was very small ($r$ = .08 and .06, respectively). The correlation was small for the domain of reading ($r$ = .19). Reading enjoyment had the biggest correlations with all three domains of achievement (i.e., achievement in both matched and non-matched subject domain), compared to the correlations for math and science enjoyment. The correlations between the achievement scores in the three domains were very large ($r$ = .78 to .85), while the correlations between the three domains of enjoyment were small to medium ($r$ = .15 to .26).

In terms of correlations between our covariates and enjoyment, most of these were not substantial in magnitude (i.e., $r < .05$). The correlation between gender and math enjoyment was very small ($r$ = .05), and the correlations between SES and reading enjoyment ($r$ = .10) and gender and reading enjoyment ($r$ = -.18) were small. These correlations imply that children with higher levels of SES showed higher levels of reading enjoyment, while girls showed lower levels of math enjoyment and higher levels of reading enjoyment compared to boys.

**Table 1**

*Correlations Between Level 1 Variables*

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Math Enjoyment | – | | | | | | | | |
| 2 | Science Enjoyment | .15 | – | | | | | | | |
| 3 | Reading Enjoyment | .18 | .26 | – | | | | | | |
| 4 | Math Achievement | .08 | .02 | .12 | – | | | | | |
| 5 | Science Achievement | .04 | .06 | .15 | .84 | – | | | | |
| 6 | Reading Achievement | .02 | .06 | .19 | .78 | .85 | – | | | |
| 7 | SES | -.02 | .02 | .10 | .35 | .37 | .37 | – | | |
| 8 | Gender | .05 | -.01 | -.18 | .01 | -.01 | -.12 | -.01 | – | |
| 9 | Age | -.01 | -.02 | -.04 | -.05 | -.05 | -.06 | -.12 | .05 | – |

*Note.* Correlations were calculated using all 5 imputed datasets and combined using Rubin's rules.

All correlations are statistically significant ($p < .05$). Gender was coded as female = 1 and male = 2.

**Table 2**

*Results From Multilevel Models Predicting Enjoyment*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.212* | .047 | .008 | .039 | .392* | .049 |
| Student-Level Ach | .169* | .014 | .098* | .012 | .241* | .019 |
| Class-Average Ach | -.173* | .020 | -.124* | .017 | -.070* | .024 |
| SES | -.030* | .004 | .016* | .003 | .054* | .003 |
| Gender | .110* | .005 | -.008 | .005 | -.323* | .005 |
| Age | -.008* | .004 | -.014* | .004 | -.019* | .003 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .243 | | .253 | | .239 | |
| Student Ach \| Class | .097 | | .120 | | .115 | |
| Intercept \| Country | .279 | | .230 | | .290 | |
| Student Ach \| Country | .080 | | .066 | | .109 | |
| Class Ach \| Country | .103 | | .076 | | .121 | |
| Residuals | .914 | | .926 | | .913 | |

*Note.* Ach = achievement.
Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

## Multilevel Analyses

Table 2 presents results from the three-level models predicting math, science, and reading enjoyment, respectively. These models include student- and class-level achievement as predictors, along with the covariates of SES, gender, and age.

### *Fixed Effects: Examination Across Domains*

The fixed effect estimates show that domain-specific student-level achievement positively predicted student-level enjoyment in each of the three subject domains, confirming Hypothesis 1. This predictive effect was largest for reading enjoyment (β = .241), followed by math (β = .169), and smallest for science enjoyment (β = .098). As hypothesized, the direction generalized over domains even though the size of the effects varied. Class-level achievement negatively predicted student-level enjoyment in all three domains, confirming Hypothesis 3. There was a larger effect for math enjoyment (β = -.173), followed by science (β = -.124) and reading (β = -.070). While the negative effect for reading enjoyment was significant, we note that the size of this effect may be considered trivial as it was smaller than 0.10 (Else-Quest et al., 2010). Overall, results showed that while students with higher achievement levels showed greater enjoyment levels in that domain, students in classes with higher class-average achievement showed lower levels of enjoyment in that domain. This confirms the HFLPE for enjoyment in our data, across all three domains (though the effect size in the reading domain may be considered trivial).

In terms of the fixed effects of the covariates, SES had a negative effect on math enjoyment and a positive effect on science and reading enjoyment. Though all three effects were significant, they were all less than .10 and can be considered to be trivial (Else-Quest et al., 2010). Similarly, though there were significant negative effects of age on enjoyment in all three

domains, all of these effects were smaller than .10. There was a positive effect of gender on math enjoyment (β = .110) and a negative effect on reading enjoyment (β = -.323). That is, girls had lower math enjoyment, but higher reading enjoyment, compared to boys.

### *Random Effects: Examination Across Countries*

**Random Variance Components.** Table 2 shows that the standard deviation of random effects for student-level achievement at the class and country levels were between .097-.120 and .066-.109, respectively. These correspond to random variance components (SD squared) of .009-.014 between classes and .004-.012 between countries. Similarly, the standard deviation of random effects for class-average achievement at the country level was between .076-.121, corresponding to random variance components of .006-.015. Following recommendations by Marsh (2016), the majority of these random variances can be considered trivially small (≤ .010) to small. Thus, going by this metric, our results are largely generalizable across countries.

**Standard Deviation of Random Effects.** Following recommendations by Marsh (2016), results indicate that the effect of student achievement on enjoyment is generalizable across countries for the domains of math and reading. That is, the standard deviations of the random effects at the country level for math (*SD* = .080) and reading enjoyment (*SD* = .109) were less than half the corresponding fixed estimates (math: β = .169; reading: β = .241). While there was a positive predictive effect of science achievement on science enjoyment, the findings suggest that this effect is likely to be less consistent across countries (β = .098, *SD* = .066). In terms of the effect of class-average achievement, results show that there was variation across countries, in relation to the size of the fixed effects from the three-level models. For all three domains, the standard deviation was larger than half the size of the corresponding fixed effect (math: β = -.173, *SD* = .103; science: β = -.124, *SD* = .076; reading: β = -.070, *SD* = .121). Thus, while there

was an overall trend of an HFLPE for all three domains, the size of the effect varied from country to country.

**Interactions with Country-Level Variables.** We consider the moderating effects of three country-level variables on the effects of student- and class-level achievement on enjoyment: (i) Human Development Index (HDI); (ii) the Education Index subcomponent of HDI; and (iii) Gross Domestic Product (GDP) per capita. The majority of these interaction effects were either non-significant or below the threshold for meaningful interaction effect sizes (i.e., $\beta < .05$; Marsh et al., 2020). The only exceptions were that the positive effect of student-level achievement on reading enjoyment was stronger in countries with higher HDIs ($\beta = .066$) and Education Index ($\beta = .076$). Thus, following Marsh et al. (2020), the effects of student- and class-level achievement on enjoyment are relatively robust to the moderators examined here. Detailed results are presented in our Supplementary Materials (Section 3, Tables S3-S5).

### *Country-Specific Estimates: Examination Across Countries*

Figure 1 presents forest plots for the effects of student-level achievement on enjoyment in each of the three subject domains, in each of the participating countries. Figure 2 presents parallel forest plots for the effect of class-average achievement. The exact estimates for each country are presented in the Supplementary Materials (Section 4, Tables S6-S7).

