2023

Multilevel Modeling of Academic Self-Concept and Moderation of the Big-Fish-Little-Pond Effect in Math and Science for TIMSS 2019 Participating Countries

Heather Marie Spangler

Multilevel Modeling of Academic Self-Concept and Moderation of the Big-Fish-Little-Pond Effect in Math and Science for TIMSS 2019 Participating Countries

by
Heather Spangler

An Applied Dissertation Submitted to the
Abraham S. Fischler College of Education
and School of Criminal Justice in Partial
Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Educational
Research and Evaluation

Nova Southeastern University
2023

Approval Page

This applied dissertation was submitted by Heather Spangler under the direction of the persons listed below. It was submitted to the Abraham S. Fischler College of Education and School of Criminal Justice and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Educational Research and Evaluation at Nova Southeastern University.

Steven Hecht, PhD
Committee Chair

Dana Mills, PhD
Committee Member

Kimberly Durham, PsyD
Dean

Statement of Original Work

I declare the following:

I have read the Code of Student Conduct and Academic Responsibility as described in the *Student Handbook* of Nova Southeastern University. This applied dissertation represents my original work, except where I have acknowledged the ideas, words, or material of other authors.

Where another author's ideas have been presented in this applied dissertation, I have acknowledged the author's ideas by citing them in the required style.

Where another author's words have been presented in this applied dissertation, I have acknowledged the author's words by using appropriate quotation devices and citations in the required style.

I have obtained permission from the author or publisher—in accordance with the required guidelines—to include any copyrighted material (e.g., tables, figures, survey instruments, large portions of text) in this applied dissertation manuscript.

Heather Marie Spangler____________
Name

November 30, 2023_________
Date

Abstract

Multilevel Modeling of Academic Self-Concept and Moderation of the Big-Fish-Little-Pond Effect in Math and Science for TIMSS 2019 Participating Countries. Heather Spangler, 2023: Applied Dissertation, Nova Southeastern University, Abraham S. Fischler College of Education and School of Criminal Justice. Keywords: Math Self-Concept, Science Self-Concept, Academic Self-Concept, TIMSS 2019, school-level BFLPE, country-level BFLPE, BFLPE moderation, Hierarchical Linear Modeling, cultural programming, social comparisons.

Self-concept is an important construct across a variety of disciplines as a facilitator of a full range of human potential. Big Fish Little Pond Effect (BFLPE) results have confirmed global generalizability for the negative effects of school- and country-averaged achievement on students' academic self-concept based on social comparisons with implications that generally discredit ability grouping, streaming or tracking. However, few studies have identified variables that ameliorate the negative effects of BFLPE. Accordingly, this study applied hierarchical linear modeling (HLM) in HLM 8.2 software to examine the effects of student, school, and country-level moderators of both school- and country-level BFLPE for STEM subjects. As a secondary analysis of TIMSS 2019 international, large-scale assessment results in math and science for a sample of 169,810 eight grade students in 5,410 school in 26 countries, these results revealed specific affective, cognitive, environmental, and financial factors that diminished and reversed the negative effects of school-and country-level BFLPE. Results also extended Marsh (2020) BFLPE-CE model and offer a standardized framework by which students can be more compatibly grouped, streamed or tracked. Furthermore, implications of these results for educational psychology suggest a hierarchical structure of the social comparison process whereby individuals have the greatest overall impact on their perceptions, but macro-level, unconscious cultural preprogramming is the overarching influence through which perceptions are filtered.

Table of Contents

Tables

Figures

Chapter 1: Introduction

Noticeably, an abundance of research has shown self-concept to be an important construct across a variety of disciplines as a facilitator of a full range of human potential and a vital element of well-being (Chiu & Klassen, 2010; Marsh et al., 2015; Marsh & Martin, 2011; Moller et al., 2009; Primavera et al., 1974; Rosenberg, 1989; Scheirer & Kraut, 1979; West & Fish, 1973; Wylie, 1979). Distinctly, self-concept is widely known "not only as an outcome, but a mediating variable that subjectively facilitates the attainment of other desirable psychological and behavioral outcomes" (Bandura, 1994; Marsh & Craven, 2006, p. 134; Marsh & Hau, 2003, p. 364; Möller et al., 2009, p. 1130). Especially, the generalizability of the reciprocal nature of academic self-concept and achievement have motivated countless educational research studies in contribution to the improvement of educational policy and learning (Marsh, 1999). Notably, academic self-concept (ASC) has been shown to be a positive predictor of achievement in math and science (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Lui & Meng, 2010; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004).

Nevertheless, every student is unique and possesses a distinguishable level of self-concept. In fact, early research has shown variability in self-concept to be due in part to influences from not only intrinsic attributes, but also influences from proximal and distal environments (Epstein, 1973; Hattie, 1992; Labenne & Greene, 1969; Marsh & O'Mara, 2010; Purkey, 1970; Rogers, 1951; Wylie, 1974). To illustrate, students worldwide attend compulsory education generally starting at the age of six and continuing for at least 13 years (Kelly et al., 2020). Logically, with so much time spent in school it is inevitable that influences from school and classroom environments could potentially affect individual students' self-concept and achievement (Hooper et al., 2017, p.68). Equally as pervasive are the influences from distal environments found within each students' country of residence (Chiu & Klassen, 2010, p. 35; Wilkins, 2004).

In practice, researchers have considered the effects of student attributes, as well as school and country contexts as a means of explaining differences in student's self-concepts. In other words, student-level, school-level, and country-level effects on self-concept have been examined to determine variability in student-level self-concept. Case in point, Arens et al. (2017) suggested to investigate the contextual influences that affect academic self-concept and achievement discreetly (p.625), stimulating countless studies that have reported on student-level, school-level and country-level contextual predictors of academic self-concept and achievement as well as reported on moderation of the positive relationship that exists between them (Arens et al.,2017; Chiu & Klassen, 2008; Hooper et al., 2013; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Tucker-Drob et al., 2014; Zheng et al.2019). For instance, Mohammadpour and Ghafar (2014) found that 40.39% of variance in Math achievement was attributed to variables at the student-level. Specifically, student-level variables such as math self-concept, family socioeconomic status (SES), students' valuing math, attitude toward math, and gender were found to have significant linkages to math achievement within schools (p.199). Likewise, in 29 countries, Mohammadpour et al. (2015) reported an averaged 43.33% of variance explained in science achievement accounted for at the student-level with greater proportions accounted for by developed countries versus developing.

As well, school-level variables such as teacher quality, school location, school climate, school size, and school SES composition were examined for associations with self-concept and academic achievement in math and science (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Coleman, 1975; Labenne & Greene, 1969; Martin et al., 2016; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Strein & Grossman, 2010). Furthermore, past research has reported that country-level factors such as how the degree of cultural individualism (Hofstede, 2003), per capita income and gross domestic product (Chiu & Klassen, 2010; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Tucker-Drob et al., 2014; Zheng et al., 2019), as well as country-wide school tracking practices (Arens et al., 2017; Martin Hooper et al., 2013; Mohammadpour & Ghafar, 2014) has impacted one's perception of his/her own academic ability and academic performance. However, much of the research on academic self-concept has focused on student and school-level effects with far fewer studies that have examined country-level effects.

Remarkably, Herbert Marsh and his colleagues have spent a good part of four decades refining a model in pursuit of a greater understanding of academic self-concept. Analogous to an amalgamation of Davis (1959) theory of Relative Deprivation and Cialdini (1980) theory of Basking in Reflected Glory with emphasis on social comparison as the frame of reference, their esteemed BFLPE theory postulates that students evaluate themselves by class, school and country averaged achievement, whereby high-achieving students placed in a high-achieving classroom or countries have reported lower academic self-concept than high achieving students placed in a mixed ability classrooms or in a lower achieving country (Marsh et al., 2019; Marsh & Parker, 1984).

Specifically, whereas earlier implications suggested a dual categorization of social comparisons on students' self-concept whereby contrast ushered negative effects of comparisons to higher averaged achievement and assimilation supported positive effects of comparisons to those with similar or lower averaged achievement (Marsh et al., 2008; Marsh & Parker, 1984; Marsh et al., 2000), BFLPE research has consistently demonstrated the presence of negative class, school and country compositional effect of averaged achievement on student self-concept, while supporting the persistence of positive relations between student-level achievement and student-level academic self-concept (Huguet, 2009; Marsh et al., 2019; Marsh & Parker, 1984; Marsh et al., 2020; Marsh et al., 2007, 2008; Nagengast & Marsh, 2012; Pekrun et al., 2019; Seaton et al., 2009, 2010). Therefore, implications for BFLPE effects on improving achievement prevailingly discredit academically selective schools, as well as ability grouping, streaming, and tracking practices (Dicke et al., 2018; Marsh et al., 2008a; Marsh et al., 2008; Trautwein et al., 2006).

It was a decade after the turn of the century, that studies began to explore mediation and moderation effects of variables on BFLPE. Marsh et al. (2008) reported a lesser negative effect of school-averaged achievement for students' academic self-beliefs such as self-efficacy, control expectations, control strategies, and effort persistence than for students' academic self-concept. Whereas, Seaton et al. (2010) examined the characteristics of student self-regulation as a potential moderator of the negative effects of school-level achievement on students' academic self-concept, results compliment a stronger association for BFLPE with surface learning strategies, high anxiety, and cooperative orientation. By exploring relationships from a unique angle, Nagengast and

Marsh (2012) confirmed academic self-concept as a mediator of negative effects of school averaged achievement on career aspirations, while Pekrun et al. (2019) suggested that academic self-concept mediated the negative effects of school averaged achievement on positive emotions and vice versa for negative emotions.

Correspondingly, more recent findings have suggested that "contextual effects matter for BFLPE, not only at the micro-contextual student- and meso-contextual school-levels, but at the macro-contextual country-level as well" (Marsh et al., 2019, p. 231, 2020). Nevertheless, utilizing cross-cultural analyses with large-scale international assessment data, results have supported global generalizations of the negative effects of BFLPE on a variety student-level predictors such as SES (Marsh & Parker, 1984, p. 198; Marsh & O'Mara, 2009; Pekrun et al., 2019; Seaton et al., 2010), gender (Marsh & Parker, 1984; Marsh et al., 2007; Pekrun et al., 2019), grade point average (Marsh & O'Mara, 2009), ability/IQ (Pekrun et al., 2019) and class standing (Huguet, 2009). However, as a "means for policy and practice to minimize BFLPE's negative impact of contrast effects and maximize the positive impact of assimilation effects" on student academic self-concept (Cheng et al., 2014; Jonkmann et al., 2012; Schwabe et al., 2019) compelling research has investigated the underlying mechanisms and intervening factors that moderate the magnitude or direction of BFLPE (Seaton et al. 2010; Dai and Rinn, 2008; Cheng 2014).

Precisely, previous examinations of BFLPE moderation have only investigated the influence of student characteristics on BFLPE, but effect sizes were unanimously small (Cheng et al., 2014; Jonkmann et al., 2012; McFarland & Buehler, 1995; Plieninger & Dickhäuser, 2015; Schwabe et al., 2019; Seaton et al., 2010; Wouters et al., 2015). For instance, Cheng et al. (2014) applied the Tymms (2004) effects size measure to report the effect sizes of significant moderating effects for several student-level moderators such as goal orientations, intrinsic and extrinsic motivations with results that ranged from −0.059 to −0.091, but in relation to 1.0 were considered small and unsubstantial (p. 570). Likewise, Seaton (2010) reported comparable moderation effects of BFLPE for similar student-level variables. Thus, Marsh and Hau (2003) and Seaton et al. (2009, 2010) suggested the incorporation of contextual variables that impact school performance such as measures of school SES, expenditures, resources, and climate are suggested be factored into future BFLPE research (Marsh & Parker, 1984; Marsh & O'Mara, 2009; Seaton et al., 2009).

Furthermore, suggestions for future BFLPE research recommends examining additional variables of interest as well as methodological improvements (Dai & Rinn, 2008; Huguet et al., 2009; Marsh & Hau, 2003; Marsh, Kong & Hau, 2000; Marsh & O'Mara, 2009; Marsh & Parker, 1984; Marshet al., 2007; Marsh et al., 2008; Morin, et al., 2014; Nagengast & Marsh, 2011; Pekrun et al. , 2019; Seaton et al., 2009, 2010; Wang, 2015). Broadly, the inclusion of student characteristics and differences were suggested to understand influences on covariates at the individual level (Marsh & Hau, 2003; Pekrun et al., 2019; Seaton et al., 2009). For the same reason, Huguet et al. (2009), Marsh et al. (2008), and Seaton et al. (2010), endorsed self-efficacy and other measures of emotional underpinnings be included as well. Additionally, Marsh and Hau (2003) and Seaton et al. (2009, 2010) favored the incorporation of confounding variables that impact school performance be factored into future BFLPE research. Specifically, measures of school SES, expenditures, resources, and climate were suggested (Marsh & O'Mara, 2009;

Marsh & Parker, 1984; Seaton et al., 2009). On the other hand, Dai and Rinn (2008) suggested that it was "possible to have multiple reference groups for social comparisons that extend beyond their local ponds (p. 293)," so directed future research to focus not only on proximal influences from school indicators, but also distal influences from country indicators such as cultural and economic dimensions too.

Research Problem

In 2019, the most recent Trends in International Math and Science (TIMSS) study reported that achievement scores in math and science for 8th graders in the United States has not significantly improved in the past four years and the achievement gap between highest and lowest achieving students has widened between 2015 and 2019 (Mullis et al., 2020). As well, TIMSS 2019 trend data revealed that since 1999, averaged scores have increased by less than 1-point. Specifically, 8th grade math scores only increased by 23-points from 1999 to 2019 and science scores only increased by 9-points. These current large-scale, international assessment results are just one example that education in the United States needs a more innovative approach to improve achievement in STEM subjects.

Remarkably, multitudes of educational research have confirmed that a reciprocally positive and mutually beneficial relationship exists between academic self-concept and achievement in math and science such that improvements in ASC would also improve achievement. Yet, issues of concern have invariably prevented the consideration of such implications for educational policy and practice. Nevertheless, the large body of educational research concerning BFLPE offers a valuable avenue to improve ASC by highlighting the effects of multilevel influences on ASC and its underlying processes.

However, while BFLPE research has clearly portrayed the presence of negative effects on ASC from social comparisons to school- and country-level achievement worldwide, clarity concerning contextual associations of multilevel predictors that moderate those negative effects has been overwhelmingly deficient. Additionally, a great deal of BFLPE investigations have applied limited statistical designs with outdated data sources and examined few multilevel contextual associations with ASC in math and science. Precisely, current BFLPE research lacks analyses that examine the global generalizability for effects of country-averaged achievement on science self-concept (L3BFLPE). As well, analyses of effects for multilevel predictors of academic self-concept in math and science are absent in current BFLPE research, as are analyses that examine cross-cultural generalizability of multilevel moderators of the negative effects of L2BFLPE and L3BFLPE in math and science.

Purpose

Therefore, it was the purpose of this study to extend the application of current BFLPE theory by contributing to current deficiencies in BFLPE research in three ways. First, this study has offered global generalizability for results of an empirical examination of the effects of both L2BFLPE and L3BFLPE on academic self-concepts in math and science. Second, this study has offered cross-cultural generalizability of results for an examination of discrete, multilevel contextual effects of multiple student- school- and country-level predictors of academic self-concept in math and science. Third, this study has contributed global generalizability of results for an empirical examination of multiple micro-, meso-, and macro-level moderators of the negative effects of L2BFLPE and L3BFLPE for STEM subjects.

Explicitly, this study synthesized partial replications of prior BFLPE for a sample of the most current large-scale international assessment results in math and science within and across 26 countries (Marsh et al., 2020; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Seaton et al., 2010). In fact, three-level hierarchical linear modeling was applied as a secondary analysis of the most recent TIMSS 2019 large-scale international results in math and science for 169,957 eighth grade students in 5,410 schools from 26 countries to investigate the existence of school- and country-level BFLPE, multilevel contextual associations with academic self-concept, as well as moderation effects of school- and country-level BFLPE in math and science. In doing so, the results of this study can inform policymakers, administrators, and practitioners alike to advance STEM policy and instruction of math and science with a greater understanding of intervening variables that could potentially minimize the negative effects of social comparisons and maximize the benefits of internal comparisons on students' perceptions of their academic ability and ultimately improve their corresponding achievement in STEM subjects.

Definition of Terms

Academic Self-Concept

The perception of one's own ability in a given academic subject (Shevelson, et al., 1976; Marsh et al., 1988; Mullis & Martin, 2013). Identified in TIMSS 2015 Grade 8 Student Questionnaire as items BSBM19A, B, C, and D for students' math self-concept (L1MSC) and as items BSBS24A, B, C, D for students' science self-concept (L1SSC)

(Foy, 2017). Student-level results were aggregated to the school-level (L2MSC and L2SSC) and country-level (L3MSC and L3SSC).

Achievement

TIMSS 2019 Grade 8 measurement of students' overall achievement in math and science. Identified as the five plausible value items BSMMAT01-BSMMAT05 for students' math achievement (L1MACH1-5) and items BSSSCI01-BSSSCI05 for students' science achievement (L1SACH1-5) (Foy, 2017). Student-level results were aggregated to the school-level (L2MACH1-5 and L2SACH1-5) and country-level (L3MACH1-5 and L3SACH1-5).

Attitude

The extent to which students like learning math and science. Identified in TIMSS 2019 Grade 8 Student Questionnaire as items BSBM16A, C, E for students' attitude toward math (L1ATM) and items BSBS22A, C, E for students' attitude toward science (L1ATS) (Foy, 2017). Student-level results were aggregated to the school-level (L2ATM and L2ATS) and country-level (L3ATM and L3ATS).