**Figure 1**

*Country-Wise Estimates and 95% Confidence Intervals of the Effect of Student-Level Achievement on Enjoyment by Domain*

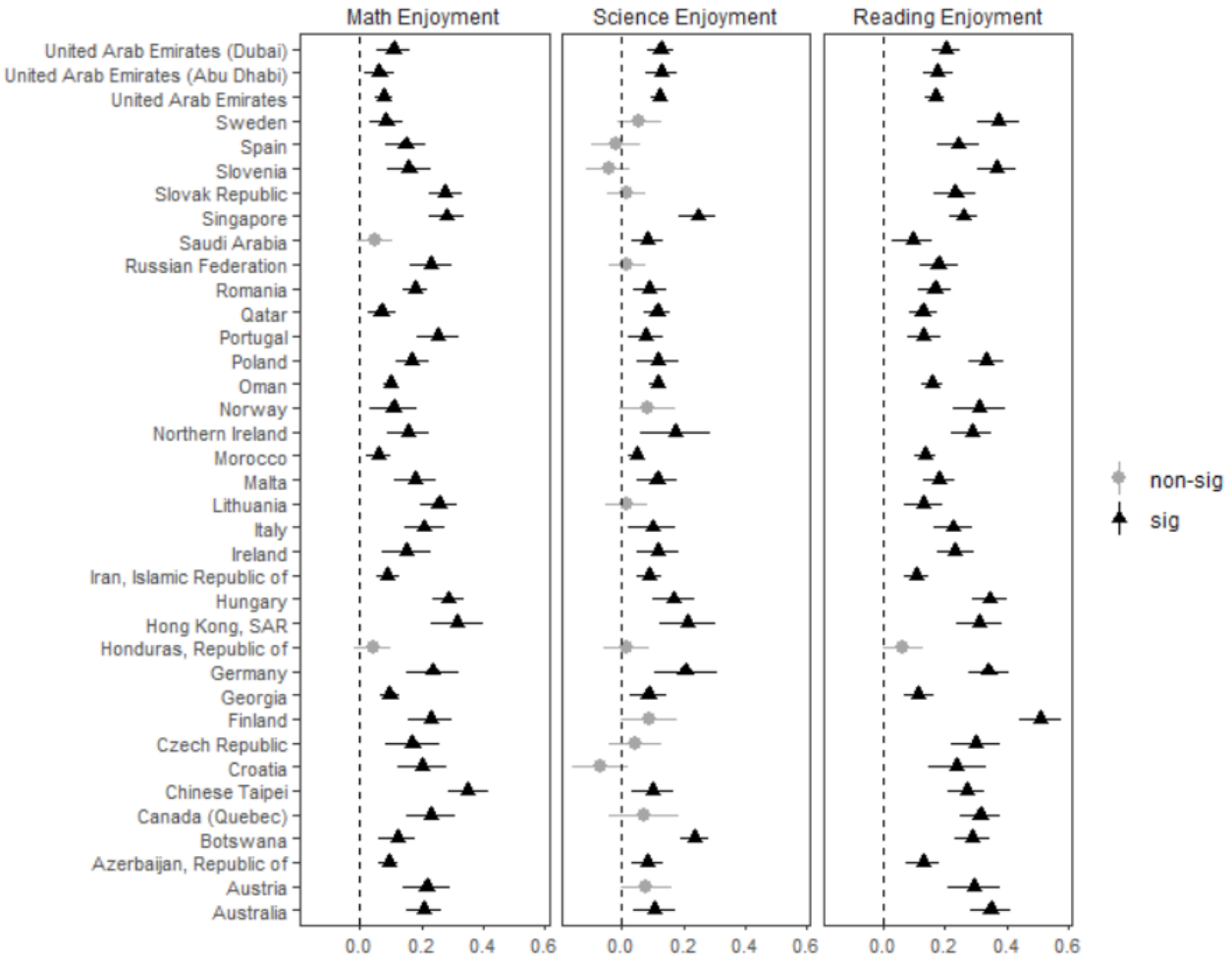
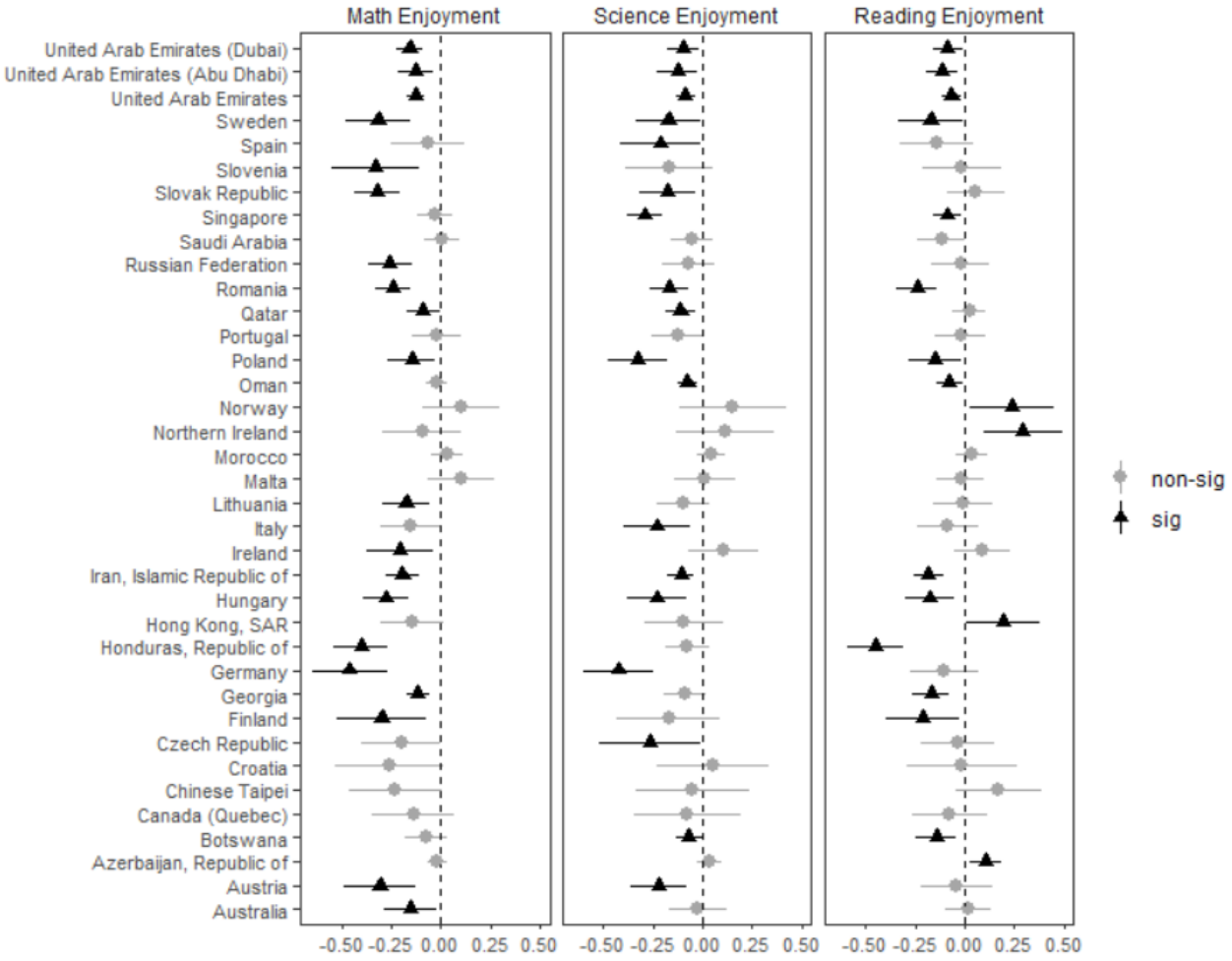


**Figure 2**

*Country-Wise Estimates and 95% Confidence Intervals of the Effect of Class-Average Achievement on Enjoyment by Domain*



**Student-Level Effects.** In terms of student-level achievement, all significant effects were positive in all three domains. The estimate was non-significant for two countries in the math domain, for thirteen countries in the science domain (with three of these non-significant effects in the negative direction), and for one country in the reading domain. Thus, these forest plots show that the effect of student achievement on enjoyment was largely consistently positive across countries, for all three domains (even though the results from the three-level models hinted at variation in the size of effects). In addition, the average effects from these models were similar to those from the multilevel models (math: $\beta = .168$, *SE* = .013; science: $\beta = .092$, *SE* = .012; reading: $\beta = .235$, *SE* = .016). This confirms our Hypothesis 2; the effect of student achievement on enjoyment is positive across countries with variation in the size of this effect.

In terms of the variation in the size of effects, the smallest effects for the math domain tended to be in Middle Eastern countries such as Morocco, Qatar, and UAE. The largest effects were in the Asian countries of Chinese Taipei, Hong Kong, and Singapore, as well as European countries like the Hungary and Slovak Republic. For the domain of science, the European countries of Croatia, Slovenia, and Spain had non-significant negative effects, with the effect for Slovak Republic, Honduras, Lithuania, and Russia being close to zero. Singapore had the largest effect, with Botswana, Hong Kong, Germany, and Northern Island following close behind. For reading enjoyment, Honduras had a non-significant effect. The smallest significant effects were seen for Saudi Arabia, Iran, Georgia, and Qatar. The European country of Finland had the largest effect, followed by Sweden and Slovenia.

**Class-Average Effects.** For class-level achievement, for both math and science enjoyment, all significant effects were negative, and non-significant effects were mostly negative. While the pattern was largely similar for reading enjoyment, four effects were

significantly positive. While the results from the three-level models demonstrated some variation in the size of the effects, the forest plots demonstrated that this variation was largely in the size of the effect rather than in its direction, especially for the domains of math and science. Again, the average effects from these models were similar to those from the multilevel models (math: $\beta$ = -.158, *SE* = .022; science: $\beta$ = -.107, *SE* = .020; reading: $\beta$ = -.048, *SE* = .023). This largely confirms our Hypothesis 4. That is, the effects of class-average achievement on enjoyment were mostly negative and the size of the negative effect varied from country-to-country.

In terms of the variation in the direction and size of effects, as mentioned above, Germany showed the largest HFLPE in the domain of math, with Honduras, Slovenia, and the Slovak Republic following close behind. Georgia had the smallest HFLPE, with the UAE, Poland, and Australia showing slightly larger effects. Croatia showed the largest non-significant HFLPE, while Norway showed the largest positive non-significant effect. Similar patterns were noticed with science enjoyment, where Germany again showed the largest HFLPE, though the next highest HFLPEs were seen in Poland, Singapore, and Czech Republic. Botswana had the smallest HFLPE with Oman, UAE, and Iran showing slightly larger effects. Finland had the largest non-significant HFLPE, while Norway showed the largest non-significant positive effect. When it came to reading, Honduras showed the largest HFLPE, followed by Romania, and Finland. UAE showed the smallest HFLPE, with Oman and Singapore showing slightly larger HFLPEs. As mentioned, four countries showed positive effects of class-average achievement on reading enjoyment. Northern Ireland had the largest of these effects, followed by Norway, Hong Kong, and the Republic of Azerbaijan.

The variations in the effects of both student- and class-level achievement do not, at face-value, show a discernable pattern that is consistent across subject domains. That said, most

countries did show consistency in the fact that, even if non-significant, the effects were in the same direction. For instance, for the effect of class-average achievement, Germany had the largest HFLPEs for math and science. While the effect for reading in Germany was non-significant, it was still negative. Similarly, Norway showed positive effects of class-average achievement in all three domains, though only the effect for reading was significant. Thus, it is possible that similar mechanisms are acting in each country, though to different levels for each domain. We speculate that the differences between math/science and reading are likely because of the difference in wording for the domain of reading (i.e., that it measures enjoyment in reading without being specific to the school context, while the measures for math and science were about enjoying learning those subjects). Further research is required to examine the processes within specific countries, such as Germany and Norway, to examine the mechanisms in detail within these countries across domains.

## Discussion

Academic enjoyment is an important educational outcome. A greater understanding of the predictors of academic enjoyment would enable the development of interventions to positively impact enjoyment, which would in turn positively promote increased attainment and other academic outcomes (e.g., Luo et al., 2016). The primary aim of this study was, therefore, to examine two main predictors of academic enjoyment—student-level achievement and class-level achievement—as well as the relative universality of these effects across three subject domains and 37 countries. Our results demonstrated that while student-level achievement positively predicted student enjoyment, the effect of class-level achievement was negative. The pattern of these results was consistent across the domains of math, science, and reading, and largely consistent across the 37 countries included in the TIMSS-PIRLS 2011 sample.

Our study extended previous research in two important ways. First, there has been no prior research exploring the multi-domain and cross-national generalizability of the predictive effect of student-level achievement on academic enjoyment. Previous research in this field has largely focused on the negative emotion of anxiety and has been limited to a single subject domain or a single country. Indeed, in the PISA publications, the OECD provides detailed reports about test anxiety and math anxiety (e.g., OECD, 2018). Thus, we extend prior research by exploring the association between achievement and the positive emotion of enjoyment. Second, there has only been one previous study positing and testing the HFLPE on enjoyment; this study was limited to the domain of math in samples of students from Germany (Pekrun et al., 2019). Thus, our study extends prior research by examining the HFLPE for the domains of science and reading in addition to mathematics. Moreover, we included samples from 36 countries in addition to Germany, and demonstrated how the effects of both student and class-level achievement on enjoyment are relatively robust across countries.

**Student-Level Achievement and Enjoyment**

Past research has shown that students' enjoyment for a given subject is usually positively related to achievement within the same domain (Camacho-Morles et al., 2021). Much of the past research on this topic has considered enjoyment (and, indeed, emotions in general) as a predictor of achievement. Students who enjoy a particular subject tend to be more engaged with the materials and channel their cognitive efforts to the activity at hand, thereby performing better. However, recent research has also shown that prior achievement leads to enjoyment (e.g., Forsblom et al., 2022; Pekrun et al., 2017). While cross-sectional, our study is consistent with such findings, as we show that achievement positively predicts enjoyment in matched subject domains. Thus, the better a student performs at a given subject, the more they tend to enjoy it.

We note that the student-level effects shown in our study are generally smaller than the mean effects reported in the meta-analysis by Camacho-Morles et al. (2021). Indeed, Camacho-Morles et al. showed that the effect was stronger in secondary school samples than primary school samples. As the students in our sample are mostly in Grade 4, this may explain the smaller effects of student achievement on enjoyment seen in our study. This difference between primary and secondary school samples may be because repeated feedback about achievement over a longer period creates a stronger link between performance and emotions. Students in secondary school have longer to gain knowledge about their level of achievement than students in primary school and, thus, may demonstrate stronger associations between their achievement and enjoyment levels. Our study provides an important addition to the literature on achievement emotion by extending past research to primary school students.

**Class-Level Achievement and Enjoyment**

Our results showed that, in general, students would have lower levels of enjoyment in a class with higher class-average achievement than if they were in a class with lower class-average achievement. This finding can be understood in terms of CVT, in combination with social comparison processes that usually operate for the Big-Fish-Little-Pond effect (Marsh & Parker, 1984). We evaluate our ability not only in relation to our own achievement levels, but also in relation to those around us. Our peers' achievement acts as a frame of reference for the judgements we make about our own ability. Such a social comparison process indicates that a student would evaluate their ability to be greater when they perform better than those around them. If they achieved top marks when their peers achieved average marks, they would feel more successful and more in control of the subject material, than if their peers also achieved top marks. Thus, the higher the class-average achievement, the lower the student's perception of

their ability. This in turn implies lower levels of control over the subject material, which is associated with lower levels of enjoyment in relation to the subject. Indeed, in a quick test of the CVT proposition, we found that when we included variables related to control and value into our model, the effects of both student- and class-level achievement on enjoyment were substantially reduced (see Supplementary Materials Section 5 for details regarding these analyses and the corresponding results [Table S8]).

Our results are again smaller in magnitude compared to the only previous study on the HFLPE (Pekrun et al., 2019). Pekrun et al.'s study demonstrated that the effect of class-average achievement on enjoyment was -.40. However, there are two important distinctions between our study and the Pekrun et al. (2019) study that might help explain the difference. The Pekrun et al. study was conducted only in the domain of math and only in the country of Germany. Therefore, the comparable beta estimate in our study is the one for the domain of math in Germany. This estimate was -.47, which is largely comparable to that reported in the Pekrun et al. study. Our study builds on this past research to provide a better estimate of the generalizable effect of group-average achievement on student enjoyment.