Climate

Measurement of school's emphasis on success (L2CLM) as reported by principals. Identified in TIMSS 2019 Grade 8 Teacher Questionnaire as items BCBG14A-M (Foy, 2017).

Cultural Classification

Binary classification of country's type of society (L3IDV). "*Individualism* pertains to societies in which the ties between individuals are loose: everyone is expected to look after him- or herself and his or her immediate family. *Collectivism* as its opposite pertains to societies in which people from birth onward are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty" (Hofstede, 2001).

Gross National Income Per Capita

Measurement of country economy in US dollars (L1IPC). Identified as demographic information in TIMSS 2019 Encyclopedia (Kelly et al., 2020).

Gender

Binary determination of student sex (L1GND). Identified in TIMSS 2019 Grade 8 Student Questionnaire as item ITSEX for male or female.

Location

School location measured as urban and densely populated, suburban and on the fringe, medium city, small town, or remote. Identified in TIMSS 2019 Grade 8 School Questionnaire as item BCBG05A (Foy, 2017).

Socioeconomic Status

Student socioeconomic status (L1SES) as measured by combined home resources and school's socioeconomic status (L2SES) as measured by percentage of students from advantaged or disadvantaged homes. L1SES was identified in TIMSS 2019 Grade 8 Student Questionnaire as items BCBG04, 05C-D. L2SES was identified in TIMSS 2015 Grade 8 School Questionnaire as items BCBG03A-B.

Tracking Practices

National educational tracking practices measured as no tracking, tracking in primary, secondary, and tertiary, or tracking for tertiary only. Reported in TIMSS 2019 Country Questionnaire as item GEN11A.

Value

Extent to which students value math and science. Identified in TIMSS 2019 Grade 8 Student Questionnaire as items BSBM20A-C, F, G, I for students' value of math and items BSBS25A-C, F, G, I for students' value of science (Foy, 2017).

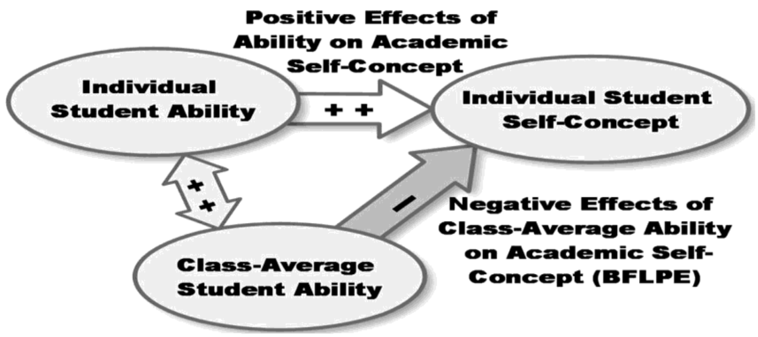
Chapter 2: Literature Review

Coined the Big Fish Little Pond Effect (BFLPE), Herbert Marsh's empirical theory has postulated that "students who attend schools where the school-averaged achievement is high tend to have lower academic self-concepts than do equally able students who attend schools with mixed or low level of achievement" (Marsh et al., 2019) (see Figure 1). In contribution to the "theoretical understand of academic self-concept and its measures," Marsh and Parker (1984) seminal contributions of the BFLPE theory asserted that individuals employ a generalized other as an external frame of reference by whom they evaluate themselves (p. 229; Festinger, 1954). In its earliest inception, the BFLPE was portrayed as an amalgamation of both Davis (1959) theory of Relative Deprivation wherein high ability students were deprived of success in a low achieving classroom (Cialdini & Richardson, 1980) and the theory of Basking in Reflected Glory that implied students regardless of their ability would desire to reflect the success of their classmates if placed in a high achieving classrooms.

Likewise, early notions as well suggested the possibility of a dual categorization of the effects of ability grouping on students' academic self-concept whereby *contrast* ushered negative effects of comparisons to higher class and school averaged achievement and *assimilation* supported positive effects of comparisons to those with similar or lower averaged achievement (Marsh et al., 2008; Marsh & Parker, 1984; Marsh et al., 2000). Though the focus of the BFLPE concerns the negative, compositional effects of school-average achievement on student academic self-concept, its establishments have credited similar negative implications for corresponding achievement based on its reciprocal association with subject-specific academic self-concept, particularly in the context of

Figure 1

Big-Fish-Little-Pond Effect (BFLPE) Model



Note. From "The Big-Fish-Little-Pond Effect: Generalizability of Social Comparison Processes Over Two Age Cohorts From Western, Asian, and Middle Eastern Islamic Countries" H.W Marsh, 2015, *Journal of Educational Psychology,* 107(1), p. 259.

tracking, ability grouping, and gifted education" (Marsh et al., 2008, p. 321). By and large, in effort to contribute a better understanding academic self-concept BFLPE research has consistently demonstrated the presence of negative school compositional effect of school-averaged achievement on student self-concept attributed to social comparisons, while supporting the persistent positive and reciprocal relationship between students' achievement and students' academic self-concept (Huguet, 2009; Marsh & Parker, 1984; Marsh et al., 2007, 2008; Nagengast & Marsh, 2012; Pekrun et al., 2019; Seaton et al., 2009, 2010). Nonetheless, academic self-concept (ASC) is primarily at the heart of BFLPE investigations as the outcome of its models.

Academic Self-Concept

Construct Validity

Today self-concept is a thriving construct widely appreciated in psychology, social science, and education "not only as an outcome, but a mediating variable that subjectively facilitates the attainment of other desirable psychological and behavioral outcomes" (Möller et al., 2009, p. 1130). Self-concept is most generally recognized as a subjective mechanism of change for receptive individuals to attain their greatest potential and influences in future endeavors by contributing to a positive personality, behavior, emotional and cognitive well-being (Chiu & Klassen, 2010; Marsh et al., 2015; Marsh & Martin, 2011; Möller et al., 2009; Primavera et al., 1974; Rosenberg, 1989; Scheirer & Kraut, 1979; West & Fish, 1973). As one of the first major scientific studies of self-concept, Raimy (1948) reported the construct as contributing to the regulation of behavior, functional interrelationships of attitudes and the explanatory principle of personality" (p. 154). Theoretically, it was later validated as operating in the phenomenal sphere that was either directly present in or accessible to awareness (Rogers, 1951; Snygg & Combs, 1949). Thereafter, without sanction, the cognitive revolution reaffirmed the "traditional centrality of self in psychological literature through rapidly expanding research" (Coopersmith, 1959; Labenne & Greene, 1969; Purkey, 1970; Rogers, 1951; Rosenberg, 1989; West & Fish, 1973, p. 195; Wylie, 1979).

Early research reviewed by Wylie (1974, 1979) reported on developmental trajectories of self-concept and its relationships with age (Bachman & O'Malley, 1977; Kaplan & Pokorny, 1970), socioeconomic status (Soares & Soares, 1969; Trowbridge, 1972), race and ethnicity (Rosenberg & Simmons, 1972; Zirkel, 1971), as well as gender and achievement (Brookover et al., 1964; Primavera et al., 1974) among others. Yet, earlier views of self-concept results were said to be contradictory, confounded, and ambiguous (Byrne, 1984; Byrne & Gavin, 1996, p. 215; Hansford & Hattie, 1982; Shavelson et al., 1976; West et al., 1980; Wylie, 1974). The construct at that time was lacking clear structure conceptualization and operationalization (Bong & Skaalvik, 2003; Hansford & Hattie, 1982). Reviews concluded that inappropriate methodological applications of nebulous operational definitions as well as psychometrically inadequate instruments of measure contributed to sweeping inconsistencies and “widespread occurrences of null or weak findings” (Byrne & Shavelson, 1986; Crowne & Stephens, 1961; Shavelson et al., 1976; Wylie, 1974, 1979, p. 691). Therefore, Shavelson et al. (1976) requested a cessation of empirical research to address issues of construct definition and theoretical structure.

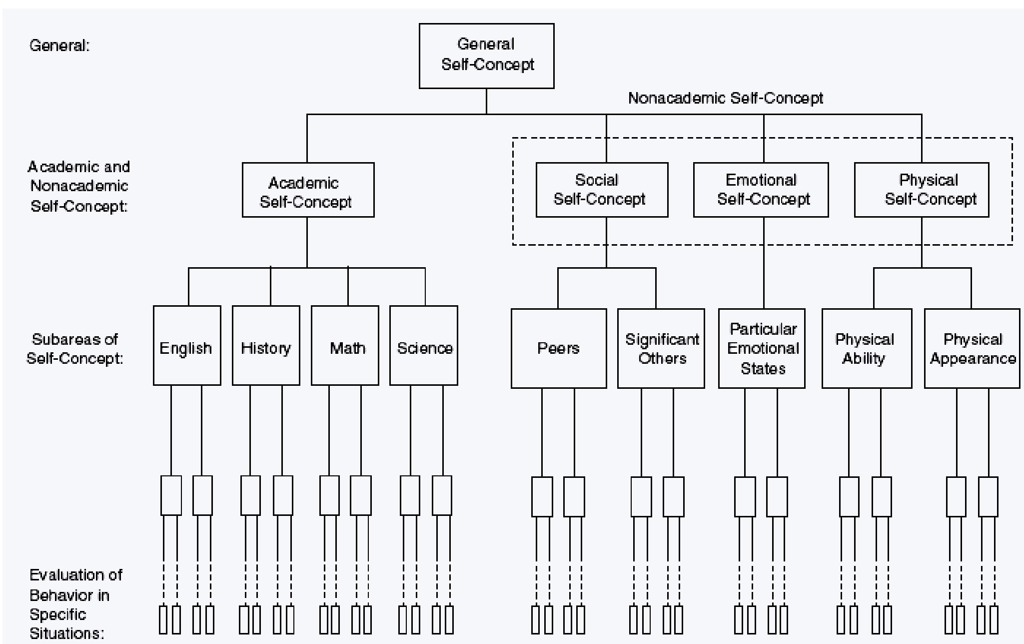
Notably, precursive definitions recognized self-concept as a “central intervening variable that mediated stimuli and behavior” (West & Fish, 1973, p. 4). It was operationally defined as the person's attitudes, feelings, and knowledge about his abilities, skills, appearance, competencies, and social acceptability (Byrne, 1984; Labenne & Greene, 1969, p. 10; Lecky, 1945; Rosenberg, 1989; West & Fish, 1973, p. 4). Additionally, such self- perceptions were understood to derived from social environment that directed behavior and subsequent self-perceptions (Epstein, 1973). Albeit, these definitions were generally accepted, but later reviews of the literature reveal no clear, concise, and universally adopted definition (Byrne, 1984, p. 428; Crowne & Stephens, 1961; Hansford & Hattie, 1982; Labenne & Greene, 1969; West & Fish, 1973; Wylie, 1974). Consequently, Shavelson et al., (1976) set out to converge commonalities consistent with then present research to develop a working definition that could be used to integrate empirical evidence and validate self-concept interpretations.

Shavelson Model. Aligned with Cooley (1902) and Mead (1934) earliest notions, their study concluded that “that self-concept is a person's perception of himself that is formed through interpretations of his experience with his environment, especially environmental reinforcements and significant others (Shavelson et al., 1976, p. 411). Moreover, Shavelson et al. (1976) offered a universally acceptable operational definition of the construct incorporating seven defining characteristics of the self-concept including that it was organized, hierarchical, stable, developmental, evaluative, multifaceted, and differentiable" (see Figure 2). Specifically, in agreement with Mead (1934) symbolic interactionism theory that assigned labels to culturally related behavioral expectations, the nature of self-concept applied a *categorization* process (Scheirer & Kraut, 1979, p. 141) based on intrinsically identified social and physical categories that often reflected cultural beliefs. Such categorization, provided as way of "organizing experiences to give them meaning" alluding to its *organized* and *multifaceted* characteristics (Shavelson et al., 1976, p. 412, p.). Likewise, its *evaluative* characteristic applied processes of comparison and evaluation were derived from Cooley (1902) and Festinger (1954) notions of social comparison whereby self-labels were intrinsically compared and ranked relative to significant others within the social sphere.

Ultimately, affective or emotional attitude towards the self were often referred to as self-esteem (Bong & Skaalvik, 2003; Coopersmith, 1959; Rosenberg & Simmons, 1972; Scheirer & Kraut, 1979, p. 141). Furthermore, *developmental* characteristics suggested "with increasing age and experience, self-concept becomes more evolved" (Kohlberg, 1969; Sears & Sherman, 1964; Shavelson et al., 1976, p. 414). Lastly, incorporating James (1890) notions, the final attributes of self- concept characterized it as *hierarchical*, *stable*, and *differentiable.* Specifically, general self-concept (GSC) was placed at the apex as the second order factor as it was most stable and more correlated with the first order factors termed academic and nonacademic factors that follow. Yet, it was less correlated with the subject-specific, lowest order factors of subdivided facets of ASC such as math, science, English, and history or nonacademic facets such as social peers and significant others, specific emotional states, or physical appearance and ability (Byrne, 1984; Shavelson & Bolus, 1982; Shavelson & Stuart, 1981).

Figure 2

Shavelson (1976) Model of Self-Concept



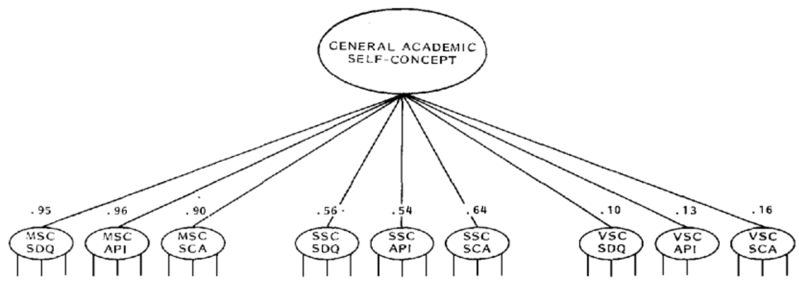
Note. From “Self-concept: Validation of Construct Interpretations” by R. J. Shavelson, J.J. Hubner, and G.C. Stanton, 1976, *Review of Educational Research,* 46(3), p.413.

However, though they were able to confirm discriminatory validity to differentiate it from other constructs, convergent validity among its facets was unable to be confirmed by any instruments available at that time. Nevertheless, consistent with Cronbach (1972) the Shavelson model set criteria for operationalizing the construct through measures of internal consistency, as well as applications of analytical methods such as factor analysis, multimethod-multitrait, and path analysis to confirm the convergent and discriminant validity of its nomonological network. As a matter of fact, Shavelson seminal research not only provided the operational definition and necessary framework for validity in instrumentation and interpretations of the self-concept construct, but also significantly influenced the future of its theoretical progression and applications to date.

Unidimensional Model. The Shavelson model's unfounded results of convergent validity were effectively challenged by Winne et al., (1977). Their ideas, like other's prior, resembled the Spearman (1904) Model of Intelligence and proposed a unidimensional structure wherein the general factor of academic self-concept overwhelmingly dominated the construct rendering it undifferentiable from other possible aspects (Coopersmith, 1967; Epstein, 1973; Rosenberg & Simmons, 1972). **S**pecifically, it was described as a unitary construct not hierarchically subdivided, but instead "structured like a daisy, whereby much of the construct is shared and undifferentiable, but individual petals or facets may be more or less relevant when related to other constructs such as achievement" (Marx & Winne, 1978, p. 100) (see Figure 3). Coincidingly, Soares & Soares (1983) refuted a hierarchical structure originating from the dominant influence of a general or temporal self and instead proposed a taxonomic structure having a collaboration of cognitive and behavioral dimensions with varying self-perceptions that are influences by social and situational experiences. However, upon further review, the unidimensional or taxonomic model was later refuted attributing it to problems in measurement and analysis typical of that time (Marsh & Hattie, 1996; Marsh & Craven, 2006).

Figure 3

Unidimensional Model of Self-Concept



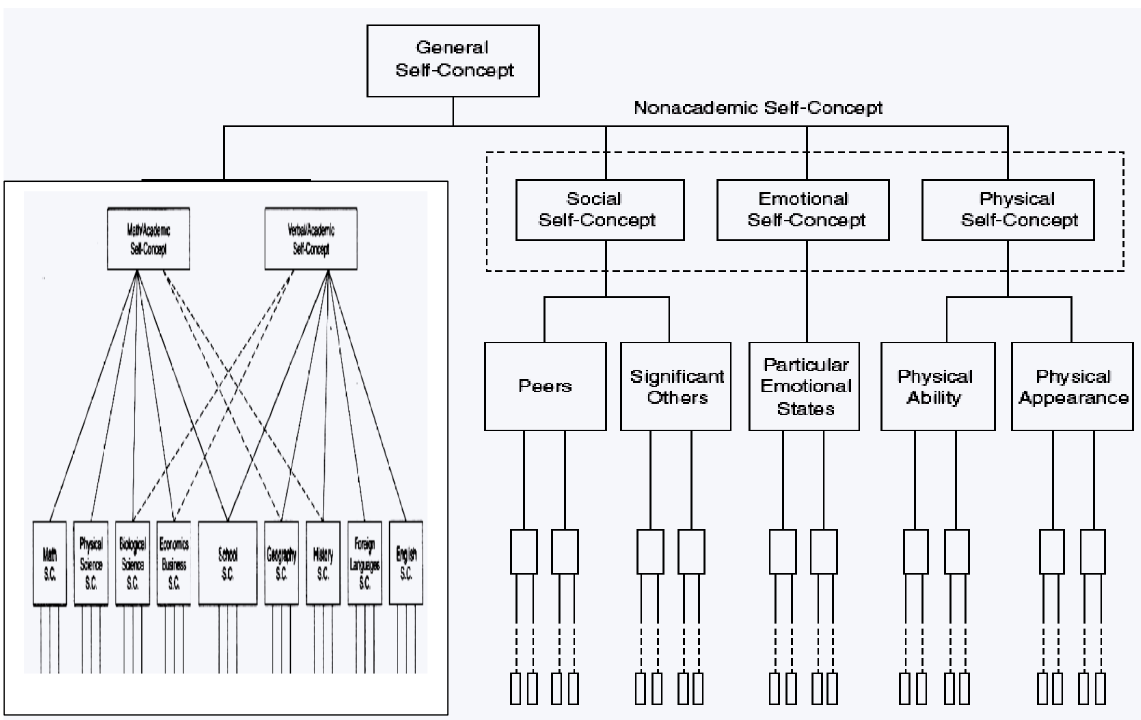
Note. From *"A multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement* by H.W. Marsh, B.M. Byrne, R.J. Shavelson,1988, *Journal of Educational Psychology,* 80(3), p. 371.

Marsh and Shavelson Model. Despite opposing views, the framework of the Shavelson (1976) model prevailed. Through the application of internal consistency measures and factor analysis to validate instrumentation designed to measure the internal facets of self-concept, the multidimensional, hierarchical, and comparative structure of self-concept was increasingly supported, especially by the research of Herbert Marsh (Byrne, 1984; Huang, 2011; Marsh & Hattie, 1996; Möller et al., 2009; Valentine et al., 2004). In fact, he argued that the determination of theoretically consistent and distinguishable domains of self-concept should be prerequisite to the study of how self-concept is related to other variables (Marsh, 2006, p. 6). His earlier research implemented an empirically designed instrument known as the Self-Description

Questionnaire (SDQ I, II, and III) to successfully replicated the relations between subject-specific facets and academic facet with strong evidence complementing the idea that developmental trajectories weakened the hierarchical structure at its lower levels due in part to increasing independence of facets (Marsh et al., 1983; Marsh et al., 1983; Marsh & O'Neill, 1984).

Figure 4

Marsh and Shavelson Model of Self-Concept



Note. The box added to the Shavelson 1976 Model figure represents the revisions by Marsh & Shavelson (1985). From "*Self-Concept Theory, Measurement and Research into Practice: The Role of Self-Concept in Educational Psychology" by* H.W Marsh, 2006, The British Psychological Association, p. 14.

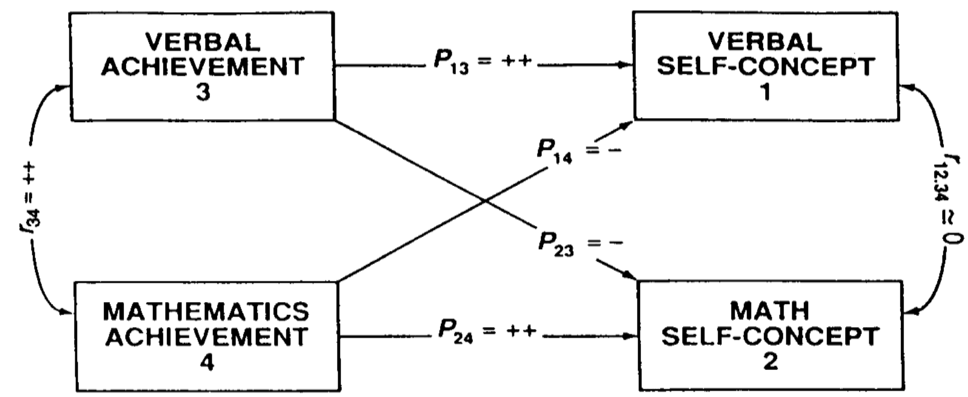
However, the Marsh and Shavelson (1985) revisions determined that the subject-specific academic self-concept facets were actually divided into two, uncorrelated factors labelled as math academic self-concept and verbal academic self-concept, whereby history and science was placed along a continuum between the two (Marsh, 1990; Marsh et al., 1988; Marsh & Shavelson, 1985) (see Figure 4). Moreover, the results not only contributed the Academic Self-Description Questionnaires (ASDQ I and II) as a valid instrument of measure, but also concluded that the academic self-concept was based on external comparisons of their ability with other students' and internal comparisons of their ability between subject domains (Marsh, 1990; Marsh et al., 1988; Marsh & Shavelson, 1985, p. 121).

Rationale Models

I/E Model. At the time of the Marsh and Shavelson (1985) revision, Marsh and Parker (1984) research regarding the internal comparison process that was similar to Byrne (1984) compensatory model examined the uncorrelated first order math and verbal facets. Consequently, influenced by Snygg and Combs (1949), Marsh (1986) proposed the Internal/External (I/E) frame of reference model that suggested students form their subject-specific, academic self-concept from internal comparisons with their own performance in the same and other academic domains, whereby positive correlations existed between ASC and achievement in the same subject domain, while negative correlations existed between ASC and achievemnt in different subject domains (see Figure 5). For instance, if students have high achievement scores in math, they would reflect high ASC in math, but low ASC in verbal domains and vice versa for high verbal achievement (Bong, 1998; Marsh, 1990; Shaalvik & Rankin, 1992). Accordingly, Jansen et al. (2014) reports substantial research that supported the I/E model through longitudinal, cross -cultural, and metanalytic studies (p.11) (Marsh et al., 2001; Marsh & Hau, 2004, 2004; Marsh & Köller, 2004; Moller et al., 2009, 2011).

Figure 5

Internal/External (I/E) Model of Academic Self-Concept



Note. From "*Self-Concept Theory, Measurement and Research into Practice: The Role of Self-Concept in Educational Psychology"* by H.W Marsh, 2006, The British Psychological Association, p. 42.

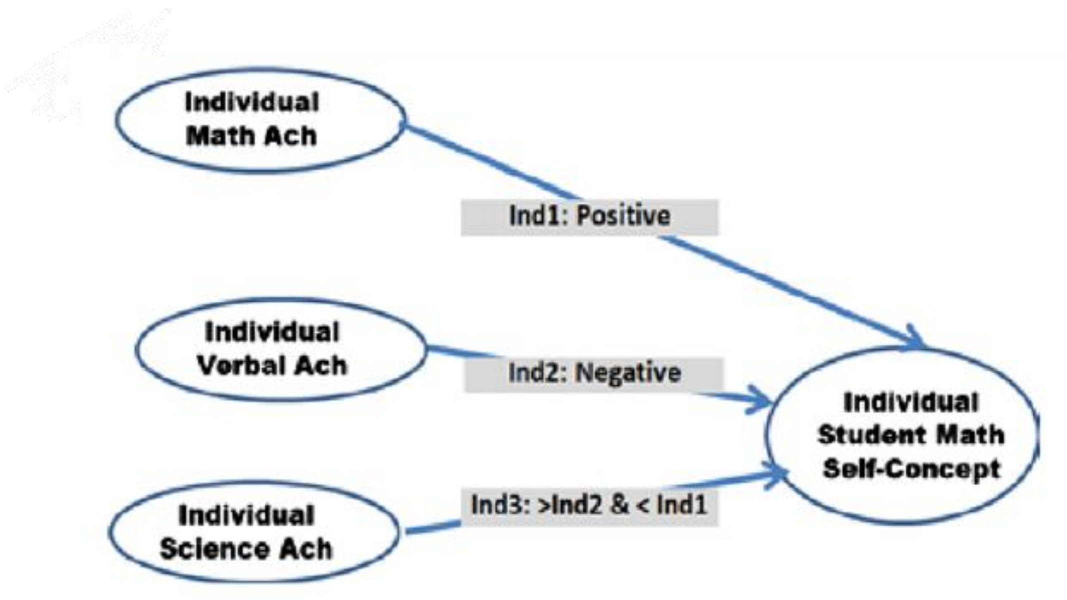
Correspondingly, more recent revisions of the I/E model not only incorporated domains other than math and verbal (Moller & Koller, 2001), but also incorporated the study of how social (Festinger, 1954), temporal (Albert, 1977) and dimensional comparisons independently and interdependently influenced the establishment of overall academic self-concept (Moller & Marsh, 2013). Generally, revised I/E models such as BFLPE illustrate that social comparison processes affect self-concept when performance is compared to others' performance in the same subject domains, (Festinger, 1954; Marsh et al., 2008; Marsh & Parker, 1984), temporal comparison affect academic self-concept when present performance is compared to past performance (Albert, 1977; Moller & Koller, 2001), and dimensional comparison theory (DCT) models illustrate effects on self-concept from internal comparisons along a subject-domain continuum (Marsh et al., 2015; Moller & Koller, 2001; Moller & Marsh, 2013) (see Figure 6).

Causal Relationships. Academic self-concept (ASC) has been recognized by numerous studies and meta-analyses for its positive relationship with academic achievement (Bandura, 1994; Byrne, 1998; Guay et al., 2003, p. 200; Hansford & Hattie, 1982; Huang, 2011; Marsh, 2005; Marsh & Craven, 2006; Marsh & O'Mara, 2009; Moller et al., 2009; Scheirer & Kraut, 1979; Valentine et al., 2004; West & Fish, 1973; Zimmerman et al., 1992). Generally, ASC has shown to positively predict level of academic attainment and course selection (Guay et al., 2003; Marsh & Yeung, 1997), intrinsic motivation and persistence (Guay et al., 2010; Skaalvik & Rankin, 1992) as well as STEM course selection (Parker et al., 2014).

Correspondingly, research examining achievement has shown an overall positive relationship with domain-specific academic self-concept as well. Comprehensively, previous research has reported that ASC positively predicted academic achievement (Moller et al., 2011; Moller & Marsh, 2013), while conversely achievement reportedly shown to positively predict academic self-concept as well (Calsyn & Kenny, 1977) with reciprocal relationships also having been reported (Byrne, 1984; Marsh, 1990). Furthermore, Marsh et al. (2007) review of longitudinal studies suggested causal relationships wherein prior academic self-concept had a positive effect on subsequent academic achievement beyond what could be explained by prior achievement. Nevertheless, Marsh (1990) had argued for the inappropriateness of such a comparison in that both paths are substantively important no matter which one is larger, while Scheirer & Kraut (1979) review of experimental self-concept interventions revealed no support that changes in ASC caused changes in achievement attributing weak theory and methodological limitations of the time to explain otherwise. Ultimately, speculation of causality between academic achievement and ASC adopted the reciprocal effects model as the established paradigm of causality (Byrne, 1984; Helmke & van Aken, 1995; Lui & Meng, 2010; Marsh, 1990; Marsh et al., 1999).

Figure 6

Dimensional Comparison Theory (DCT) Model



Note. From *"Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: Pan-Human Cross-Cultural Generalizability over 68 Countries" by* Marsh, 2020, European Journal of Personality, 34, p. 180-202.

Though educational research has extensively substantiated the positive relationship between ASC and achievement (Ma & Kishor, 1997; Marsh, 1986; Reyes, 1984; Shavelson & Bolus, 1982), negative relationships have also been established (Bong & Skaalvik, 2003; Hansford & Hattie, 1982; Huang, 2011; Wilkins, 2004). For instance, Huang (2011) path analysis found that ASC and academic achievement effects were mixed. Whereas results supported the effect of prior ASC on subsequent academic achievement, the magnitude and significance of not only the effects of prior academic achievement on subsequent ASC, but also those reciprocal relations depended greatly on the number of predictors in the regression model (p.525). By and large, a great debate existed concerning the endorsement of conflicting theories of causality arguing for self enhancement, skill development, or reciprocity to determine whether or not academic achievement is important in the formation of ASC, whether ASC influences subsequent achievement, or whether they are mutually beneficial.

Early on, Byrne (1984) suggested that three conditions were necessary to make causal inferences including the establishment of a statistical relationship, time precedence, and nonspuriousness from undue influence of unforeseen extraneous variables. However, like Wylie (1979) earliest review of self-concept research, a later review identified major concerns of research on causality whereby differences in design, varying measurements in the number of time points, grade level, time intervals, assessment of ASC by multiple versus single item assessments, assessment of achievement by cumulative grade point averages or grades versus achievement test scores, as well as issues of subject domain (Marsh, 1986; Shavelson & Bolus, 1982) were indicated (Helmke & van Aken, 1995, p. p.625).

Self Enhancement Model. The self-enhancement model illustrated a positive relationship between self-concept and achievement wherein a higher self-concept would cause a subsequently high achievement (Calsyn & Kenny, 1977). With implications for emphasis on school policy to increase self-concept as a means of addressing achievement differences, Shavelson & Bolus (1982) investigated causal predominance using structural equation modeling including three hierarchical measures of self-concept and three subject-specific achievement measures of junior high school students. In all instances, the dominant causal link originated from either ASC or subject-specific self-concept and achievement measures. Though early results were disputed based on flawed methodology, Marsh (1990) reanalyzed results of the Youth in Transition study and reported grade averages in Grades 11 and 12 that were significantly affected by previous academic self-concept measures with no effect of prior grades on subsequent measures of academic self-concept. As well, Byrne (1998) provided support for the causal effects of prior academic self-concept on subsequent achievement in English with causal links reversed in Math and Science, suggesting subject-specific disparity relating to Marsh's I/E model of ASC.

Skill Development Model. The skill development models portrayed self-concept as the consequence of academic achievement with implications for policy to emphasize curriculum structure and content over self-concept development as a means of improving achievement (Calsyn & Kenny, 1977). In contribution, Skaalvik and Valås (1999) examined math and verbal scores of three-cohorts of Norwegian primary and middle school students and found that prior math achievement significantly affected prior self-concept in all three cohorts. Though Helmke and van Aken (1995), Lee and Kung (2018) and Burns et al. (2020) found evidence of reciprocal relationships in math, the skill-development model had a stronger effect in all those studies. Specifically, Helmke and van Aken (1995) reported that "later achievement depends almost completely only on prior math achievement and not on prior math self-concept," regardless of whether grades or standardized test scores were included as achievement measures for early elementary students (p.634). Also, Lee and Kung (2018) reported similar results for high school Taiwanese students and recommended investigating potential intervening variables that mediated the longitudinal effects on achievement. However, Burns et al. (2020), scrutinized results of the reciprocal effects model (REM) as it often has adopted cross lagged panel designs that do not account for between-person variances. Subsequently, they applied a random intercept, cross lagged panel design that examined such variance for 1st year undergraduate students that only supported the skill development model.

Reciprocal Effects Model (REM). Marsh (1990) contributed the widely adopted "either or" debate about the causal predominance of ASC and ACH to limitations in statistical techniques available at that time. To subside the debate, he offered substantial evidence using multiple indicators of achievement in English, math, and science across 3-year span for Catholic school boys. His results reflected a reciprocal relationship whereby prior ASC significantly affected subsequent achievement and prior achievement significantly affected subsequent ASC. In culmination, Marsh et al. (1999) review refuted Byrne (1984, 1986) support for a null model and rather confirmed the consistency of reciprocal relationships regardless of obvious methodological limitations. In conclusion, they offered an updated version of ideal standards of REM studies that should include the measure of academic self-concept and academic achievement in at least two time points, multiple indicators for all latent constructs, sufficiently large and diverse sample to justify the use of confirmatory factor analysis (CFA) and generalization of results, as well as data to fit a variety of CFA models that would incorporate measurement error and test for likely residual covariation among measured variables (p.161).

Henceforth, subsequent research demonstrated support for the REM model of ASC and achievement with evidence of discriminant validity regarding relations to self-esteem (Marsh & Craven, 2006) as well as generalizability across different student characteristics such as age (Guay et al., 2003), gender (Lee & Kung, 2018; Marsh et al., 2005; Marsh & Yeung, 1997), achievement tracking (Arens et al., 2017; Seaton et al., 2014) and diverse cultures (Marsh & Martin, 2011, p.67), while Valentine et al. (2004) found no support for country as a moderator of REM.

Overall, REM findings synthesize the implications of self-enhancement and skill development models to recommend the implementation of a dual intervention approach to enhance both self-concept and achievement as mutually beneficial to a wide array of student populations. Subsequently, today REM is the most widely applied theoretical model of causality for the study of the relationship between ASC and ACH. Importantly, implications of REM can be applied to BFLPE results such that results for academic self-concept can be extended to encompass corresponding subject-specific achievement.

Big Fish Little Pond Effect (BFLPE)

Social Comparisons

Distinguishably, Festinger (1954) Social Comparison Theory has laid the framework for most frame of reference studies. Whereas his theory asserted that “humans have an innate desire to evaluate one’s own opinions and abilities in comparison to the opinions and abilities of others " (p. 118), subsequent theories have applied the concept of social comparisons in a more specific context. For instance, Davis (1966) Frog Pond Effect supported the theory of Relative Deprivation that suggested a detrimental effect on self-evaluations based on comparisons of those more successful. Based on this notion, in his investigation of college students' career aspirations, he concluded that those attending high-ability schools would have lower GPAs. Conversely, he reported that GPA rather than school-quality would determine career aspirations due to the influence of comparisons with others' GPA. He suggested as well that such comparisons influenced one's perception of academic ability (ASC) over and above the influence of school quality, thereby demonstrating a negative compositional effect of individual- and group-level comparisons (p. 30).

As such, "the empirical BFLPE theory was simply a specific example of simikar frame-of-reference effects that have been studied in psychology" (Marsh, 1984, p. 281). By applying similar concepts to that of social comparison theory (SCT) and Davis's Frog Pond Effect, BFLPE contended that "students apply social comparisons to the average of academic accomplishments of other students within their school to form a frame of reference against which to evaluate their own academic accomplishments" (Marsh et al., 2008, p. 324). Specifically, if equally able students attended different ability schools, they would each have different levels of self-concept based on their local frame of reference comparisons. This pointed out that the major difference between SCT and BFLPE was in the choice of generalized other that was less specific in SCT than that of localized, social comparisons to classmates as in the BFLPE.