We note here that when examining the effect of group-average achievement on student outcomes (especially self-concept), past research has sometimes included the cross-level interactions between student- and group-level achievement to examine whether individual student achievement moderates the effects of group-level achievement on the outcome (e.g., Seaton et al., 2010). We, therefore, examined such an interaction with our data as well (Supplementary Materials Section 6, Table S9). The results showed a positive interaction effect for math and reading, but a negative interaction effect for science. However, we note that while these interaction effects were significant, they were very small ($< |.027|$) and likely not of

practical importance (and also did not change the direction of main effects). As our sample size was large, even very small effects that are not meaningful can be statistically significant. A growing body of empirical research in self-concept (see Marsh et al. 2021) supports this interpretation in that the interactions between individual- and group-level achievement are consistently small or non-significant, and not even consistent in direction.

### Gender Differences in Enjoyment

We also found gender differences in enjoyment, particularly for the domain of reading (the differences for age and SES were either non-significant or trivial in size). Specifically, girls showed higher levels of reading enjoyment compared to boys, while they also had slightly lower levels of math enjoyment and showed no differences in relation to science enjoyment. This pattern was found both through zero-order correlations and multilevel models that accounted for the effects of student- and class-level achievement, SES, and age. These differences are in line with previous research showing that girls show greater reading enjoyment compared to boys (Brozo et al., 2014), but lower levels of math enjoyment (Frenzel et al., 2007), even though the levels of math achievement were comparable. It is possible that girls showed lower levels of math enjoyment even when controlling for their math achievement because of lower competence beliefs. Indeed, Frenzel et al. (2007) provided some preliminary evidence to suggest that girls' lower competence beliefs and perceptions of the value of math helped explain the gender differences in math enjoyment. Future research is required to examine the reasons for gender differences in enjoyment of different subject domains, particularly in the domain of reading.

### Generalizability of Results

As mentioned in the Introduction, Pekrun (2006) stated that the processes underlying the formation of academic emotions and their consequences operate according to relative

universality. That is, the processes themselves are the same for everyone (i.e., universal). However, the degree to which these processes operate varies from person to person, depending on their context (i.e., is relative). The results of this study are largely consistent with this proposition. Specifically, the effect of student-level achievement on enjoyment was largely positive across domains and countries, but the size of this effect varied. Similarly, the effect of class-average achievement on student enjoyment was mostly negative across domains and countries (with some exceptions), but the specific size of this effect varied. Thus, our results generally support the relative universality principle.

### *Examination Across Domains*

The effect of student-level achievement on enjoyment was the strongest in the reading domain, followed by the math and science domains. In contrast, the effect of class-average achievement was the strongest in the math domain, followed by the science and reading domains. The contrasting results for the reading domain are likely due to the way enjoyment was measured. For the math and science domains, the item asked students whether they enjoyed learning math/science. In contrast, in the reading domain the item simply asked whether students "enjoyed reading". Thus, for the reading domain, the enjoyment item was not specific to the school context, but rather assessed reading enjoyment in general.

In terms of the effects of student-level achievement, the stronger effect for the reading domain may be because students who are good at reading likely also read outside of school for enjoyment, thus strengthening the link between reading competency and enjoyment through frequent positive reading experiences. For class-average achievement, the lack of focus on the school context reduces the likelihood of the social comparison process operating in the students' judgment of their enjoyment (Marsh et al., 2019). Thus, it is less likely that the achievement of

peers played a role in students' responses to the enjoyment items. In contrast, as the math and science items were specific to the learning context, it is more likely that the achievement of peers played a role here. Thus, the frame-of-reference effect of class-average achievement was less relevant for the reading enjoyment item than for the math and science items.

Past research has suggested that relations of achievement and enjoyment tend to be stronger in math compared to other subjects such as reading and science (Camacho-Morles et al., 2021). Our research is in opposition to these findings. This may be due to the lack of education-context information in the items, as mentioned above. It must also be noted that the Camacho-Morles et al. meta-analysis was heavily weighted towards math, and therefore, had lesser information about the association between achievement and enjoyment in the domains of science and especially reading. A consequence is the smaller aggregate samples in non-math domains, and thus a greater probability for second-order sampling error (Hunter & Schmidt, 2004) which lessens the power of the analysis. It is, therefore, possible that with enough power for reading and science domains, the results of the meta-analysis may have been similar to what is reported here.

### *Examination Across Countries*

The effect of student-level achievement on enjoyment showed relative universality across countries. That is, in general, students who had high levels of achievement also showed greater levels of enjoyment. Similarly, given the same level of student-level achievement, students in classes with high average levels of achievement tended to show lower levels of enjoyment. While the direction of these effects was largely consistent across countries, the sizes of these effects showed some variation. We also showed that there were no consistent interactions between student- and class-level achievement with the country-level variables of HDI, Education

Index, and GDP across domains. This is concordant with previous research examining moderators for the negative effect of group-average achievement on self-concept (e.g., Marsh et al., 2021). The only exception was the slightly stronger effect of students' reading achievement on reading enjoyment in countries with a higher HDI and Education Index. Further research is required to elucidate the reasons for the differences in the size of effects for the other domains, and for the variation in the effect of class-average achievement on enjoyment.

As mentioned earlier, CVT proposes that student achievement positively predicts enjoyment because of increased perceived control over the subject matter, and that higher class-average ability reduces this sense of control (Pekrun, 2006; Pekrun et al., 2019). Given the consistency of the direction of effects across countries, our findings suggest that these processes underlying the two effects are likely similarly across countries. This is an important finding given that the samples used in this study are not restricted to WEIRD countries, thus providing a strong basis for replicability and generalizability of results than has been provided before.

We note that there are a few exceptions to this general pattern of findings, especially for the HFLPE in the reading domain. As mentioned earlier, this could be due to the lack of class-context in the measure of reading enjoyment. Still, future research is required to ascertain what is different, if anything, about the very few countries that showed non-significant effects or effects in the opposite direction to the general trend. Future research is also required to examine why differences in the size of the effects across countries occurred, and whether these differences can be explained through country-level differences in educational systems or policies.

**Limitations, Directions for Future Research, and Practical Implications**

While we consider enjoyment as an outcome in this study, we note that causality cannot

be determined with cross-sectional data. There is evidence to suggest that the links between achievement and enjoyment are reciprocal (Forsblom et al., 2022; Pekrun et al., 2017); increases in achievement lead to increases in enjoyment and vice versa. Unfortunately, there is a dearth of cross-national datasets that are also longitudinal in nature. While such datasets would be ideal for future research, we can currently only amalgamate the knowledge we have from research using cross-national, cross-sectional data with research using single-country, longitudinal data. Thus, our study adds to past longitudinal research on the effect of student and class-average achievement on enjoyment by examining these relations across multiple countries.

Our study was limited in relation to the breadth of emotions covered, as we focused specifically on the positive emotion of enjoyment. Past research has shown that student-level achievement is positively and reciprocally related to various positive emotions, and negatively and reciprocally related to negative emotions (Pekrun et al., 2017), with the opposite pattern for group-average achievement (Pekrun et al., 2019). Thus, in line with the findings of our study, we would expect a similar pattern of relations with other positive emotions across domains and countries, and the opposite pattern for negative emotions. In line with relative universality (Pekrun 2006), however, research would need to attend to possible variation of these effects across contexts.