Furthermore, Marsh et al., (2000) highlighted the possibility of effects of local frame of reference comparisons that demonstrated deprived effects as well as counterbalancing effects such as that of Cialdini (1976) theory of Basking in Reflected Glory whereby an *assimilation* effect represented students desire to reflect those that were more successful. Their research found evidence for assimilation in that there was a positive effect of school status on self-concept for high school students in Hong Kong that which only exacerbated the contrast effect of school-achievement and self-concept. However, Hughet et al. (2009) argued that direct social comparisons in BFLPE were only implied, inferring that the exact reference group chosen by each student was indefinite (p. 164). Resultingly, their research concurred with the simultaneous presence and subsequent counterbalancing effects of both *contrast* and *assimilation* effects, but also suggested that "it seems reasonable to conclude that beyond the relatively uncontrollable comparisons underlying the BFLPE, students may still exercise considerable choice over the target with whom they compare themselves, with sometimes a beneficial effect on their academic self-concept" (p. 165). Similarly, Marsh and Hau (2000) and Marsh et al. (2000) reported the counterbalancing of an increased negative effect of upward comparison with an increased sense of pride for students placed in a high averaged achievement classroom.

Generalizability

Overall, BFLPE studies have substantially supported the generalizability of the negative effects of school-averaged achievement on students' academic self-concept when controlling for individual ability effects across a variety of controlled variables such as student-level socioeconomic status (SES), gender, grade point average, aspirations, other self-beliefs, and academic interests (Marsh et al., 2008; Marsh & Parker, 1984; Marsh & Craven, 2002; Marsh & O'Mara, 2009; Pekrun et al., 2019; Plieninger & Dickhäuser, 2015; Seaton et al., 2009a, 2010; Trautwein et al., 2006). For instance, using IQ and SDQ results, Marsh and Parker (1984) seminal BFLPE study applied analysis of variance design for a sample of sixth grade Australian students initially to explain the uncorrelated academic domains of the Shavelson (1976) model of ASC, but also found that lower income and lower ability schools had a higher academic self-concept than higher income and higher ability schools with larger negative effects found after controlling for individual level SES and ability. Supplementarily, using more complex structural equation models with longitudinal data, Marsh and O'Mara (2009) longitudinal study was able to show that long term total negative effects of BFLPE were more negative than direct effects across SES, grade-point average (GPA), as well as occupational and educational aspirations. Additionally, Seaton (2010) found BFLPE to generalize across 16 student characteristics, including those of academic self-regulation and SES by applying multilevel modeling with PISA 2003 data.

By the same token, Pekrun et al. (2019) in three studies that applied multilevel modeling and longitudinal designs over the course of one year for German fifth and tenth graders, reported negative compositional effects of class-averaged achievement on individual level self-concept in math across gender and grade-level (p. 172). They reported as well that "individual achievement positively predicted enjoyment and negatively predicted anger, anxiety, and hopelessness, whereas class-average achievement had negative compositional effects on enjoyment but positive compositional effects on negative emotions" (p. 174). Comparatively, Marsh (2014) applied multilevel latent variable models of cross-cultural data and confirmed that the BFLPE generalized across gender for Saudi Arabian and American students using TIMSS 2007, though Saudi girls did outperform Saudi boys in math. Of the few contrasting reports, Plieninger and Dickhäuser (2015) applied multilevel modeling to examine a German sample of PISA science data with results that indicated a lower self-concepts in females when controlling for individual achievement.

Furthermore, BFLPE was also found to generalize across 32 countries for other self-beliefs such as general self-efficacy, control expectations, control strategies, and effort persistence (Marsh et al., 2008). Likewise, Trautwein et al. (2006) confirmed that self-concept and academic interests such as intrinsic value, personal importance, and attainment value, were "negatively predicted by school-averaged achievement and positively predicted by student-averaged achievement with students in lower tracks showing more academic interest than those in higher tracks" (p. 803).

Nonetheless, multiple studies have supported cross-cultural and longitudinal generalizability of BFLPE as well (Marsh et al., 2019; Marsh et al., 2020; Marsh & Hau, 2003; Marsh et al., 2007; Marsh & O'Mara, 2009; Pekrun et al., 2019; Seaton, 2007; Seaton et al., 2009). For example, in the first large-scale, international study using PISA data for 26 countries, Marsh and Hau (2003) reported pan-human validity of BFLPE proposing further research to investigate "student-level characteristics that predict students who may benefit from academically selective schools" (p. 375). Likewise, Seaton (2009) argued that most BFLPE research had been conducted in individualistic and developed countries, so reported generalizability of BFLPE using PISA data across 41 countries. Marsh (2008) was able to generalize BFLPE across 26 countries using PISA

2000 data to confirm that school-averaged achievement had negative effects on ASC with little effect on other self-belief constructs. Also, Marsh et al. (2007) reported German high school students that the negative compositional effects of school-averaged achievement were pervasive at the end of high school and continued two and four years later. Likewise, a longitudinal study of high school sophomores, seniors, and students two years after high school, demonstrated the pervasiveness of BFLPE on academic outcomes such as standardized test scores, self-concept, coursework selection, academic effort, school grades, educational and occupational aspirations, and college attendance. His research highlighted that though the effect sizes were small, no positive effects of BFLPE were found to be significant for the any of the academic outcomes measured.

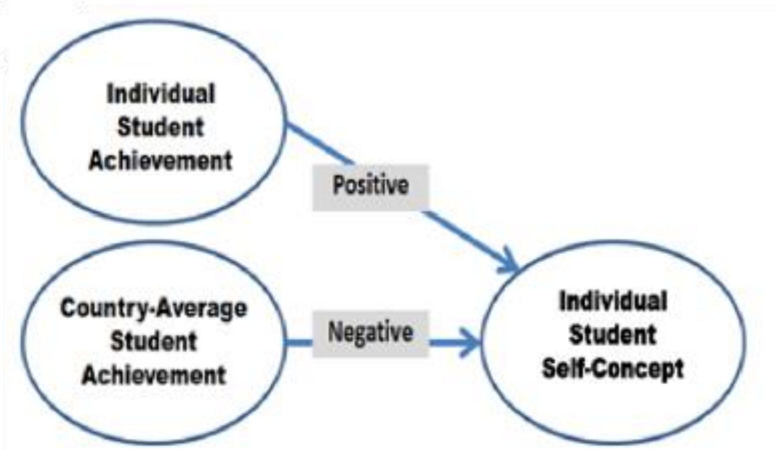
Accordingly, Marsh (2008) recognized BFLPE as an "inherently multilevel phenomenon" (p. 324) wherein "contextual effects matter…not only at the student- and local school level, but remarkably even at the macro-contextual country-level" (Marsh et al., 2019, p. 231; Marsh et al., 2020). Specifically, Marsh et al. (2019) reported that at the student-level studies showed cross-cultural generalizability of BFLPE across 68 countries for year in school relative to age, whereby due to "relative position within a given frame of reference" there was a negative effect of class-averaged achievement on "de facto acceleration such as starting early or skipping a grade" and positive effects for "de facto retention such as starting late or repeating a grade" (p. 233). Similarly, they found negative effects on ASC not only from comparisons to school-averaged of achievement, but also similar negative effects on ASC from comparisons to country-averaged achievement (see Figure 7). Resultingly, they contributed an explanation for the previously reported paradoxical cross-cultural effect whereby students in high achieving countries showed lower self-concept then those from lower achieving countries (Chiu & Klassen, 2010; Lui & Meng, 2010; Marsh et al., 2014; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004).

Expressively, this cross-cultural paradox has been exemplified as students from the USA having lower country-level achievement, yet higher student-level ASC then students from East Asia who have higher country-level achievement (Marsh et al., 2019, p.233). Subsequently, Marsh et al. (2020) has attributed the "paradoxical country-level frame of reference effects" to Bronfenbrenner's ecological model that posits student-level attributes are influenced by proximal processes of the immediate social and cultural environment during development (p. 187). Conclusively, these studies recommended further examination of the cross-cultural generalizability of the contextual effects of country-frame of reference effects on ASC.

Collectively, results of BFLPE research not only contribute to the understanding of how classroom or school level achievement affects students' perception of their academic ability and corresponding achievement, but also it provides "important policy implications for academic tracking, streaming and selective schools" (Marsh et al., 2008, 2014, p. 796; Marsh et al., 2008). Accordingly, Hattie (2002) clarified that tracking was a means of minimizing differentiation in the classroom, wherein teachers can address collective needs and refine instruction accordingly, while Byrne (1988) contested that such an organization inadvertently nurtures social comparisons (p. 46). However, "the nature and extent of tracking differs across countries, states, and/or school districts, making the term somewhat ambiguous." Moreover, school-level organizational practices were reportedly categorized as either explicit, between school-level tracking or implicit, within school-level tracking (Salchegger, 2016; Trautwein et al., 2006, p. 789; Wouters et al., 2012). Whereas explicit between-school tracking is more stable and caters to specific populations of students often according to homogeneous achievement and coinciding future career paths, implicit within-school tracking is mostly malleable and implements the streaming of students into achievement based, homogenous classrooms or clusters of a schools that caters to a variety of academic abilities and opportunity.

Figure 7

Paradoxical Cross-Cultural BFLPE Model



Note. From *"Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: Pan-Human Cross-Cultural Generalizability over 68 Countries" by* Marsh, 2020, European Journal of Personality, 34, p. 180-202.

Dominantly, BFLPE results discredit previous notions of the positive assimilation effects of tracking on students ASC and according to the reciprocal effects model, subsequent achievement as well (Dicke et al., 2018; Marsh et al., 2008a; Marsh et al., 2008; Trautwein et al., 2006). However, studies have contended explanatory stipulations. For instance, Byrne (1988) found greater negative effects of BFLPE when comparisons were made with students from a higher track rather than immediate classmates, Marsh et al., (2015) determined a local dominance effect with more impact from proximal comparisons with classmates rather than school-wide, and Hattie (2002) found no effect of school grouping on student learning outcomes supporting Arens et al., (2017) notion that attributed differences to disparity in the quality of instruction for low and high ability tracks. As well, Salchegger (2016) found no effect of implicit, within-school tracking, but a more pronounced BFLPE in schools with earlier explicit tracking practices. Regardless, "although BFLPE does predict the negative effect of school-average ability, it clearly does not assume that this is the only influence on student ASC or achievement" (Marsh et al., 2008, p. 335).

Suggested Improvements

Indeed, as Marsh (2008) concurred, "significant moderators of the BFLPE have important implications in terms of better understanding the BFLPE and how to ameliorate its negative consequences for students" (Seaton, 2010, pg. 423) as well as improve the current state of policy and instruction, yet available research and generalizable evidence for such is minimal and inconclusive. Therefore, as a result of such a narrow scope of focus and few alternative implications of the BFLPE, recommendations for future research commonly advised the consideration of additional contextual variables of interest as well as methodological improvements (Dai & Rinn, 2008; Huguet, 2009; Marsh & Parker, 1984; Marsh et al., 2000, 2007, 2008; Marsh & Hau, 2003; Marsh & O'Mara, 2009; Morin et al., 2014; Nagengast, & Marsh, 2011; Pekrun et al., 2019; Seaton et al., 2009, 2010; Wang, 2015). Broadly, the inclusion of student characteristics and differences are suggested to understand influences on covariates at the individual level (Marsh & Hau, 2003; Pekrun et al., 2019; Seaton et al., 2009). For the same reason, Huguet (2009), Marsh et al. (2008) and Seaton et al. (2010) endorsed other measure of perception and emotional underpinnings be included as well. Although the negative effects of classroom or school-averaged achievement on student ASC has been found to generalize over gender as well, Marsh and O'Mara (2009) concurred that it should continue to be analyzed as well as the inclusion of science self-concept too (Marsh et al., 2007).

Furthermore, Marsh and Hau (2003) and Seaton et al. (2009, 2010) favor the incorporation of contextual variables that impact school performance be factored into future BFLPE research. Specifically, measures of school SES, expenditures, resources, and climate are suggested (Marsh & Parker, 1984; Marsh & O'Mara, 2009; Seaton et al., 2009a). On the other hand, Dai and Rinn (2008) directs future research to focus not only on proximal influences from school indicators, but also distal influences from country indicators such as cultural and economic dimensions as it is indeed "possible to have multiple reference groups for social comparison that extend beyond their local ponds (p. 293)." Moreover, methodological improvements are favorably recommended as well. Though causal inferences among indicators of BFLPE are in high demand, longitudinal data is needed yet not always available (Nagengast & Marsh, 2012; Seaton et al., 2010).

Additionally, being that prior research has often conducted analyses using correlational, multilevel or structural equation modeling designs, researchers have prompted more complex and sophisticated approaches. Specifically, Nagengast and

Marsh (2012) suggested that the BFLPE model should be explicitly specified at three levels. As well, Dai and Rinn (2008) and Seaton et al. (2010) requested additional designs to use moderation and mediation to contribute a better understanding of implications to improve theory, policy and practice. As such, it is the intent of this study to highlight and apply the valuable implications offered by previous self-concept research in attempt to identify potential moderators and enhance current BFLPE models.

Moderation

Granted, many studies have clearly confirmed the negative effects of BFLPE, there have been so few studies reporting contradictory results. For instance, Huguet (2009) demonstrated the coexistence of both *contrast* and *assimilation* effects resulting from BFLPE (Festinger, 1954; Marsh & Parker, 1984). Correspondingly, Marsh et al. (2008) study found that ability classmates may have negatively affected student academic self-concept, but did not affect student metacognitive responses. Most notably, Marsh et al. (2000) and Marsh and Hau (2003) reported the counterbalancing of an increased negative effect of upward comparison with an increased sense of pride for students placed in a high averaged achievement classroom. Even fewer studies concern effects occurring at the country-level such as those regarding cultural differences and level of economic development to support the cross-cultural generalizability of BFLPE found by Seaton et al. (2009).

Resultingly, as a "means for policy and practice to minimize BFLPE's negative impact and maximize the positive impact" on student academic self-concept, compelling research has investigated the underlying mechanisms and intervening factors that moderate the magnitude or direction of the negative effects of school-averaged and country-averaged achievement (Cheng et al., 2014; Dai & Rinn, 2008; Jonkmann et al., 2012; Marsh et al., 2020; Schwabe et al., 2019; Seaton et al., 2010). Correspondingly, in a critical review of BFLPE research, Dai and Rinn (2008) asserted that Herbert Marsh's studies inappropriately prioritized the social comparison aspects and neglected to "incorporate contextual, developmental, and individual differences" as potential moderators (p. 283). Marsh et al. (2008) in his response to that article, contends that his "model does not posit that individual and school-averaged ability are the only determinants of ASC" and encourages the investigation of "contextual characteristics that moderate the size" (p. 323). Albeit, subsequent moderation studies have generally investigated mostly proximal influences of individual characteristics rather than distal influences from classroom or school with very few significant results have been reported (Jonkmann et al., 2012; Wouters et al., 2015).

Notably, moderation studies that have investigated the influence of student characteristics such as performance goals, personality traits, gender, sense of belonging, ability on BFLPE, have had the greatest contribution to the research, yet effect sizes are unanimously small (Cheng et al., 2014; Jonkmann et al., 2012; McFarland & Buehler, 1995; Plieninger & Dickhäuser, 2015; Schwabe et al., 2019; Seaton et al., 2010; Wouters et al., 2015). For instance, students' performance goals and goal orientation have showed significant moderation, but motivation and engagement exacerbated the contrast effects of classroom-level BFLPE (Cheng et al., 2014; Wouters et al., 2015). As well, Jonkmann et al. (2012) reported on the moderation of Big 5 personally traits, reporting that narcissism reduced the BFLPE, while neuroticism increased the negative effects on ASC. Plieninger and Dickhäuser (2015) suggested that females were more substantially impacted by the BFLPE as they relied more on comparisons with classroom performance, while Marsh et al. (2020) rejected the "bright student hypothesis" such that higher ability students reflected lower ASCs rather than being immune to the effects of BFLPE as the hypothesis suggests (p. 186).

Additionally, in an extension of Seaton (2007) doctoral dissertation, Seaton et al. (2010) investigated 16 student characteristics including student SES as measured by composite home and parent variables, academic self-regulations such as motivations, self-efficacy, learning strategies implored, social orientations such as sense of belonging, as well as emotional dimensions. Though recognized as the most extensive BFLPE moderation study at that time, results found BFLPE to generalize across most variables, though a stronger negative BFLPE effect was discovered for students who utilized surface learning strategies, experienced high anxiety, and implemented cooperative social orientations (p.424). On the other hand, even as examinations of classroom compositional effects were scarce, Schwabe et al., (2019) examined moderation of instructional practices by comparing individualized and collectively differentiated instructional practices with collective approaches, but no significant change in magnitude or direction of BFLPE were reported.

Contrastingly, mediation studies have also explained the underlying mechanisms of BFLPE. For instance, Marsh et al. (2008) reported a lesser negative effect of school-averaged achievement (L2ACH) for students' academic self-beliefs (self-efficacy, control expectations, control strategies, and effort persistence) than for students' academic self-concept (L1ASC). By exploring relationships from a unique angle, Nagengast and Marsh (2012) confirmed academic self-concept as a mediator of negative effects of L2ACH on career aspirations, while Pekrun et al. (2019) suggest that academic self-concept mediates the negative effects of L2ACH on positive emotions and vice versa for negative emotions.

Self-Concept and Achievement

An abundance of research has shown the immense benefits of a positive academic self-concept (ASC) on affective processes and achievement in math and science (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Jansen et al., 2014, 2015; Lui & Meng, 2010; Marsh, 1986; Marsh & Martin, 2011; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004). Though STEM education initiatives only just began in the early 20th century, both science and mathematics have historically covered most school education in STEM subjects. Therefore, predicting academic achievement within these subjects and examining factors that affect achievement has mounted noticeable evidence through a theoretical lens (Bandura, 1994; Byrne & Shavelson, 1986; Marsh, 1986; Marsh & Martin, 2011; Möller & Pohlmann, 2010; Shavelson & Bolus, 1982; Taijfel & Turner, 2004; Turner & Reynolds, 2012)

Additionally, educational researchers have conducted investigations in math and science regarding the relationships between achievement and subject specific self-concepts, attitudes and values. Many studies investigated student–level factors such as gender, ethnicity, subject specific self-concept, attitudes, and values (Areepattamannil et al., 2011; Jansen et al., 2014, 2015; Lee & Kung, 2018; Ma & Kishor, 1997; Wilkins, 2004), though fewer studies explored school-level factors such as parent support, school's academic expectations, and perceived engagement of teachers (Areepattamannil et al., 2011; Chen, 2005; Ma & Kishor, 1997). Comparatively, even fewer studies have considered country-level factors such as aggregated performance, national self-concept, and national social economic status (Lui & Meng, 2010; Wilkins, 2004). The fewest studies simultaneously analyzed relationships between students' achievement and other variables at all three levels including student –level, school-level and country- level (Chui & Klassen, 2008; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014).

Therefore, Arens et al. (2017) recommends that those intending to investigate the relationship between ASC and ACH would benefit most from considering the contextual influences that affect each of those variables discreetly as well (p.625). Similarly, early reviews of research suggested the presence of other variables affecting the relationship between ASC and ACH (Byrne, 1984, p. 451; Hansford & Hattie, 1982). Though contemporary research has generally adopted the positive and reciprocal association between ASC and ACH, conclusions for further study have shown interest in investigating other variables that contribute to the moderation of the relationship (Arens et al., 2017; Guay et al., 2010; Helmke & van Aken, 1995; Marsh et al., 1999). Particularly, Byrne (1984) advocated for studies that focus on diverse student populations reference groups and that include other important variables, such as status, IQ, ethnicity, peer and parental influences (p.451). Likewise, Wilkins (2004) emphasized that the influence of other variables depended on the level of analysis, such that the analysis of contextual variables at the student-level will differ from those at the school- or country-level.

Student-Level Influences

ASC and achievement are both student-level aspects that have been shown to be influenced by a variety of contextual factors. For instance, Arens et al. (2017) pointed out that socioeconomic status (SES), IQ, and gender were found to affect both ASC and achievement, while (Hansford & Hattie, 1982) reported the effects of grade level, socioeconomic status (SES), ethnicity, ability level, and varying characteristics of measurement instruments moderated the relationship between ASC and achievement. In fact, student-level ASC was found to be a strongest positive predictor of achievement in math and science (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Lui & Meng, 2010; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004). Explicitly, Areepattamannil et al., (2011) in a multilevel study of science motivation, self-concept and instructional practices, 92% of the variance in science achievement was attributed to students. Similarly, using hierarchical linear modeling (HLM), Mohammadpour (2012) reported on eighth grade Singapore students TIMSS 2007 math results indicating that 23.40% of variance in math achievement was accounted for at the student-level. Specifically, whereas 20.37% of the student-level variance was attributed to significant linkages with math self-concept, attitude toward math, gender and SES, neither valuing math, educational aspirations, language spoken at home nor time spent on homework or out of school had an effect on achievement (p. 513).

Alike, using similar results for eighth grade students in 49 countries, Mohammadpour and Ghafar (2014) found that 40.39% of variance in math achievement was attributed to student-level differences whereby math self-concept (MSC), SES, valuing math, attitude toward math, and gender showed significant linkages to math achievement within schools (p.199). Additionally, using a similar sample in 29 countries, Mohammadpour et al. (2015) reported that on average 43.33% of variance in science achievement based on TIMSS 2007 results was found at the student-level with greater proportions accounted for by developed countries versus developing. Furthermore, 12.99% of the total student-level variance in science achievement was explained by significant relationships with science self-concept, valuing of science, gender, SES, time spent on academic, non-academic, and household chores, whereas attitude toward science and school, educational aspirations as well as language spoken at home showed no significant associations (p. 455).

Gender. Generally, research concerning Gender differences in ASC, achievement, and its moderating effect on the relationship between them has confirmed a prevailing polarity of higher male self-concept in math (Arens et al., 2017; Helmke & van Aken, 1995; Lee & Kung, 2018; Mohammadpour & Ghafar, 2014; Nagy et al., 2006) and science (Jansen et al., 2014; Marsh et al., 2015; Mohammadpour et al., 2015; Ruschenpöhler & Markic, 2019; Schroeders & Jansen, 2020; Wilkins, 2004), whereas females reflected higher self-concepts in reading (Marsh & Hau, 2003; Strein & Grossman, 2010), English, foreign languages (Marsh et al., 2015), and biology (Nagy et al., 2006). As well, cross-cultural results by Chiu and Xihua (2008) found that boys outperformed girls by 15 points on the 2002 Program for International Student Assessment (PISA), while Mohammadpour et al. (2015) found gender to be significant in only 14 of 29, such that nine of those males reined superior compared to five countries where females performed better.

Contrastingly, other studies found few differences in the strength of the relationship for males and females (Hansford & Hattie, 1982; Jansen et al., 2014; Ma & Kishor, 1997). According to Hooper et al. (2013) “TIMSS has shown that there is no large overall difference in average mathematics and science achievement between boys and girls across participating countries (p. 81). However, those results varied from country to country and the gap was present in the first stream of science achievement results in 1995 and 1999 favoring males. Yet overall, reported gender differences have been rationalized by both the Stereotype Threat Paradigm (Arens et al., 2017) in which female students would underperform or underestimate themselves in settings dominated by gender stereotypes as well as the Balanced Identity Theory (Ruschenpöhler & Markic, 2019) suggesting that if a person identifies strongly with his or gender group and a gender stereotype exists then the person will be likely to develop a corresponding self-concept” (p. 42). Likewise, Chiu and Klassen (2010) attributed gender dominance in masculine societies to weakening female self-concept and achievement, complimenting Good et al. (2012) suggestion that negative stereotypes minimized sense of belonging and contributed to less motivation and effort in male dominate subjects with impacts on subsequent self-concept and achievement (Hofstede, 2001).

Age and Grade Level. Educational research has acknowledged that self-concept not only becomes more differentiated with a clearer distinction of academic facets (Shavelson et al., 1976), but also transitions with increased age from a more positive and malleable state for younger children (Marsh et al., 1998) to a more systematic and stable state (Lee & Kung, 2018; Ruschenpöhler & Markic, 2019). Specifically, Marsh et al. (2020) reviewed studies that found a difference in academic self-concept due to relative age of students also referred to "red-shirting effect" (p. 186). This effect suggests that students who may have started school at a younger age due to birthday month, showed higher self-concepts than classmates that started school at an older age.

However, in terms of the moderation effect of age on the relationship between ASC and ACH results are inconsistent. Initially, Hansford and Hattie (1982) found a significant moderation effect supporting the developmental progression and stronger relationship of ASC and ACH from preschool through secondary school, but Huang (2011) and Moller et al., (2009) later found a negative relationship whereby the correlation between ASC and achievement is higher for younger students.

SES. Generally a positive relationship between socioeconomic status (SES) and achievement has been reported in the literature (Areepattamannil et al., 2011; Chui & Klassen, 2008; Mohammadpour et al., 2015; Strein & Grossman, 2010; Yang, 2003). Especially, measures of parent socioeconomic status such as those compositing level of education and home resources were shown to be the most influential factor affecting student achievement (Yang, 2003). To point out, Chiu and Xihua (2008) and Areepattamannil et al., (2011) found that students with more familial resources and books at home had greater achievement, whereas Mohammadpour et al. (2015) found those same relationships to be strongest in developed countries. Additionally, Strein and Grossman (2010) found family SES to be a strong moderator of student-level achievement and ASC in reading.

Affect. Affect refers to students' feelings about academic subjects, classroom aspects, and themselves as learners (Reyes, 1984). Several studies have highlighted that The National Council of Teachers in Mathematics (NCTM) and National Research

Council (NRC) mentioned the importance of the influence of effect on learning and predicting educational outcomes (Ma & Kishor, 1997; Osborne et al., 2003; Wang, 2007). In regard to self-concept, affective variables have been reported to affect students' academic self-concept with implications for students' subsequent motivation, decision-making, judgement, aspiration and achievement as well as comparable effects across age and culture (Reyes, 1984; Zhang et al., 2016). More recently, the Collaborative for Academic, Social, Emotional Learning (CASEL) has advocated the enhancement of social-emotional competencies such as emotional processes and cognitive regulation as a means of boosting achievement. By the same token, Hart et al. (2020) contends a similar relationship in regard to achievement, whereas autonomous academic motivation and awareness of the importance was found to mediate the relationship between academic self-concept and academic achievement (Guay et al., 2003).

Value. "Students' achievement and intrinsic motivation in learning mathematics can be affected by whether they find mathematics to be interesting, valuable, or an important subject for success in school and for the future career aspirations" (Mohammadpour & Ghafar, 2014; Mullis et al., 2004). In other words, the more students value the academic task at hand, the more motivated they will be to achieve higher scores (Arens et al., 2019). For instance, O'Mara et al. (2006) showed a significant relationship exists between intrinsic value and later attainment value with ASC in math and English, while Mohammadpour et al. (2015) reported that valuing science yielded the strongest link to achievement in Botswana and Qatar. Correspondingly, the presence of a positive relationship between high ASC and high achievement was shown to contribute to a greater intrinsic value when compared to students with lower ASC (Valentine et al., 2004). Additionally, Guay et al. (2010) contended that the nature of the variables measuring academic interest determined the direction of relationship between ASC and achievement. Markedly, the more complex motivational spectrum of measure used, the more distinction between the effects of academic awareness and academic value on achievement was found, highlighting the possibility that students may find academic tasks to be boring, yet still find value in its necessity.

Self-Efficacy. Self-efficacy (SE) refers to the belief of one's ability to successfully perform certain tasks (Bandura, 1994). It is said that people with high academic self-efficacy reflect greater interest, intrinsic motivation, engagement in academic tasks, and achievement, as well as approach difficult tasks with greater confidence (Bandura, 1994; Mimi Bong & Skaalvik, 2003; Huang, 2011; Marsh et al., 1991; Pajares, 1996; Valentine et al., 2004). Uniquely, SE differs from ASC in several ways. For instance, whereas self-efficacy is more future oriented, malleable, assessed through internal goal referenced frames and only concerned with what can be done with the skills one already has, self-concept is more past oriented, stable, assessed by external frames of reference such as social comparisons, and primarily concerned with evaluating intrinsic skills and abilities (Arens et al., 2019; Bong & Skaalvik, 2003; Jansen et al., 2015; Marsh et al., 1991; Parker et al., 2014; Valentine et al., 2004). Comparatively, theoretical distinction is prevalent, but tests of construct validity have inconsistently confirmed statistical discrimination between the two constructs often due to variation of instrumentation and nature of wording. For instance, Valentine et al., (2004) review pointed out that assessments of self-concept often encompassed composite measures of self-esteem and self-efficacy to include items for evaluative self-feelings and perception of capability, whereas others measured the constructs discretely.

Nevertheless, studies have shown substantial relationships existed between academic achievement and self-efficacy (Chang, 2012; Gao et al., 2020; Huang, 2011; Marsh, 2005; Marsh & Martin, 2011; Pajaras & Miller, 1994; Pajares, 1996; Valentine et al., 2004). Precisely, in a study using path analysis to compare Math self-efficacy with Math self-concept as predictors of math achievement, Pajaras and Miller (1994) found the correlation between math self-efficacy and math achievement was slightly higher than that between math self-concept and math achievement. Chang (2012) found fifth graders had 70% confidence in their ability with evidence of a positive relationship between math self-efficacy (MSE) and math achievement, while Chiu and Xihua (2008) investigated motivational effects on math achievement scores across 41 countries and found that with a 10% increase in self-efficacy, students scored one to three points higher in math achievement. These results corroborated Zimmerman et al. (1992) earlier indications that "self-efficacy influenced not only students' setting of academic goals for themselves, but also their achievement of these goals" (pg. 637). However, Gao et al. (2020) reported that the effects of self-efficacy mediated the relationship between teaching practices and achievement, as well as highlighted in their literature review that there were few studies that examined large-scale, internationally representative data or focused on middle school students (p. 386).

Attitude. Attitude toward science can be defined as a person's predisposition positive or negative likeness, feelings and beliefs about school science subjects and science tasks (Hacieminoglu, 2016; Osborne et al., 2003; Zimmerman et al., 1992).

Papanastasiou and Papanastasiou (2004) emphasized that “the examination of the attitudes toward science is especially important since attitudes can influence student’s educational achievement in ways that reinforce higher or lower performance” (p. 240). With this definition being extended to include math as well, it has been shown that higher achievement scores are related to positive attitudes toward math and science (Chen et al., 2018; Ma & Kishor, 1997; Mohammadpour, 2012; Reyes, 1984).

Notably, Mullis et al. (2008) indicated that “developing positive attitudes toward mathematics is an important goal of the mathematics curriculum in many countries” (p.173), therefore results reflected that “average mathematics achievement was highest among students with a higher indexed levels of positive attitudes with next highest among those at the medium level and lowest at the low level” (p. 174). Comparably, Chen et al. (2018) found that positive attitude toward math uniquely predicted math achievement, even after multiple other affective factors were considered. Further, Hacieminoglu (2016) supported conclusions that positive attitudes contributed to higher achievement in science, adding results of causality to confirm that attitudes in science influenced achievement, rather than achievement influencing attitudes.

School-Level Influences

Investigating the compositional effects of school-wide characteristics on student-level learning outcomes has been of interest to educational researchers (Coleman, 1975; Strein & Grossman, 2010). Though such research has been dominated by the examination of school-averaged ability on student-level ASC also known as the Big-Fish-Little-Pond Effect (BFLPE), other school-level variables have been examined as well for associations with self-concept and academic achievement in math and science (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Coleman, 1975; Labenne & Greene, 1969; Martin et al., 2016; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Abdul Ghafar, 2014; Strein & Grossman, 2010).

Despite the fact that school effects often account for less variation in ASC than those of student-level variables, effects of school-level variables are valuable indicators to be examined. For instance, Strein and Grossman (2010) study of the effects of school type, size, and average SES on students' reading and math ASC found that about 10% of the variance in ASC was due to school-level effects. As for achievement, the earliest report noted relatively higher effects of school location and school SES on student achievement in literature, science, and reading for 14 year old students accounted for 26% of variance versus 22% variance explained for 10 year old's (Coleman, 1975). As well in a series of three-level, HLM reports, school-level variables such as climate, size, and SES accounted for 53.25%, 15.85%, and 25% variation in math achievement (Chiu & Klassen, 2010; Mohammadpour, 2012; Mohammadpour & Ghafar, 2014) as well as 14% and 19.78% variation in science achievement attributed to school instructional practices and demographics (Areepattamannil et al., 2011; Mohammadpour et al., 2015).

School SES. Most of the literature regarding the effects on school-level SES have revolved around its positive association with student-level achievement, especially since the Coleman et al. (1966) report that called attention to financial inequality of schools at that time (Armor et al., 2018; Caponera & Losito, 2016; Coleman et al., 1966; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Willms, 2010; Yang, 2003). Specifically, Armor et al. (2018) suggested that much of the measurements of school-level SES were the result of the aggregation of student-level

SES. For their study using three, state-wide achievement measures in math and reading for third to eighth graders, they found significant cross-sectional effects for school SES and student achievement as measured by scores averaged across schools, grades, and years, yet those effects evaporated once longitudinal data was introduced.

On the other hand, Caponera and Losito (2016) focused on the impact of contextual factors such as availability of school resources on math achievement in high and low SES schools, which like Yang (2003), reported varying results between countries with effects from high SES schools explaining a greater proportion of variability in achievement between schools. Similarly, at the school-level, Mohammadpour (2012) found that student-level educational resources were a stronger predictor of math achievement for TIMSS 2007 eighth graders, whereby students in classrooms with a greater percentage of "economically advantaged students" (TIMSS measure of school SES) reflected higher achievement scores. Further investigation of the same data set using an international population reported a positive relationship between school-resources and achievement in all but the Czech Republic (Mohammadpour et al., 2015), while another complementary investigation confirmed similar results for science achievement (Mohammadpour & Ghafar, 2014).

Correspondingly, Willms, (2010) examined the student- and school-level SES compositional effects on 2006 PISA results in science, reporting a 58-point disparity in achievement scores for school SES compositional effects that accounted for 4.4%, 58.8%, and 12.7% of variation in achievement among students, schools, and countries. Their conclusions suggested a lack of consideration for additional school-level contextual variables such as school type and instructional practices in the aggregation of school compositional effects. Additionally, other results found that low school SES reduced the positive effects of student-level influences on science achievement (Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Zheng et al., 2019), while school resources was a significant predictor of science achievement between schools in science (Mohammadpour et al., 2015), but not between countries for math (Mohammadpour & Ghafar, 2014) nor was school SES a significant predictor of math achievement when aggregated from the student-level (Mohammadpour et al., 2015).

On the other hand, of the vary few studies that reexamined the effects of school SES on ASC, results reported a negative relationship between school-level SES and student-level ASC though effect sizes were negligible (Marsh, 1987; Marsh & Parker, 1984). For example, Strein and Grossman (2010) reported that "students in schools with higher percentages of minorities had *slightly* greater ASCs than for students in schools with fewer minorities, after considering the effects of individual and school-wide achievement" (p.6). Conclusively, inconsistencies in results can be attributed to variations in measurement.