We also note that our study utilized single-item measures of enjoyment. As mentioned in the methods section, past research suggests that it is appropriate and reliable to utilize single-item measures (Allen et al., 2022; Gogol et al., 2014). Still, there are issues of measurement error that are better dealt with using multiple-item scales. Our enjoyment items were taken from a scale assessing learning attitudes in those domains, with the scale for each domain including 6 items. We reran our main models with this scale as the outcome. Results were very similar (see

Supplementary Materials Section 7, Table S10) to our main models with the enjoyment items as the outcome. However, care is required when using multiple-item measures as all items in the scale might not measure the same emotion. For instance, as noted earlier, measures that assess enjoyment and boredom within the same scale (which was also the case with the learning attitudes scale) tend to confound two different emotions (e.g., He et al., 2019). Thus, future research is required to confirm our results using multiple-item measures of enjoyment that are also true measures of this construct.

Despite these limitations, our findings were robust across domains and countries, and thus, have large-scale implications for policy and practice. Our findings suggested that if two students have the same level of achievement, with one student in a class with high achievers and the other student in a class with lower achieving students, then the first student is likely to report lower levels of enjoyment compared to the second student. That is, students tend to enjoy learning a subject less when they are surrounded by high achievers.

As we argue in our paper, and has been shown in prior research (Pekrun et al., 2019), emotions such as enjoyment are impacted by the level of control a student feels in that subject, as well as how much value or importance they place on it. In our example above, this would imply that the first student feels as though they have less control in that subject domain and may also not find gaining knowledge in that subject valuable in and of itself. Thus, to increase students' beneficial emotions such as enjoyment, interventions could increase students' opportunities for success and highlight the importance of learning the subject.

To increase opportunities for experiencing success, students could be trained to compare themselves not with other students, but relative to their own improvements over time. That is,

students could be taught to use intrapersonal standards to evaluate achievement (as opposed to interpersonal standards) and endorse mastery goals (as opposed to competition goals; Pekrun, Cusack, et al. 2014). Past research has indeed shown a positive effect of achievement on enjoyment over time and that this relationship is reciprocal. That is, prior achievement positively predicts enjoyment at a later time point, which in turn positively predicts subsequent achievement (Pekrun et al., 2017, 2022). Thus, focusing on mastery would not only improve students' levels of enjoyment, but also their own achievement over time.

In terms of value, teachers can stress the importance of the subject being taught and highlight how valuable gaining knowledge in that subject is. For instance, it is important to learn math because math helps improve basic logic and problem-solving skills that are useful to traverse through many situations in everyday life. Past research has indeed shown intrinsic value to be positively correlated with enjoyment (Goetz et al., 2006). Further, research from Self-Determination Theory highlights several benefits of engaging in behaviors because those behaviors (in this case, engaging in subject matter) are personally important to the individual, such as increased task persistence (Ryan & Deci, 2017).

The negative effect of class-level achievement on enjoyment may suggest that selective schooling—where students are chosen based on their high achievement levels thereby increasing the average achievement levels of the group—may harm students' academic enjoyment (similar to self-concept: Parker et al., 2021). However, we note that efforts to reduce selective schooling can have unintended consequences, such as increases in course-by-course tracking (Chmielewski, 2014). Instead, policy makers could focus on improving achievement equally for all children or reducing the gap between the highest achievers and lowest achievers. Indeed, research has suggested that countries that have lower variance in achievement (i.e., students

within a country tend to have similar achievement levels) also tend to have higher average achievement (Parker et al., 2018).

Given the generalizability of our results, these implications may be relevant for multiple subjects and across different socio-cultural contexts as represented by different countries. Still, given the lack of research on the HFLPE, future research is required to confirm whether any teacher-level or policy-level constructs truly moderate the effect of group-average achievement on academic emotions.

## Conclusion

We demonstrate the relative universality of positive effects of individual student achievement, and negative effects of class-average achievement, on academic enjoyment. We do so across the subject domains of mathematics, science, and reading, as well as across 37 countries. Our findings add to the literature on achievement emotions by highlighting two important predictors of enjoyment that operate across domains and cross-nationally. Interventions aimed at increasing positive academic emotions can, thus, focus on increasing student-level achievement while helping students to compare their achievement with their own past achievement, rather than in relation to the students' peer groups.

## References

Addabbo, T., Di Tommaso, M.L. & Maccagnan, A. (2016) Education Capability: A Focus on Gender and Science. *Social Indicators Research, 128*, 793–812. https://doi.org/10.1007/s11205-015-1057-8

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*(1), 4-12. https://doi.org/10.1016/j.cedpsych.2010.08.001

Akpınar, E., Yıldız, E., Tatar, N., & Ergin, Ö. (2009). Students' attitudes toward science and technology: an investigation of gender, grade level, and academic achievement. *Procedia-Social and Behavioral Sciences*, *1*(1), 2804-2808. https://doi.org/10.1016/j.sbspro.2009.01.498

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment*, *38*(1), 1-5. https://doi.org/10.1027/1015-5759/a000699

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin, 147*(2), 134-168.https://doi.org/10.1037/bul0000307

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Behrens, C. C., Dolmans, D. H., Gormley, G. J., & Driessen, E. W. (2019). Exploring undergraduate students' achievement emotions during ward round simulation: a mixed-method study. *BMC Medical Education, 19*(1), 1–7. https://doi.org/10.1186/s12909-019-1753-1.

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*, 175–184. https://doi.org/10.1509/jmkr.44.2.175

Brozo, W. G., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, gender, and engagement: Lessons from five PISA countries. *Journal of Adolescent & Adult Literacy, 57*(7), 584-593. https://doi.org/10.1002/jaal.291

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review, 33*(3), 1051-1095.https://doi.org/10.1007/s10648-020-09585-3

Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction, 63*, 101209. https://doi.org/10.1016/j.learninstruc.2019.05.006.

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

Csikszentmihalyi, M. (2000). *The contribution of flow to positive psychology.* American Psychological Association

De Naeghel, J., Valcke, M., De Meyer, I. *et al.* The role of teacher behavior in adolescents' intrinsic reading motivation. *Read Writ* **27**, 1547–1565 (2014). https://doi.org/10.1007/s11145-014-9506-3

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

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin, 136*(1), 103. https://doi.org/10.1037/a0018053

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2022). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology*, *114*(2), 346–367. https://doi.org/10.1037/edu0000671

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218-226.

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*(4), 497-514. https://doi.org/10.1007/BF03173468

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156-168. https://doi.org/10.1177/2515245919847202

Goetz, T. (2004). *Emotionales Erleben und selbstreguliertes Lernen bei Schülern im Fach Mathematik [Students' emotions and self-regulated learning in mathematics]*. Utz.

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology, 33*(1), 9-33. https://doi.org/10.1016/j.cedpsych.2006.12.002

Goetz, T., Pekrun, R., Hall, N., & Haag, L. (2006). Academic emotions from a socialcognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction. British Journal of Educational Psychology, 76, 289–308. https://doi.org/10.1348/000709905X42860

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*(3), 188-205. https://doi.org/10.1016/j.cedpsych.2014.04.002

Gruder, C. L. (1971). Determinants of social comparison choices. *Journal of Experimental Social Psychology, 7*(5), 473-489. https://doi.org/10.1016/0022-1031(71)90010-2

Gunderson, E. A., Hamdan, N., Sorhagen, N. S., & D'Esterre, A. P. (2017). Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Developmental Psychology, 53*(6), 1188. https://doi.org/10.1037/dev0000282

Guo, J., Hu, X., Elliot, A. J., Marsh, H. W., Murayama, K., Basarkod, G., Parker, P. D., & Dicke, T. (2022). Mastery-approach goals: A large-scale cross-cultural analysis of antecedents and consequences. *Journal of Personality and Social* Psychology. https://doi.org/10.1037/pspp0000436

Guo, J., Hu, X., Marsh, H. W., & Pekrun, R. (2021). Relations of epistemic beliefs with motivation, achievement, and aspirations in science: Generalizability across 72 societies. Journal of Educational Psychology. https://doi.org/10.1037/edu0000660

He, J., Barrera-Pedemonte, F., & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice, 26*(4), 369-385.https://doi.org/10.1080/0969594X.2018.1469467

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*, 1-47. https://doi.org/10.18637/jss.v045.i07

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*(1), 156-170. https://doi.org/10.1037/a0015558

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage.