Urbanicity. "Educational research has examined rural/urban differences in achievement with many believing that students from smaller and rural schools receive an education inferior to that of students from larger urban or suburban schools due to shortages of resources" (Areepattamannil et al., 2011; Coleman, 1975; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Triandis, 1989; Young, 1998, p. 387) or differences in cultural practices (Triandis, 1989; Young, 1998; Zhang et al., 2016). Likewise, a similar position held true for school location relations to self-concept (Trowbridge, 1972). Specifically, Trowbridge (1972) using one of the earliest measures of general self-concept as measured by the Coopersmith Self-Esteem Inventory, indicated a negative relationship for school location wherein rural children had higher self-concepts than urban and suburban children.

Nonetheless, Young (1998) results using TIMSS data for secondary Australian students showed that location had a stronger effect on mathematics achievement than for science where it accounted for 21.5% of its variability. Yet still, Mohammadpour and Ghafar (2014) reported location as the strongest school-level predictor with a 20-point disparity favoring urban students. Similarly, school location was statistically linked to science achievement in nine mostly developed countries where urban students outperformed rural students (Mohammadpour & Ghafar, 2014, p. 461). Contrastingly, Areepattamannil et al. (2011) found trivial results for school location that only accounted for an additional 2.2% of variance between schools compared to a greater influence of student-level demographics on student achievement (p.247).

Climate. Educational researchers and practitioners assert that supportive school and classroom climates can positively influence the academic outcomes of students, thus potentially reducing academic achievement gaps between students (Berkowitz et al., 2017, p. 425; Caponera & Losito, 2016). Likewise, earlier recommendations for future research regarding the causality of ASC and achievement suggested investigating "subject-specific differences as well as relationship for different classroom compositions and social climates (Helmke & van Aken, 1995, p. 636). Precisely, Berkowitz et al. (2017) reviewed 78 articles and concluded that although there was great variability in measurement and definition of school climate, "positive school climate contributed to higher academic achievement and decreased the negative influence of poor SES background characteristics and other risk factors on academic achievement" (p.457).

Correspondingly, school climate has been recognized as a multidimensional construct with various aspects to define it. For example, using TIMSS 2007 measures of teacher job satisfaction and their expectation of students' success, school and home connectedness, and students' efforts, it was determined that students score higher in schools where the principals and teachers describe a positive school climate positively (Mohammadpour & Ghafar, 2014, p. 196). Similar results found that school climate as perceived by school principals was one of the strongest predictors of science achievement in 16, mostly developed, countries (Mohammadpour et al., 2015) and attributed to over 12-point improvement in math scores across countries as well (Mohammadpour & Ghafar, 2014).

Country-Level Influences

Country dynamics and cultural norms serve as an influence in one's life (Chiu & Klassen, 2010; Hofstede, 2001; Wilkins, 2004). Whereas not only the degree of collectivism (Hofstede, 2003), but also economic classifications as measured by per capita income and gross domestic product (Chiu & Klassen, 2010; Mohammadpour et al., 2015; Mohammadpour & Abdul Ghafar, 2014; Tucker-Drob et al., 2014; Zheng et al., 2019), as well as country-wide school tracking practices (Arens et al., 2017; Hooper et al., 2013a; Mohammadpour & Ghafar, 2014) has impacted students' perception of self and achievement.

Collectivist/Individualist Culture. Notably, Wilkins (2004) concluded that the relationship between ASC and achievement depended on the level of analysis with results suggesting that there may be cultural differences that influence individual and country outcomes. For instance, individualistic western cultures “not only display higher self-concept, but also lower performance” due to the notion that “students in more individualistic societies value their individuality more, choose downward comparisons more often, and have higher self-concept while students in more collectivist societies often seek upward comparisons and have lower self-concept” (Chiu & Klassen, 2010, p. 4; Chiu & Xihua, 2008; Kashima et al., 1995; Rhee et al., 1996; Triandis, 1989).

Income Per Capita. Specifically, Chiu and Klassen (2010) found that the positive link between math self-concept and math achievement was stronger in wealthier countries, whereby when a country’s GDP per capita exceeded the mean by 10%, its students averaged six points higher in mathematics compared to if inequality exceeded 10% results would be four points lower. However, no cultural value was linked to students’ mathematics achievement, showing no support for the collectivist achievement hypothesis with limitations attributed to the absence of other possible relevant variables such as degree of tracking to explain differences. Similarly, Tucker-Drob et al. (2014) reported a “stronger association of science interest with science achievement in higher GDP countries, such that in very-low-GDP countries, the association between science interest and science achievement was essentially flat” (p. 2054). In the same light, Mohammadpour and Ghafar (2014) found that country-level SES was the main predictor of national achievement accounting for about 89% of country-level variation in math achievement with a one-scale point increase of national SES attributing to a 108-point increase in math achievement (p. 207).

Country Tracking Practices. Contrastingly, in an analysis of research on the effects of ability grouping on achievement, little effect was found with minimal benefits for advantaged students either (Hattie, 2002). On the other hand, Arens et al. (2017) asserted that positive effects of tracking practices for high-ability students is due in part to improved instructional quality and fewer disciplinary problems in the classroom compared to less instructional support provided to students in lower achievement track schools (p.623). Yet, Dicke et al. (2018) results indicated that multilevel modeling of school compositional effects reduced the positive effects of ability grouping on achievement, known as the spillover effect, implying there was a negative effects of school-averaged achievemnt on academic self-concept in support of BFLPE research.

The Present Study

For forty years, BFLPE researchers have produced staggering confirmations of global generalizability for the negative effects of school-averaged achievement (L2BFLPE) and country-averaged achievement (L3BFLPE) on student-level academic self-concept with robustness across a variety of multilevel control variables. Yet, far fewer BFLPE studies have examined intervening variables that could potentially moderate the negative effects and maximize the positive effects of L2BFLPE or L3BFLPE and its subsequential, reciprocal impact on corresponding achievements. However, a multitude of previous research has reported the positive effects of students-, school-, and country-level variables on achievement in math and science, while to a lesser extent previous research has reported the positive effects of similar multileveled variables on academic self-concept as well.

Therefore, it was the intent of this study to advance current BFLPE research by investigating a diverse range of multilevel moderation effects on L2BFLPE and L3BFLPE. Specifically, this study assumed the implications of the Reciprocal Effects

Model to discretely apply the positively associated student-, school-, and country-level variables that were reported in previous research concerning academic self-concept or achievement as moderators of L2BFLPE and L3BFLPE in math and science. Subsequently, these results contribute to BFLPE theory and inform STEM policy and practice math and science with a greater understanding of intervening variables that ameliorate the negative effects of social comparisons with school- and country-averaged achievement.

Research Questions

Research Question 1 (RQ1)

Does school- and country-level BFLPE in math and science exist across 26 TIMSS 2019 countries (Marsh et al., 2019)?

Research Question 2 (RQ2)

Is student-level academic self-concept in math and science significantly associated with *student-level* achievement, gender, self-concepts, socioeconomic status, value and attitude toward math or science, *school-level* achievement, socioeconomic status, location, climate, academic self-concept, value and attitude toward math or science and *country-level* achievement, income per capita, classification of individualism, tracking practices, self-concepts, value and attitude toward math or science across 26 TIMSS 2019 countries (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014)?

Research Question 3 (RQ3)

Is school- or country-level BFLPE moderated by student-, school, or country-level variables found to be significantly associated with student-level self-concept in math and science across 26 TIMSS 2019 countries (Seaton, 2010)?

Chapter 3: Methodology

Sample

To address the research questions, this study conducted a secondary analysis of results from the most recent TIMSS 2019 large-scale, international assessment of math and science achievement for eighth graders including results of corresponding contextual questionnaires as well (Eklöf, 2007; Lui & Meng, 2010; Marsh et al., 2015; Marsh et al., 2014; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004; Yang, 2003). Specifically, in Grade 8, TIMSS 2019 assessed a total of 262,998 students from 8,760 schools in 46 countries (International Association for the Evaluation of Educational Achievement (IEA, 2016). Of those, 169,957 students in 5,410 schools from 26 countries that participated in both Grade 8 math and science as a single integrated subject (SIS) will be incorporated in this cross-national comparative analysis.

Collectively, nine East Asian and Pacific countries (EAP) including Australia, Republic of Korea, Chinese Taipei, Hong Kong SAR, Japan, Republic of Korea Malaysia, New Zealand, Singapore as well as ten Middle East and North African (MENA) countries including Bahrain, Egypt, Islamic Republic of Iran, Israel, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, and United Arab Emirates were represented in the sample. Likewise, five European and Central Asian (ECA) countries including England, Ireland, Italy, Norway, and Turkey and the only Sub-Saharan African (SSA) country of South Africa were also represented in the sample. Chile was the only Latin American country represented and the United States was the only North American country represented in the sample as well (World Bank, 2020).

Correspondingly, the sample comprehensively represented the national targeted populations of each participating country defined as “all students enrolled in the grade that represented eight years of schooling counting from the first year providing the mean age at the time of testing is at least 13.5 years as standardized by 2011 International Standard Classification of Education (ISCED)” (LaRoche et al., 2020, p. 3.4). Generally, school participation ranged from 623 schools in United Arab Emirates to 134 schools in New Zealand and student participation ranged from 22,334 students in United Arab Emirates to 3,265 students in Hong Kong SAR (see Table 1). Primarily, each participating country was assessed in Grade 8 that which corresponds to Year 8 in Australia, Year 9 in New Zealand and England, Basic 8 in Chile, Secondary 2 in Singapore and Hong Kong SAR, Lower Secondary Grade 3 in Italy, Middle School Grade 2 in Republic of Korea, and Second Year in Ireland. Uniquely, South Africa and Norway assessed Grade 9 to maintain trend comparisons and best match curricula (Mullis et al., 2020, p. 511). Additionally, average ages of participants ranged from 14.5 years old in Japan to 13.7 years old in Kuwait, while of those assessed in ninth grade average ages were 15.5 years old in South Africa and 14.7 years old in Norway. Overall, gender was distributed equally among samples in each country.

Sampling Method

Overall, this study analyzed only a sample of the total student population assessed in TIMSS 2019. Uniquely, TIMSS 2019 was the first cycle to begin transitioning to a computer based assessment platform which gave countries the option to assess using the paperTIMSS format or the new eTIMSS format (LaRoche et al., 2020, p. 3.1). Though the content among the two versions was similar, unavoidably there were differences due to the mode of administration, but results were made comparable through "bridge" data that was also collected as a link between the two versions. Additionally, in effort to produce unbiased, accurate, and comparable international results, TIMSS National Research Coordinators (NRC) in collaboration with Statistics Canada and IEA Hamburg, selected participants using a complex two-stage, stratified, random sampling design. The NRC coordinated all sampling operations within their home country as well as provided Statistics Canada with documentation needed to conduct national sampling calculations including detailing enrollment coverage and school-level exclusions of unrepresentative populations often based on limitations of disabilities, language, and school size.

Explicitly, Initial stages of sampling required the development and implementation of a unique National Sampling Plan to detail the National Target Population with identification and specifications of the national target grade. Though TIMSS identified the Grade 8 International Target Population as all students in their eighth year of formal schooling, it was recommended for countries to select the next higher grade if the average age was less than that described by ISCED or for curricular and comparative needs. Sequentially, to ensure that specific groups of schools in the sample were proportional represented, the NRC worked cooperatively with Statistics Canada and the IEA Hamburg to identify demographic variables by which the target schools would be explicitly or implicitly stratified (p 3.12). Thereafter, the School Sampling Frame was created as a spread that listed "all schools in the country that have students enrolled in the target grade" (p. 3.14) accompanied by detailed classroom and sampling frame specifications including school measures of size (MOS) and average class size that was authorized upon evaluation of compliance with standards set by TIMSS & PIRLS International Study Center.

Successively, in the first sampling stage, schools that were listed on the National Sampling frame were grouped according to the explicit national stratification variables then each group was sorted according to implicit stratum variables as well as MOS. Then, random-start fixed interval systematic sampling was simultaneously applied with probabilities of school selections proportional to its size (PPS) to draw random samples of schools that were discretely designated for field tests, data collection, or replacement (p.3.29). Generally, in the first-stage of random sampling, larger schools had a higher probability of being sampled. However, the second-stage of random sampling applied Within-School Sampling Software (WinW3S), so one intact class from the sampled schools were sampled inversely proportional to school size which fostered equal probability of student selections within each school (p. 3.11). Characteristically, intact classrooms endorsed greater evaluation of curricular and instructional experiences as responses to corresponding contextual questionnaires reflected those of students and teachers within the same class.

Uniquely, sampling precision requirements mandated that "national student samples reflected standard errors no greater than 0.035 standard deviations from the country's mean achievement" which was usually attained with a sample of 150 schools and 4000 students with the addition of 1,500 students for the bridge data sample. Yet, some countries preferred to sample more than one class per school to "increase student sample size and estimation of school effects" (p. 3.9). Additionally, though 100% participation rate was the goal, there must have been at least 85% school participation accompanied by 95% classroom participation and 85% student participation or a combined participation rate of 75% for a sample to acceptable for data collection (p. 3.10).

Distinctively, due to unequal sample sizes, TIMSS 2019 provided sampling weights to reflect accurate proportional representation of the population from which the sample was drawn (Rutkowski et al., 2010). Students' sampling weights mirrored the inverse of the probability for a student in a sampled class to be selected with adjustments for nonparticipation that reflected selection probabilities and sampling outcomes at three levels including school, class (within school), and student (within class). Exclusively, all student data that was reported in TIMSS 2019 were weighted using overall student sampling weights which was calculated as the product of the final weighting components for the three levels and only portioned for participating students whereas nonparticipating students are weighted at 0 (Fishbein et al., 2021). For instance, total student weight (TOTWGT) "summed to the student population size in each country" whereby larger countries were be represented with a greater proportion than smaller (Fishbein et al., 2021, p. 83).

However, TOTWGT inflated sample sizes to reflect the total population of participants, so to limit bias for larger countries in cross-comparative studies, TIMSS 2019 provided transformations of TOTWGT including HOUWGT and SENWGT as viable options for accurate proportional representations of student-level data for cross-country analyses. Additionally, concerning analyses wherein countries must be treated equally, Senate Weight (SENWGHT) offered uniform weighted sample sizes of 500, while House Weight (HOUWGHT) "ensured that the weight sample corresponded to the actual sample size in each country" (p. 84). Likewise, schools were sampled with a probability proportional to school size and the School Weight (SCHWGT) was calculated by combining basic school weight with the school nonparticipation adjustment. The final school weight was recommended for use wherein school is the only unit of analysis (p. 84).

Instrument

Designed by the International Association for the Evaluation of Educational Achievement (IEA) and its TIMSS & PIRLS International Study Center at Boston College in conjunction with an international cooperative of government agencies and research institutions, TIMSS 2019 was the seventh iteration of the large scale, international, standardized assessment of achievement in math and science. Since its initiation in 1995, TIMSS has assessed students' achievement in Grade 4 and Grade 8 every four years in math and science with the additional support of contextual questionnaire responses to capture influential national, home, school, classroom, and individual student contexts. Accordingly, trends in the effects of education policy, structure, and curriculum among global education systems were monitored, compared, and utilized to "help make informed decision about how to improve science and math achievement" and STEM education worldwide (Mullis, 2017, p. 3).

Uniquely, in addition to maintaining standardization consistent with previous TIMSS assessments, TIMSS 2019 provided four new key features. Most notably, this iteration marked the beginning of a transition to a computer-based assessment. Specifically, for the first time, the assessment was offered to fourth and eighth graders as an option of either the new eTIMSS format or the original paperTIMSS format. eTIMSS presented an innovative "engaging, interactive, and visually appealing format to more efficiently "assess complex areas and perform translation, assessment delivery, data entry, and scoring," while the paperTIMSS replicated the standard format from previous cycles (Cotter et al., 2020, p. 1.1). Supplementary, eTIMSS versions for fourth and eighth grade included the introduction of extended Problem Solving and Inquiry (PSI) tasks designed to simulate real world process manipulations in math and virtual scientific investigation (p. 1.1).

Furthermore, in effort to control for a mode effect between the eTIMSS and paperTIMSS, TIMSS 2019 collected comparable "Bridge" data of an additional sample from eTIMSS participants by administering the same trend items carried over from TIMSS 2015 in a paper version to "calibrate linkage and safeguard trend reports" (LaRoche et al., 2020, p. 3.1). Finally, with the success of TIMSS 2015 Numeracy assessment, TIMSS 2019 offered the option of administering a less difficult version of the fourth-grade math assessment. However, this study only collected data from TIMSS 2019 Grade 8 math and science achievement assessments, contextual questionnaires and *TIMSS 2019 Encyclopedia* as they presented the most appropriate measures of cross-national data synonymous with the variables addressed in the research questions of this study.

Assessment Development

At its core, the TIMSS 2019 Curriculum Model guided the development of achievement assessment frameworks and items from the complimentary supplemental contextual questionnaires as well (Mullis, 2017, p. 4). Overall, the three features of the model included the *intended curriculum*, the *implemented curriculum,* and the *attained* *curriculum* for each participating country (p. 4). Specifically, the *intended curriculum* was exhibited in the *TIMSS Encyclopedia* as a chapter prepared by each participating country summarizing the characteristics of their national education structure, curricula, and policies. Like its predecessor, TIMSS 2019 embedded elements of the country chapters in the contextual curriculum questionnaires to highlight curriculum expectations and instructional practices unique to each country. Correspondingly, the *implemented curriculum* unique to each country was highlighted in the school, teacher, and classroom contextual questionnaires, and the *attained curriculum* that was unique to each country was displayed within student achievement assessments as well as within student contextual questionnaires.