Luo, W., Ng, P. T., Lee, K., & Aye, K. M. (2016). Self-efficacy, value, and achievement emotions as mediators between parenting practice and homework behavior: A control-value theory perspective. *Learning and Individual Differences, 50*, 275-282. https://doi.org/10.18637/jss.v045.i07

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*(2), 256-273.https://doi.org/10.1037/edu0000059

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*(2), 256–273. https://doi.org/10.1037/edu0000059

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

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

Marsh, H. W., 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*(2), 180-202. https://doi.org/10.1002/per.2232

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331-353. https://doi.org/10.1037/edu0000281

Marsh, H. W., Xu, K. M., Parker, P. D., Hau, K. T., Pekrun, R., Elliot, A., ... & Basarkod, G. (2021). Moderation of the big-fish-little-pond effect: Juxtaposition of evolutionary (darwinian-economic) and achievement motivation theory predictions based on a delphi approach. *Educational Psychology Review*, *33*(4), 1353-1378. https://doi.org/10.1007/s10648-020-09583-5

Muis, K. R., Psaradellis, C., Lajoie, S. P., Di Leo, I., & Chevrier, M. (2015b). The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, *42*, 172–185. https://doi.org/10.1016/j.cedpsych.2015.06.003.

Murphy, C. & Beggs, J. (2003). Children's perceptions of school science. *School Science Review*, *84*, 109-116.

OECD (2018). *PISA 2015: Results in Focus.* OECD. Retrieved from https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Parker P. D., Marsh H. W., Jerrim J. P., Guo J., Dicke T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal*, 55(4), 836–858. https://doi.org/10.3102/0002831218760213

Parker, P., Dicke, T., Guo, J., Basarkod, G., & Marsh, H. (2021). Ability stratification predicts the size of the big-fish-little-pond effect. *Educational Researcher, 50*(6), 334-344. https://doi.org/10.3102/0013189X20986176

Peixoto, F., Sanches, C., Mata, L., & Monteiro, V. (2017). "How do you feel about math?": Relationships between competence and value appraisals, achievement emotions and academic achievement. *European Journal of Psychology of Education, 32*(3), 385-405. https://doi.org/10.1007/s10212-016-0299-4

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315-341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. *Big Theories Revisited, 2*, 162-190.

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. https://doi.org/10.1016/j.learninstruc.2013.09.002

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531-549. https://doi.org/10.1037/a0019243

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*(5), 1653-1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., Marsh, H. W., Suessenbach, F., Frenzel, A. C., & Goetz, T. (2022). School grades and students' emotions: Longitudinal models of within-person reciprocal effects. Learning and Instruction. Advance online publication. https://doi.org/10.1016/j.learninstruc.2022.101626

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*(1), 166-185. https://doi.org/10.1037/pspp0000230

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raccanello, D., Brondino, M., Moè, A., Stupnisky, R., & Lichtenfeld, S. (2019). Enjoyment, boredom, anxiety in elementary schools in two domains: Relations with achievement. *The Journal of Experimental Education, 87*(3), 449–469. https://doi.org/10.1080/00220973.2018.1448747.

Respondek, L., Seufert, T., Stupnisky, R., & Nett, U. E. (2017). Perceived academic control and academic emotions predict undergraduate university student success: Examining effects on dropout intention and achievement. *Frontiers in Psychology, 8*, 243. https://doi.org/10.3389/fpsyg.2017.00243

Rogiers, A., Van Keer, H., & Merchie, E. (2020). The profile of the skilled reader: An investigation into the role of reading enjoyment and student characteristics. *International Journal of Educational Research, 99*, 101512. https://doi.org/10.1016/j.ijer.2019.101512

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* John Wiley & Sons. https://doi.org/10.1002/9780470316696

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *American Psychologist, 53*(10), 1101-1110. https://doi.org/10.1037/0003-066X.53.10.1101

Simon, R. A., Aulls, M. W., Dedic, H., Hubbard, K., & Hall, N. C. (2015). Exploring student persistence in STEM programs: a motivational model. *Canadian Journal of Education, 38*(1), n1.

Smith, J. K., Smith, L. F., Gilmore, A., & Jameson, M. (2012). Students' self-perception of reading ability, enjoyment of reading and reading achievement. *Learning and Individual Differences, 22*(2), 202-206. https://doi.org/10.1016/j.lindif.2011.04.010

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483-493. https://doi.org/10.1016/j.jad.2017.11.048

Wan, Z. H., & Lee, J. C. K. (2017). Hong Kong secondary school students' attitudes towards science: A study of structural models and gender differences. *International Journal of Science Education, 39*(5), 507-527. https://doi.org/10.1080/09500693.2017.1292015

ONLINE SUPPLEMENTARY MATERIALS

## Supplementary Materials 1: Country-by-Country Sample Sizes

**Table S1**

*Sample Sizes for Each Country in the Combined TIMSS-PIRLS Data*

| Country | n |
|---|---|
| Azerbaijan, Republic of | 4728 |
| Australia | 5186 |
| Austria | 4366 |
| Botswana | 4133 |
| Chinese Taipei | 4265 |
| Croatia | 4156 |
| Czech Republic | 4298 |
| Finland | 4248 |
| Georgia | 4586 |
| Germany | 3827 |
| Honduras, Republic of | 3604 |
| Hong Kong, SAR | 3802 |
| Hungary | 5047 |
| Iran, Islamic Republic of | 5569 |
| Ireland | 4154 |
| Italy | 4013 |
| Lithuania | 4461 |
| Malta | 3397 |
| Morocco | 7401 |
| Oman | 10155 |
| Norway | 2901 |
| Poland | 4859 |
| Portugal | 3785 |
| Qatar | 4030 |
| Romania | 4485 |
| Russian Federation | 4346 |
| Saudi Arabia | 4426 |
| Singapore | 6208 |
| Slovak Republic | 5363 |
| Slovenia | 4372 |
| Spain | 4058 |
| Sweden | 4290 |
| United Arab Emirates | 14339 |
| Northern Ireland | 3331 |
| United Arab Emirates (Dubai) | 5893 |
| United Arab Emirates (Abu Dhabi) | 4100 |
| Canada (Quebec) | 3902 |

## Supplementary Materials 2: Descriptive Statistics

**Table S2**

*Descriptive Statistics for the Study Variables*

| Variable | Mean (*SD*) | n Missing (%) |
|---|---|---|
| Enjoyment | | |
| Math | 3.40 (0.89) | 4,640 (2.58%) |
| Science | 3.52 (0.82) | 5,497 (3.05%) |
| Reading | 3.40 (0.91) | 4,889 (2.71%) |
| Student-Level Achievement | | |
| Math | 478 (111) | 0 (0%) |
| Science | 473 (121) | 0 (0%) |
| Reading | 474 (117) | 0 (0%) |
| Class-Average Achievement | | |
| Math | 478 (85) | 0 (0%) |
| Science | 473 (93) | 0 (0%) |
| Reading | 474 (90) | 0 (0%) |
| SES | 9.95 (2.03) | 21,310 (11.83%) |
| Age | 10.29 (0.82) | 452 (0.25%) |
| Gender | n (%) | 3 (0.002%) |
| Female | 88,712 (49%) | |
| Male | 91,369 (51%) | |

*Note.* Achievement scores here use the first plausible value and all descriptives are based on raw, unimputed data.
Total sample size = 180,084.

## Supplementary Materials 3: Country-Level Moderators

**Table S3**
*Results From Multilevel Models Predicting Enjoyment, Including Interactions with HDI*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.201* | .041 | .024 | .027 | .404* | .041 |
| Student-Level Ach | .167* | .014 | .097* | .012 | .235* | .015 |
| Class-Average Ach | -.171* | .020 | -.121* | .016 | -.075* | .022 |
| SES | -.029* | .004 | .016* | .003 | .054* | .003 |
| Gender | .110* | .005 | -.007 | .005 | -.323* | .005 |
| Age | -.008* | .004 | -.014* | .004 | -.019* | .003 |
| HDI | -.131* | .040 | -.168* | .026 | -.154* | .040 |
| Student Ach*HDI | .021 | .014 | .008 | .012 | .066* | .015 |
| Class Ach*HDI | -.016 | .020 | -.037* | .016 | .050* | .021 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .243 | | .253 | | .239 | |
| Student Ach \| Class | .097 | | .120 | | .115 | |
| Intercept \| Country | .244 | | .155 | | .242 | |
| Student Ach \| Country | .079 | | .067 | | .087 | |
| Class Ach \| Country | .106 | | .073 | | .110 | |
| Residuals | .914 | | .926 | | .912 | |

*Note.* Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Ach = achievement; HDI = Human Development Index.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

**Table S4**

*Results From Multilevel Models Predicting Enjoyment, Including Interactions With Education Index*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.194* | .043 | .032 | .027 | .414* | .040 |
| Student-Level Ach | .165* | .013 | .098* | .012 | .231* | .014 |
| Class-Average Ach | -.168* | .019 | -.118* | .017 | -.075* | .023 |
| SES | -.029* | .004 | .016* | .003 | .054* | .003 |
| Gender | .110* | .005 | -.008 | .005 | -.323* | .005 |
| Age | -.008* | .004 | -.014* | .004 | -.019* | .003 |
| Education Index | -.129* | .043 | -.174* | .027 | -.174* | .040 |
| Student Ach*EI | .031* | .014 | .001 | .012 | .076* | .014 |
| Class Ach*EI | -.036 | .020 | -.035* | .017 | .039 | .023 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .243 | | .253 | | .239 | |
| Student Ach \| Class | .097 | | .120 | | .115 | |
| Intercept \| Country | .251 | | .154 | | .233 | |
| Student Ach \| Country | .075 | | .067 | | .080 | |
| Class Ach \| Country | .099 | | .074 | | .116 | |
| Residuals | .914 | | .926 | | .913 | |

*Note.* Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Ach = achievement; EI = Education Index component of the Human Development Index.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

**Table S5**

*Results From Multilevel Models Predicting Enjoyment, Including Interactions With GDP Per Capita*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.213* | .046 | .007 | .038 | .391* | .048 |
| Student-Level Ach | .169* | .014 | .098* | .012 | .241* | .018 |
| Class-Average Ach | -.174* | .020 | -.124* | .016 | -.070* | .023 |
| SES | -.029* | .004 | .016* | .003 | .054* | .003 |
| Gender | .110* | .005 | -.008 | .005 | -.323* | .005 |
| Age | -.008* | .004 | -.014* | .004 | -.019* | .003 |
| GDP | -.059 | .046 | -.070 | .037 | -.047 | .048 |
| Student Ach*GDP | -.006 | .014 | .021 | .012 | .024 | .018 |
| Class Ach*GDP | .009 | .020 | -.027 | .015 | .035 | .021 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .243 | | .253 | | .239 | |
| Student Ach \| Class | .097 | | .120 | | .115 | |
| Intercept \| Country | .274 | | .222 | | .289 | |
| Student Ach \| Country | .081 | | .064 | | .108 | |
| Class Ach \| Country | .106 | | .073 | | .117 | |
| Residuals | .914 | | .926 | | .913 | |

*Note.* Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Ach = achievement; GDP = Gross Domestic Product per capita.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

## Supplementary Materials 4: Country-by-Country Student-Level and Class-Average Predictive Effects

**Table S6**

*Country-by-Country Estimates and Standard Errors for Student-Level Achievement Predicting Enjoyment, From Two-Level Models Presented in Main Text Forest Plots.*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| Country | β | SE | β | SE | β | SE |
| Australia | .207* | .029 | .105* | .034 | .348* | .033 |
| Austria | .216* | .039 | 0.079 | .042 | .294* | .044 |
| Azerbaijan, Republic of | .093* | .016 | .085* | .026 | .128* | .027 |
| Botswana | .121* | .030 | .236* | .022 | .287* | .030 |
| Canada (Quebec) | .226* | .040 | .073 | .058 | .314* | .035 |
| Chinese Taipei | .350* | .033 | .100* | .034 | .269* | .031 |
| Croatia | .202* | .039 | -.069 | .046 | .239* | .047 |
| Czech Republic | .170* | .044 | .044 | .043 | .299* | .040 |
| Finland | .228* | .036 | .089* | .045 | .509* | .034 |
| Georgia | .096* | .017 | .086* | .032 | .114* | .025 |
| Germany | .234* | .043 | .205* | .052 | .341* | .033 |
| Honduras, Republic of | .042 | .031 | .015 | .037 | .063 | .033 |
| Hong Kong, SAR | .316* | .043 | .213* | .046 | .311* | .037 |
| Hungary | .286* | .026 | .167* | .035 | .345* | .030 |
| Iran, Islamic Republic of | .089* | .018 | .089* | .020 | .106* | .021 |
| Ireland | .152* | .041 | .116* | .036 | .233* | .030 |
| Italy | .209* | .033 | .099* | .039 | .228* | .032 |
| Lithuania | .256* | .031 | .016 | .034 | .130* | .033 |
| Malta | .178* | .034 | .115* | .034 | .181* | .027 |
| Morocco | .060* | .021 | .048* | .013 | .134* | .018 |
| Northern Ireland | .155* | .036 | .173* | .057 | .287* | .033 |
| Norway | .108* | .039 | .082 | .047 | .309* | .044 |
| Oman | .100* | .012 | .115* | .012 | .158* | .017 |
| Poland | .168* | .027 | .118* | .034 | .333* | .029 |
| Portugal | .250* | .035 | .078* | .029 | .131* | .028 |
| Qatar | .069* | .023 | .115* | .023 | .127* | .023 |
| Romania | .179* | .021 | .090* | .028 | .167* | .028 |
| Russian Federation | .229* | .034 | .017 | .030 | .178* | .032 |
| Saudi Arabia | 0.048 | .028 | .083* | .026 | .094* | .033 |
| Singapore | .279* | .029 | .245* | .031 | .259* | .024 |
| Slovak Republic | .276* | .028 | .014 | .032 | .234* | .035 |
| Slovenia | .158* | .037 | -.043 | .036 | .370* | .032 |
| Spain | .147* | .034 | -.019 | .041 | .244* | .035 |
| Sweden | .085* | .029 | .057 | .036 | .374* | .035 |
| United Arab Emirates | .078* | .014 | .123* | .013 | .168* | .016 |
| United Arab Emirates (Abu Dhabi) | .062* | .025 | .128* | .027 | .177* | .026 |
| United Arab Emirates (Dubai) | .108* | .027 | .126* | .020 | .201* | .023 |