Commonly, TIMSS math and science achievement items and contextual questions were updated and improved with each subsequent iteration through a series of standardized procedures and field tests that best illustrated the current state of international policy and practice (Hooper, 2016; Mullis et al., 2016). Specifically, TIMSS 2019 math and science achievement items were similar to those from TIMSS 2015 with improvements that reflected both new formats as well as the most current curricula, standards, and frameworks described by participating countries in *TIMSS 2015 Encyclopedia* (Centurino & Jones, 2017, p. 29; M. Lindquist et al., 2017, p. 13). Notably, TIMSS 2019 eighth grade assessment "required developing and field testing 325 new math and science items for both formats as well as seven PSI tasks" (Cotter et al., 2020, p. 1.1).

Similarly, context questionnaires were developed using collaboration and reviews among TIMSS & PIRLS Study Center staff, policy experts from TIMSS 2015

Questionnaire Item Review Committee (QIRC), and the NRCs that reviewed the existing TIMSS 2015 questionnaires to determine what changes were needed (Hooper et al., 2017, p. 59). Particularly, improvements to TIMSS 2019 focused on “enhancing measures of teacher instructional quality, addressing areas relevant to technology, and reducing response burden on teachers” (Mullis & Fishbein, 2020, p. 2.5). For instance, many additions addressed technology use for instruction and assessment, cyber bullying and social media, while other additional elements measured student tracking practices, school-working conditions, as well as teacher collaboration and confidence were deleted (p. 2.6).

Frameworks and Design

Achievement Items. All math and science achievement items were created around a two-dimensional framework consisting of both content and cognitive dimensions. The content dimension addressed the general subject matter that was assessed, while the cognitive dimension assessed general thinking processes (Lindquist et al., 2017, p. 14; Mullis, 2017, p. 7). Grade 8 cognitive dimensions for math and science similarly represented knowing, reasoning, and applying domains, yet content domains differed, as did the percentage of assessment items dedicated to each domain (see Figure 7 for assessment domain details). Additionally, eTIMSS 2019 included PSI “scenarios that presented students with adaptive and responsive ways of integrating and applying process skills in math and science by following a series of steps toward a solution” (Mullis, 2017, p. 8).

Due to the massive size of resources and time needed to assess all participants on 211 math items, 220 science items, eTIMSS and paperTIMSS 2019 employed a matrix-sampling design that condensed the assessed items into 28 item blocks. Notably, 16 of the 28 item blocks were transferred directly from TIMSS 2015 for consistency in monitoring trends and 12 new item blocks were developed specifically for TIMSS 2019 (Fishbein et al., 2020, p. 10.2; Martin et al., 2017, p. 84). Overall, 14 items blocks were dedicated to math achievement items and 14 blocks for science achievement items with each having an equal distribution of content and cognitive dimensions assessed.

Distinctively, each item block included 12-18 assessment items that represented 22.5 minutes of assessment time with an estimated 18 value points. In total, 28 item blocks were distributed among 14 student booklets such that each booklet included four blocks with two blocks in math and two blocks science of which two blocks represent trend items and two represent newly developed items. Generally, student booklets were organized with various combinations of item blocks whereby each item appeared in two booklets (p. 88). Student booklets were distributed to intact classrooms and each student completes the two parts of the booklet in two 45-minute intervals with 30 minutes allotted afterwards to complete student contextual questionnaires. Similarly, eTIMSS design presented block counterparts resembling those in paperTIMSS as closely as possible except for the options to drag and drop, sort and other digital adaptations (p. 88).

Additionally, eTIMSS included 25 math items for three PSI tasks and 29 science items from two PSIs with the bridge booklets that included 117 mathematics and 122 science trend items (Fishbein et al., 2020, p. 10.2). Therefore, instead of assessing 14 items blocks, eTIMSS design incorporated two additional block combinations for PSI tasks providing 16 item blocks that were equally distributed among student digital assessments using the same within-school sampling software as paperTIMSS.

Overall, items in both formats were presented in the form of selected response or constructed response questions (Cotter et al., 2020, p. 1.26). Explicitly, instructions were visibly illustrated a sample question with the answer clearly marked. Selected response items elicited single responses, multiple responses, or a series of responses. Most selected responses questions were worth 1-point value, though some received 2-score points and were marked as either correct (2-point value), partially correct (1-point value) or incorrect 0-point value). Uniquely, selected responses covered 48% of math items and 60% of science items of both assessment formats and domains in eighth grade .

Contrastingly, constructed response questions were worth one- or two-value points and required a written explanation or interpretation with fully correct, partially correct, or incorrect scoring assigned. Markedly, constructed responses covered 52% of math items and 40% of science items for both assessment formats across both domains. Furthermore, upon the completion of the assessments extensively trained scorers referenced a standardized scoring guide and applied standardized scoring procedures that were uniformly implemented for scorers among all participating countries.

Contextual Questionnaires. In addition to achievement assessments, TIMSS 2015 administered contextual questionnaires to eighth grade students, their teachers, principals, and the NRCs that measured the influences of community, home, school, and individual contexts. Explicitly, these background questionnaires measured these contexts as latent construct measures that were represented as composite measures of itemized, scaled responses (Hooper et al., 2017; Mullis & Fishbein, 2020). Generally, measures reflected five broad areas including national and community contexts, home contexts, school contexts, classroom contexts, as well as student characteristics and attitudes toward learning (Hooper et al., 2013, p. 62).

Sequentially, context questionnaires were administered after the completion of Part 1 and Part 2 of the achievement assessments. Generally, student questionnaires required 15-30 minutes to complete. They measured students' responses regarding home and school lives, climate for learning, demographic information, home environment, as well as self-perceptions and attitudes toward math and science" (Martin et al., 2013, p. 96). Additionally, teacher questionnaires required 30 minutes to complete and collected responses from teachers concerning teacher, classroom, and instructional characteristics, while school questionnaires require 30 minutes to complete and captures principals' responses regarding instructional and student characteristics, school resources, as well as parental and staff involvement (p. 97). Furthermore, the NRC from each country highlighted national contexts with responses to the curriculum questionnaire that concerned country educational systems and policies, while additional country contextual information provide by NRC was placed in the *TIMSS 2019 Encyclopedia* (pg. 98; Hooper et al., 2017, p. 60).

Item Statistics and Reliability

TIMSS and PIRLS International Study Center examined the internal consistency of TIMSS 2019 math and science achievement assessment including eTIMSS, paperTIMSS, eTIMSS PSIs, and the paper "bridge" booklets with measures of reliability and comparisons of item statistics within and across all participating countries (Fishbein et al., 2020, p. 10.1). Particular attention focused on any changes in trend items from TIMSS 2015 and differences between eTIMSS and "bridge" trend items (p. 10.20). For instance, item statistics that was reviewed for both multiple choice and constructed response items included overall test reliability as measured by total number of responses per country (N), item difficulty as measured by average percentage correct (DIFF), item discrimination (DISC) as measured by correlation between response and total score on all items, percentages (P) as measured by percentage of student choosing either selected or constructed response options including those not reached and omitted, point biserials (PB) as measured by correlations between the selection of each response option and total response options administered, item difficulty as measured by Rasch one-parameter IRT model (RDIFF), as well as reliability for human scores constructed response items as measured by percentage agreement on score and code as measured by inter-scorer agreements on randomly sampled booklets within and across countries. All item statistics were reported in TIMSS 2019 Data Almanacs for each grade.

Consecutively, all item statistics were compared across all countries and any inconsistencies were subsequently flagged when point-biserial correlations were not ordered, item difficulty was less than .25 or exceeded .95, item discrimination was less than .10, items registered easier or harder than international average, percentage of students choosing one selected response or constructed response value were less than 10%, score reliability on constructed response was less than 85% (Fishbein et al., 2020, p. 10.7). Upon the flagging of an item, flaws or inaccuracies in national translation documents and printed booklets were subsequently removed from that country's database (p. 10.17). Distinctly for TIMSS 2019 only minimal items were removed that had resulted from translation or printing discrepancies with smaller amounts of items removed due to "severe differential item functioning" (10.19).

In like manner, Cronbach's Alpha reliability coefficients and principal component analysis of Context Questionnaire scale items was employed to ensure reliability, validity, and comparability across countries (Yin & Fishbein, 2020, p. 16.12). Specifically, reliability measures included all construct scales (i.e. emphasis on success, student confidence, student value, etc.) from both subjects and grades in every country (Mullis et al., 2016, p. 15.10). Overall, results reported reliability coefficients that were mostly acceptable "with almost all above 0.7 and many above 0.8" (p. 16.12). Furthermore, whereas principal component analysis summarized the amount of variance that the construct accounted for in each item, results indicated a high percentage of item variance that was accounted for by each construct confirming that each item could be measured by a single construct. As well, high and positive loadings indicated "a strong correlation between each item and the measured construct in every country" (p. 16.12).

Measurement Scaling

Achievement scores in math and science were scaled within a range from 0 to 1,000 with a mean of 500, standard deviation of 100, and scores generally falling within the range between 300 and 700 score points. Educational Testing Service (ETS) and U.S. National Assessment of Educational Progress (NAEP) designed the TIMSS 2019 scaling method to integrate Item Response Theory (IRT) and marginal estimation (von Davier, 2020, p. 11.2). Using a three-parameter IRT measurement model for multiple choice questions scored as correct or incorrect, a two-parameter IRT measurement model for constructed response items with only two score options, and a partial-credit model for constructed response items, achievement scores were calibrated from the probability of an item being answered correctly or incorrectly based on the proficiency of a student (p.

11.2). However, being that the assessments were designed using a matrix sampling technique wherein each student only answers a portion of the total number of available questions, precise measures of students' individual proficiency could not be rendered. Therefore, TIMSS employed plausible values as an estimation of the proficiency distributions.

Since its inception TIMSS has applied Rubin Rule (1987) of addressing missing data through Estimation Maximization to impute the remaining values of those omitted from matrix sampling of student achievement. Therefore, proficiency estimates or plausible values were assigned as five random selections from the conditional distribution of imputed values that has been created from measures of assessment responses and model parameters for the items (von Davier, 2020, p. 11.1). Distinctly, TIMSS 2019 also accounted for a mode effect that often "manifested as differential item functioning by some of the items across modes, which can affect measurement invariance and may cause undesirable changes in comparability of proficiency scores" (p. 11.9). However, cross-mode correlations between item location parameters from both paperTIMSS and eTIMSS items were very high, which suggested a strong link and comparability of results that were reported on the same scale (p. 11.14).

Additionally, TIMSS 2019 applied a scale anchoring method to produce content referenced interpretations of achievement results by identifying student competencies in terms of different locations along the achievement benchmark scale" (Mullis & Fishbein, 2020, p. 15.1). Categorically, TIMSS & PIRLS International Study Center collaborated with the expert Science and Mathematics Item Review Committee to distinguish four benchmarks that described content-referenced competencies as indicators of international achievement. Conclusively, cut points identified the benchmarks and scores within a five-point range attained benchmark recognition. For instance, Low Intermediate Benchmark cut point was located at score value 400 with a score range from 395 to 405, Intermediate International Benchmark cut point is located at 475 with range from 470 to 480, High International Benchmark cut point is located at 550 with a range from 545 to 555, and the Advanced International Benchmark cut point is located at 625 with range from 620 to 630 (p. 15.3).

In a similar fashion, TIMSS & PIRLS International Study Center applied Rasch IRT partial-credit model for scaling Context Questionnaires. Precisely, using ConQuest 2.0, item calibrations estimated item parameters (delta), deviations from delta (tau), and infit item statistics for the entire assessment population that was senate weighted for equal contribution of a sample size of 500 in each country (p. 16.3). Thereafter, like the achievement tests, benchmarks were assigned at unique cut points for each polytomous construct scale to determine high, medium, and low values on the construct. Notably, raw scores for polytomous item responses were calculated by assignment corresponding number values in ascending order to response items within each scale (p. 16.5). Standardly, center points were located at 10 reflecting the mean across all countries and scale units were set to two reflecting the standard deviation across all countries (p. 16.4). However, differences in item response options (i.e. degree of agreement or magnitude of construct present) required cut point values to be in locations unique to each construct scale. Correspondingly, “the IRT calibration and scoring procedures for trend scales were the same as those used for the newly developed context scales” (p. 16.9).

Measures

This section identifies each variable that was included in this study including the variable question number, variable name, variable description and where applicable the scale range and cut point values indicated on TIMSS 2015 assessment and questionnaires (see Appendix A). Notably, all relevant variables information was collected from "TIMSS 2019 User Guide for the International Database (Fishbein et al., 2021), its corresponding supplements (T19_UG_Supplement 1) and codebook (T19_G8_Codebook) as well as "TIMSS 2019 International Results in Mathematics and Science" (Mullis et al., 2020). Subsequently, itemized achievement and contextual questionnaire results were extracted using the IDB Analyzer (Version 4.0) (IEA, 2017).

Student-Level Measures

Student Identifier (IDSTUD). An eight-digit identification code that uniquely identified each student participant (Fishbein et al., 2021, p. 86).

Student Total Weight (TOTWGT). Standardized numerical value assigned to each participant. When summed for students within the same school (see IDSCHL), TOTWGT reflects the total amount of participants assessed within that school (Fishbein et al., 2021, p. 83).

Student Gender (L1GND). Gender was measured from the TIMSS 2019 Grade 8 Student Questionnaire variable ITSEX (Eklöf, 2007; Marsh et al., 2014; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004).

Student Math Achievement (L1MACH1-5). Individual math achievement scores variables L1MACH1, L1MACH2, L1MACH3, L1MACH4, L1MACH5 were measured as the five plausible values for the Grade 8 (B), student (S), mathematics (M), overall achievement scale (MAT) (Caponera & Losito, 2016; Chiu, 2012; Eklöf, 2007; Foy, 2017, p. 56; Lui & Meng, 2010; Marsh et al., 2015; Marsh et al., 2014; Mohammadpour & Ghafar, 2014; J. Wang, 2007; Z. Wang, 2015; Wilkins, 2004). Plausible values BSMMAT01, BSMMAT02, BSMMAT03, BSMMAT04, BSMMAT05 were assigned to each student to represent overall math achievement results (Fishbein et al., 2021, p. 71).

Student Science Achievement (L1SACH1-5). Individual student science achievement variables L1SACH1, L1SACH2, L1SACH3, L1SACH4, L1SACH5 were measured by the five plausible values for the Grade 8 (B), student (S), science (S), overall achievement scale (SCI) (Chiu, 2012; Fishbein et al., 2021, p. 71; Gao et al., 2020; Mohammadpour et al., 2015; Wilkins, 2004). Plausible values BSSSCI01, BSSSCI02, BSSSCI03, BSSSCI04, BSSSCI05 were assigned to each student to represent overall science achievement results (Fishbein et al., 2021, p. 71).

Student Self-Concepts in Math (L1MSC). Self-concept in math was measured as a composite of responses to four items from the TIMSS 2019 Grade 8 *Students Confident in Mathematics* questionnaire including "I usually do well in math" (BSBM19A), Math is more difficult for me than many of my classmates" (BSBM19B), "Math is not one of my strengths" (BSBM19C), "I learn things quickly in math" (BSBM19D) (Caponera & Losito, 2016; Chiu, 2012; Lui & Meng, 2010; Marsh et al., 2015; Marsh et al., 2014; Mohammadpour & Ghafar, 2014; Salchegger, 2016; J. Wang, 2007; Wang, 2015; Wilkins, 2004). This Likert-type questionnaire measured the magnitude of agreeableness to math-specific statements based on a scale that ranged from "agree a lot" that received one-point raw score value, "agree a little" received two-point raw score value, "disagree a little" received three point raw score value, and "disagree a lot" received four point raw score value (Mullis et al., 2020). Explicitly, *Students Confident in Mathematics* scale included nine total items with the upper cut points valued at 12.1 indicating that students with an equal or higher raw score value were identified as *very confident in mathematics,* the lower cut point value at 9.5 indicating that those with an equal or lesser value were identified as *not confident in mathematics,* and all of those with raw scores located between the two cut points were identified as *somewhat in mathematics* (Mullis et al., 2020)

Student Self-Concept in Science (L1SSC). Self-concept in science was measured as a composite of responses to four items from the TIMSS 2019 Grade 8 *Students Confident in Science* questionnaire including "I usually do well in science" (BSBS24A), Science is more difficult for me than many of my classmates" (BSBS24B), "Science is not one of my strengths" (BSBS24C), "I learn things quickly in science" (BSBM24D) (Chiu, 2012; Mohammadpour et al., 2015; Wilkins, 2004, p. 200). This Likert-type scales measured magnitude of agreeableness to science-specific statements on a scale identical to that of L1MSC (Mullis et al., 2020). Similarly as well, the *Students Confident in Science* scale included eight items wherein for general/integrated science cut points were located at 10.2 and 8.2. However, for consistency among previous study results, only a portion of the items from the total scale will be applied in this study (Heck et al., 2014, pp. 132–138). Notably, Gao et al. (2020) measured self-efficacy using similar variables from the Students Confident in Science scale. Though, these items will be applied as measures of self-concept and self-efficacy will not be measured in this study.