**Table S7**

*Country-by-Country Estimates and Standard Errors for Class-Level Achievement Predicting Enjoyment, From Two-Level Models Presented in Main Text Forest Plots.*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| Country | β | SE | β | SE | β | SE |
| Australia | -.151* | .068 | -.025 | .076 | .016 | .058 |
| Austria | -.306* | .093 | -.223* | .072 | -.040 | .093 |
| Azerbaijan, Republic of | -.017 | .025 | .034 | .031 | .107* | .042 |
| Botswana | -.073 | .055 | -.068* | .033 | -.143* | .052 |
| Canada (Quebec) | -.138 | .105 | -.077 | .137 | -.075 | .097 |
| Chinese Taipei | -.227 | .120 | -.049 | .146 | .172 | .109 |
| Croatia | -.258 | .140 | .053 | .145 | -.012 | .144 |
| Czech Republic | -.199 | .102 | -.265* | .131 | -.032 | .094 |
| Finland | -.297* | .116 | -.172 | .133 | -.212* | .095 |
| Georgia | -.116* | .029 | -.088 | .055 | -.169* | .047 |
| Germany | -.461* | .097 | -.421* | .090 | -.102 | .087 |
| Honduras, Republic of | -.402* | .070 | -.076 | .056 | -.451* | .072 |
| Hong Kong, SAR | -.140 | .081 | -.094 | .101 | .196* | .093 |
| Hungary | -.273* | .058 | -.233* | .077 | -.176* | .065 |
| Iran, Islamic Republic of | -.193* | .041 | -.109* | .034 | -.182* | .037 |
| Ireland | -.208* | .086 | .105 | .089 | .088 | .070 |
| Italy | -.148 | .078 | -.232* | .086 | -.085 | .080 |
| Lithuania | -.173* | .061 | -.098 | .066 | -.009 | .078 |
| Malta | .104 | .086 | .012 | .077 | -.020 | .061 |
| Morocco | .032 | .041 | .046 | .036 | .034 | .041 |
| Northern Ireland | -.094 | .102 | .116 | .128 | .295* | .101 |
| Norway | .104 | .101 | .153 | .138 | .238* | .109 |
| Oman | -.020 | .027 | -.076* | .025 | -.075* | .034 |
| Poland | -.146* | .062 | -.327* | .074 | -.151* | .068 |
| Portugal | -.020 | .064 | -.126 | .067 | -.021 | .066 |
| Qatar | -.089 | .043 | -.111* | .039 | .024 | .042 |
| Romania | -.240* | .043 | -.170* | .049 | -.241* | .051 |
| Russian Federation | -.257* | .056 | -.071 | .069 | -.019 | .074 |
| Saudi Arabia | -.008 | .046 | -.052 | .052 | -.118* | .062 |
| Singapore | -.030 | .044 | -.290* | .044 | -.087* | .036 |
| Slovak Republic | -.319* | .060 | -.174* | .072 | .057 | .073 |
| Slovenia | -.328* | .113 | -.169 | .111 | -.013 | .102 |
| Spain | -.067 | .095 | -.211* | .103 | -.143 | .095 |
| Sweden | -.315* | .085 | -.172* | .083 | -.171* | .085 |
| United Arab Emirates | -.123* | .023 | -.087* | .025 | -.065* | .026 |
| United Arab Emirates (Abu Dhabi) | -.124* | .044 | -.127* | .050 | -.112* | .041 |
| United Arab Emirates (Dubai) | -.156* | .035 | -.097* | .040 | -.087* | .038 |

## Supplementary Materials 5: A Test of the CVT Proposition

While a proper test of the CVT proposition would require a mediation model with longitudinal data, we conducted a quick test to examine whether the inclusion of proxy measures of control and value would reduce the fixed effects of student- and class-level achievement on enjoyment. The TIMSS-PIRLS 2011 dataset included a measure of self-concept, which can be thought of as "competence appraisal" and has been used in past research as a proxy for perceived control. The closest item measuring something akin to value was a single item measuring importance in doing well in math and science ("It is important to do well in mathematics/science"), but this item was not available for reading. We included the self-concept scale and importance items to our models, the results of which are presented in Table S8.

Self-concept had a positive effect on enjoyment in all three domains, while importance had a positive effect on enjoyment in math and science (as noted, this item was not available for reading). Importantly—and consistent with CVT—the inclusion of self-concept and importance substantially reduced the effects of student-level and class-average achievement on enjoyment. Specifically, the effect of student-level achievement on enjoyment was now close to zero for math and science, and substantially reduced for reading (changing from .240 to .097). The effect of class-average achievement on enjoyment was now non-significant for all three domains. The combination of direct effects of self-concept and importance, combined with substantially reduced direct effects of student- and class-level achievement, support the interpretation that the effects of achievements were largely mediated by self-concept and importance, similar to the pattern of effects in the original HFLPE paper by Pekrun et al. (2019). In addition, the residual variances were smaller as well (previously ~.920), implying that the inclusion of self-concept and importance helped explain some of the variance in enjoyment.

**Table S8**

*Results From Multilevel Models Predicting Enjoyment, With Self-Concept and Importance as Additional Predictors*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.115* | .036 | -.056 | .029 | .351* | .046 |
| Student-Level Ach | -.057* | .010 | -.050* | .013 | .097* | .017 |
| Class-Average Ach | -.019 | .012 | -.016 | .012 | .002 | .023 |
| SES | -.071* | .003 | -.025* | .005 | .022* | .003 |
| Gender | .053* | .004 | .026* | .004 | -.293* | .005 |
| Age | .005 | .003 | .000 | .003 | -.013* | .003 |
| Self-Concept | .417* | .002 | .344* | .003 | .267* | .002 |
| Importance | .206* | .002 | .278* | .003 | - | - |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .185 | | .073 | | .223 | |
| Student Ach \| Class | .084 | | .129 | | .114 | |
| Intercept \| Country | .214 | | .168 | | .272 | |
| Student Ach \| Country | .051 | | .073 | | .097 | |
| Class Ach \| Country | .050 | | .051 | | .121 | |
| Residuals | .797 | | .822 | | .883 | |

*Note.* The importance item was only available for math and science domains.
Ach = achievement.
Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

## Supplementary Materials 6: Interaction Between Student-Level and Class-Average Achievement

**Table S9**

*Results From Multilevel Models Predicting Enjoyment, With Interaction Between Student-Level and Class-Average Achievement.*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.211* | .046 | .008 | .039 | .393* | .048 |
| Student-Level Ach | .168* | .012 | .098* | .012 | .240* | .018 |
| Class-Average Ach | -.178* | .020 | -.122* | .016 | -.076* | .023 |
| SES | -.030* | .004 | .016* | .003 | .054* | .003 |
| Gender | .110* | .005 | -.007 | .005 | -.323* | .005 |
| Age | -.009* | .004 | -.013* | .004 | -.020* | .003 |
| Student Ach * Class Ach | .026* | .007 | -.027* | .007 | .019* | .007 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .244 | | .253 | | .239 | |
| Student Ach \| Class | .098 | | .119 | | .116 | |
| Intercept \| Country | .275 | | .232 | | .287 | |
| Student Ach \| Country | .069 | | .068 | | .102 | |
| Class Ach \| Country | .105 | | .070 | | .119 | |
| Residuals | .914 | | .926 | | .912 | |

*Note.* Ach = achievement.
Fixed effects marked with an asterisk are statistically significant at $p < .05$.
Random effects are shown in standard deviations.
The labels for these random effects represent the intercept/variable being random at the class and country levels.

## Supplementary Materials 7: Learning Attitudes Scale as Outcome

**Table S10**

*Results From Multilevel Models Predicting Learning Attitudes*

| | Math | | Science | | Reading | |
|---|---|---|---|---|---|---|
| | β | *SE* | β | *SE* | β | *SE* |
| **Fixed Effects** | | | | | | |
| Intercept | -.067 | .045 | .139* | .037 | .486* | .046 |
| Student-Level Ach | .265* | .012 | .248* | .016 | .307* | .016 |
| Class-Average Ach | -.173* | .021 | -.122* | .022 | -.076* | .024 |
| SES | -.006 | .004 | .044* | .003 | .115* | .003 |
| Gender | .045* | .005 | -.073* | .005 | -.362* | .005 |
| Age | -.032* | .003 | -.035* | .003 | -.014* | .003 |
| **Random Effects (*SDs*)** | | | | | | |
| Intercept \| Class | .283 | | .291 | | .000 | |
| Student Ach \| Class | .102 | | .111 | | .187 | |
| Intercept \| Country | .267 | | .219 | | .276 | |
| Student Ach \| Country | .069 | | .094 | | .094 | |
| Class Ach \| Country | .107 | | .110 | | .135 | |
| Residuals | .905 | | .901 | | .909 | |

*Note.* Ach = achievement.

Fixed effects marked with an asterisk are statistically significant at $p < .05$.

Random effects are shown in standard deviations.

The labels for these random effects represent the intercept/variable being random at the class and country levels.