Student Attitudes Toward Math (L1ATM). Students' attitudes toward math was measured by a composite of responses to three items from the TIMSS 2019 Grade *Students Like Learning Mathematics* questionnaire including "I enjoy learning math" (BSBM16A), "Math is boring" (BSBM16C), "I like Math" (BSBM16E) (Caponera & Losito, 2016; Marsh et al., 2014; Mohammadpour & Ghafar, 2014). This Likert-type scale measured the magnitude of agreeableness to math-specific statements with corresponding raw score values similar to those of L1MSC scale previously described (Mullis et al., 2020). Explicitly, TIMSS 2019 Grade 8 *Students Like Learning Mathematics* scale originally included nine total items with the upper cut point for the math scale valued at 11.4 indicating that students with an equal or higher raw score value are identified as *very much like learning mathematics,* the lower cut point value is at 9.4 indicating that those with an equal or lesser raw score value are identified as *do not like learning mathematics*, and all of those with raw scores located between the two cut points were identified as *somewhat like learning mathematics.*

Student Attitudes Toward Science (L1ATS). Students' attitudes toward science was measured by a composite of responses to three items from the TIMSS 2019 Grade "*Students like Learning Science"* including "I enjoy learning science" (BSBS22A), "Science is boring" (BSBS22C), "I like Science" (BSBS22E) (Mohammadpour et al., 2015; Papanastasiou & Papanastasiou, 2004). This Likert-type scale measured the magnitude of agreeableness to math-specific statements with corresponding raw score values similar to those of L1SSC scale previously described (Mullis et al., 2020). Similar to L1ATM scale cut points, the upper cut point for general/integrated science was located at 10.6 and the lower cut point was located at 8.3.

Student Valuing of Math (L1VOM). Students' valuing of math was measured by a composite of responses to six items from the TIMSS 2019 Grade 8 "*Students Value Mathematics*" questionnaire including "I think learning math will help me in my daily life" (BSBS20A), "I need math to learn other school subjects" (BSBS20B), "I need to do well in math to get to the university of my choice" (BSBS20C), "It is important to do well in math to get ahead in the world" (BSBS20F), "Learning math will give me more job opportunities when I am an adult" (BSBS20G), "It is important to do well in math" (BSBS20I) (Caponera & Losito, 2016; Eklöf, 2007; Mohammadpour & Ghafar, 2014)**.** This Likert-type scale measured agreeableness similar to that of the L1MSC scale previously described (Mullis et al., 2020) Explicitly, TIMSS 2019 Grade 8 Students Value Mathematics scale included nine total items with the upper cut point for the math scale valued at 10.3 indicating that students with an equal or higher raw score value are identified as *strongly value mathematics*, the lower cut point value is at 7.8 indicating that those with an equal or lesser raw score value are identified as *do not value mathematics*, and all of those with raw scores located between the two cut points were identified as *somewhat value mathematics* (Mullis et al., 2020).

Student Valuing of Science (L1VOS). Students' valuing of science was measured by a composite of responses to six items from the TIMSS 2019 Grade 8 "*Students Value Science"* questionnaire including "I think learning science will help me in my daily life" (BSBS25A), "I need science to learn other school subjects" (BSBS25B), "I need to do well in science to get to the university of my choice" (BSBS25C), "It is important to do well in science to get ahead in the world" (BSBS25F), "Learning science will give me more job opportunities when I am an adult" (BSBS25G), "It is important to do well in science" (BSBS25I) (Mohammadpour et al., 2015). This Likert-type scale measured agreeableness similar to that of the L1SSC scale previously described (Mullis et al., 2020). Similar to L1VOM scale cut points, the upper cut point for *Students Value Science* scale was located at 10.8 and the lower cut point was located at 8.5.

Student Socioeconomic Status (L1SES). Student's socioeconomic status was measured by the composite of three resource variables derived from the responses to five general background questions on the TIMSS 2019 Grade 8 S*tudent Questionnaire* (SQG). including "About how many books are there in your home?" (BSBG04), derived variable BSDG05S that included "Do you have your own room at home?" (BSBG05C) and "Do you have an internet connection at home?" (BSBG05D), as well as derived variable BSDGEDUP that included "What is the highest level education completed by your female guardian" (BSBG06A), "What is the highest level education completed by your male guardian" (BSBG06B) (Caponera & Losito, 2016; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Cut points for student resources scale variables indicated that a student with a score greater than or equal to 12.2 had *many resources* if they reported having more than 100 books, at least two home study supports, and at least one parent/guardian that finished university, those with a score no higher than 8.4 had *few resources* and those with any score in between had *some resources* (Mullis et al., 2020).

School-Level Measures

School Identifier (IDSCHL). A four-digit identification code that uniquely identifies each school included in the sample (Fishbein et al., 2021, p. 86).

School-Averaged Math Achievement (L2MACH1-5). School achievement in math variables L2MACH1, L2MACH2, L2MACH3, L2MACH4, and L2MACH5 were aggregated from the five student plausible values of overall math achievement including L1MACH1, L1MACH2, L1MACH3, L1MACH4, L1MACH5 (Chiu, 2012; Marsh et al., 2015; Marsh et al., 2014; Mohammadpour & Ghafar, 2014; Salchegger, 2016; Wang, 2007; Wang, 2015).

School-Averaged Science Achievement (L2SACH1-5). School achievement in science variables L2SACH1, L2SACH2, L2SACH3, L2SACH4, and L2SACH5 were aggregated from the five student plausible values of overall science achievement including L1SACH1, L1SACH2, L1SACH3, L1SACH4, L1SACH5 (Chiu, 2012; Mohammadpour et al., 2015).

School Socioeconomic Status (L2SES). School SES was measured by principle's response to two questions on the TIMSS 2019 Grade 8 *School Questionnaire* (ScQ) including "Approximately what percentage of students in your school come from economically advantaged (BCBG03A) and disadvantages (BCBG03B) homes?" (Caponera & Losito, 2016; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Responses were scaled in ascending order ranging from one-point value that represented the lowest range of 0 to 10% and 5 that represented the highest range of more than 50%. *More affluent* schools reported more than 25% student composition from affluent backgrounds while *more disadvantaged* schools reported more than 25% of the student composition from a disadvantaged background. All other responses were identified as *neither more affluent nor disadvantage* (Mullis et al., 2020).

School Location (L2LOC). School location was measured by the principal's response to one question on the TIMSS 2019 Grade 8 *School Questionnaire* (ScQ) including "Immediate area of school location? (BCBG05B)" (Foy, 2017;

Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Explicitly, the school location scale was measured on the school questionnaire as urban and densely populated, suburban and on the fringe, medium city, small town, or remote.

School Climate (L2CLM). School climate was measured as a composite of responses to eleven items on the TIMSS 2019 *Teacher Questionnaire* concerning *school's emphasis on academic success.* Items included "teacher's understanding of curricular goals" (BCBG14A), Teacher's degree of success in implementing the school's curriculum" (BCBG14B), "Teacher's expectations for student achievement" (BCBG14C), "Teacher's ability to inspire students" (BCBG14D), "Parental involvement in school activities" (BCBG14E), "Parental commitment to ensure that students are ready to learn" (BCBG14F), "Parental expectations for student achievement" BCBG14G, "Parental support for student achievement" BCBG14H, "Students' desire to do well in school" (BCBG14I), "students' ability to reach schools' academic goals" (BCBG14J), and "students respect for classmates who excel (BCBG14K) (Caponera & Losito, 2016; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Explicitly, the *Emphasis on Success* scale placed cut points at 13.1 to represent teacher's report of a school with *very high emphasis,* reports less than or equal to a score of 9.6 represent school's with *medium emphasis*, and scores between those values represent *high emphasis* (Mullis et al., 2020).

School-Averaged Self-Concept in Math (L2MSC). School-level measure of self-concept in math that was aggregated from student-level variable L1MSC.

School-Averaged Self-Concept in Science (L2SSC). School-level measure of self-concept in science that was aggregated from student-level variable L1SSC.

School-Averaged Attitudes Towards Math (L2ATM). School-level measure of attitude toward math that was aggregated from student-level variable L1ATM.

School-Averaged Attitudes Towards Science (L2ATS). School-level measure of attitude toward science that was aggregated from student-level variable L1ATS.

School-Averaged Value of Math (L2VOM). School-level measure for value of math that was aggregated from student-level variable L1VOM.

School-Averaged Value of Science (L2VOS). School-level measure for value of science that was aggregated from student-level variable L1VOS.

Country-Level Variables

Country Identifier (IDCNTRY). Unique Six-digit country identification code based on the ISO 3166 classification (Fishbein et al., 2021, p. 86).

Country-Averaged Math Achievement (L3MACH1-5). Country-level measure of math achievement L3MACH1, L3MACH2, L3MACH3, L3MACH4, L3MACH5 that were aggregated from student-level plausible values for math achievement L1MACH1-5 (Chiu, 2012; Marsh et al., 2014; Mohammadpour & Ghafar, 2014; Wilkins, 2004).

Country-Averaged Science Achievement (L3SACH1-5). Country-level measure of science achievement L3SACH1, L3SACH2, L3SACH3, L3SACH4, L3SACH5 that was aggregated from the student-level plausible values for science achievement L1SACH1-5 (Chiu, 2012; Mohammadpour et al., 2015; Wilkins, 2004)

Cultural Classification (L3IDV). National IDV scale ranges from 0-100 with 0 representing extreme collectivism and 100 representing extreme individualism (Hofstede, 2001).

Gross National Income Per Capita (L3IPC). Country IPC was measured in US dollars and collected from the demographic information provided in *TIMSS 2019 Encyclopedia* (Kelly et al., 2020).

Tracking Practices (L3TRK). National tracking practices were measured from response to question GEN11A of TIMSS 2019 *Country Questionnaire.* Responses were grouping within 3 categories including "no tracking practices," "tracking practices for primary, secondary, and tertiary," tracking practices for tertiary only."

Country-Averaged Self-Concept in Math (L2MSC). Country-level measure of self-concept in math that was aggregated from student-level variable L1MSC.

Country-Averaged Self-Concept in Science (L2SSC). Country-level measure of self-concept in science that was aggregated from student-level variable L1SSC.

Country-Averaged Attitudes Towards Math (L2ATM). Country-level measure of attitude toward math that was aggregated from student-level variable L1ATM.

Country-Averaged Attitudes Towards Science (L2ATS). Country-level measure of attitude toward science that was aggregated from student-level variable L1ATS.

Country-Averaged Value of Math (L2VOM*).* Country-level measure for value of math that was aggregated from student-level variable L1VOM.

Country-Averaged Value of Science (L2VOS). Country-level measure for value of science that was aggregated from student-level variable L1VOS.

Procedures

Study Design

This quantitative study implemented an explanatory, correlational design to explain associations between and among variables of interest, account for variance at multiple levels, and investigate moderation effects using TIMSS 2019 international large-scale, cross-sectional achievement and survey data (Creswell & Guetterman, 2019, p. 345). Uniquely, the TIMSS 2019 data included in this study reflected a nested hierarchical structure wherein students were nested within schools that were nested within countries. Resultingly, analysis of student outcomes “may be misleading unless consequences of group membership are evaluated” (Bickel, 2007, p. 3). Therefore, this design implemented hierarchical linear modeling (HLM) to account for the dependance of student-level scores within the school- and country-levels. Specifically, single -level analysis of variance assumes all variables to be independent of each other. However, in this study, student-level variables were dependent on school and country-level membership, so HLM was most appropriate to reduce spurious results or inflated standard errors (Raudenbush et al., 2019; Tabachnick & Fidell, 2013, p. 787).

TIMSS 2019 Data Collection. Student achievement and background data for this study was collected from the TIMSS 2019 International Database. “To ensure the consistency and uniformity of approach that was necessary for high-quality, internationally comparable data” TIMSS & PIRLS International Study Center, IEA HAMBURG, the IEA Secretariat, Statistics Canada, and the NRC for each participating country collaboratively developed standardized operating procedures that were followed to administer, collect and report TIMSS 2019 results (Johansone, 2020, p. 6.1).

Uniquely, TIMSS 2019 operations and procedures were adapted to control for mode effects of paperTIMSS, eTIMSS and "paperBridge" assessment integrations. For both paperTIMSS and eTIMSS countries, operating procedures were detailed with step-by-step instructions within the seven operations procedures units, supplementary support staff manuals, and necessary software systems that were utilized uniformly throughout the assessment process by all participating countries. Notably, field tests were conducted on all assessment guides and subsequent evaluations of all procedural materials were modified, updated, and translated upon authorization prior to standardized international administrations.

Generally, the NRC was responsible for the organization and supervision of the overall assessment from sampling, scoring, and reporting (p. 6.6). Specifically, the NRC was required to obtain school participation and sample classes, prepare assessment materials and translations, prepare and collect data protection declarations for participants in the European Union and European Economic Area, identify and train school coordinators, manage test administration and scoring of constructed response items, as well as create TIMSS 2019 data files (p. 6.8), and complete Survey activities questionnaire. As well, on site, the school coordinator conducted administration procedures as per instructed by the School Coordinator Manual to obtain voluntary consent and parental permissions, complete student and teacher tracking forms, organize, distribute and collect assessment instruments and materials, train test administrators and return test materials to the national center (p. 6.11). Additionally, the test administrators distributed and proctored assessments as per the Test Administration Manual to maintain strict testing times and conditions.

Students were assigned test booklets according to a "systematic distribution plan implemented by WinW3S sampling software" that linked and labeled student booklets with teacher identifiers for tracking purposes. If participation rate was below 90 percent on the day of administration, a make-up assessment was provided for absent participants before all completed assessment materials were submitted to the national center for scoring (p. 6.11). Upon receipt, scoring was conducted by previously trained scoring staff under the supervision of NRC with inter-rater reliability measures utilized (p. 6.14). Following data collection, data-entry staff referenced IEA HAMBURG international codebooks to enter sampling data into WINW3S database as well as contextual questionnaire responses and achievement item data into the national database using IEA Data Management Expert (DME) software (p. 6.16). Complimentarily, eTIMSS results including constructed response were automatically captured over the IEA eTIMSS server and countries could access the Online Data Monitor to monitor submissions (p. 6. 16).

As well,/ the Online SurveySystem (OSS) offered an online contextual questionnaire platform that required no additional manual data entry. Subsequently, NRCs performed periodic reliability checks of data entry staff through multiple reentries to maintain quality control and assurance of entry accuracy before submitting final scores. Furthermore, upon submission of scores, NRCs completed survey activity questionnaires to evaluate experiences throughout the entire assessment process for necessary improvement considerations on future TIMSS assessments (p. 6.16).

Data Preparation

TIMSS 2019 offered efficient access to all raw data via the IEA IDB Analyzer. Correspondingly, *TIMSS 2019 User Guide for the International Database* provided step-by-step instructions for accessing, merging, and analyzing data using the IDB Analyzer (Fishbein et al., 2021). Therefore, the IDB Analyzer (version 4.0) was employed to locate and merge individually selected math and science achievement items as well data from the school, student, and teacher background questionnaires. Though limited analysis was possible with the IDB analyzer, this study only applied the software to select, retrieve, and merge the data specific to this study into a single SPSS as it was not capable of executing the complex HLM analyses needed to address these research questions. Subsequently, SPSS 27 software was applied for all data preparations to screen and clean then analyze raw data for HLM assumptions (see Appendix B for SPSs syntax) (Green & Salkind, 2017; Tabachnick & Fidell, 2013; Raudenbush et al., 2019). Additionally, RStudio was applied for confirmatory factor analyses (R Core Team, 2017). Thereafter, to address research questions, HLM 8 was applied for all multilevel analyses and Microsoft Excel for necessary manual calculations.

Data Preparation Procedures (see Appendix B for corresponding syntax).

1. Raw data items were selected in IDB Analyzer and the Merge Module was applied to create transferrable SPSS files that only included the itemized raw data from the 26 countries included in this study (Fishbein et al., 2021).
2. All raw data items were merged in SPSS 27 with the *merge variables* and *merge cases* commands to create a single overall data file with variables named in their original TIMSS 2019 coded state (T19G8COMPLETE.data.sav) (Green & Salkind, 2017, p. 55).
3. L3TRK variable was created from variable GEN11A (CQG-11B) (see

Exhibit 17: National Policies Regarding Examinations with Consequences for Students as reported by National Research Coordinator). Dichotomous yes/no responses were recalibrated on an ordinal range from 1-3 (1: “no tracking”; 2: “tracking practices for tertiary placement only”; 3: “tracking practices for primary and/or secondary as well as tertiary placement”). To do so, ID country was recoded as a new variable whereby country ID values were changed to country’s tracking measures.

4. L3IPC was created from info available in TIMSS 2019 Encyclopedia (see Exhibit 1: Selected Characteristics of TIMSS 2019 Countries). To do so, ID country was recoded as a new variable whereby country ID values were changed to country’s gross national income per capita in US dollars.

5. L3IDV was created from Hofstede’s individualism continuum that identified country’s level of individualism as a value that ranged from 1-100 such that 1 represented extreme collectivism and 100 represented extremely individualistic (https://www.hofstede-insights.com/product/compare-countries/). ID country was recoded as a new variable whereby country ID values were changed to the value of each country’s level of individualism.

6. Recoded “omitted or invalid” raw data values 9 and 99 for all ITEMS except plausible values as well as recoded derived value 8 into a new value “SYSMIS” so it would be considered as part of the following missing value analysis.

7. Applied missing value analysis (MVA) then ESTIMATION

MAXIMIZATION (EM) to impute missing data as all missing data was initially less than 5% missing (Marsh, 2019, 2020).

8. Examined Descriptive Statistics of imputed data

9. Computed L1 and L2 Derived Variables (see User Guide Supplement 3 section 2.1 pg. 22-23) **BSDG05S** = BSBG05C + BSBG05D; 0 “neither own room nor internet connection IF (BSBG05C = 2 AND; 3 > 1 “either own room or internet connection”; 2 >2 “both own room and internet connection.” **BSBGEDUP =** max (BSBG06A, BSB06B) original scale retained 1 “Some primary or lower secondary,” 2 “lower secondary”, 3 “upper secondary”, 4 “upper secondary, non-tertiary”, 5 "university or higher". RECODED: 8 > 0 "don’t know"; 6 > 5 "university or higher; 7 > 5 "university or higher". **BCDGSBC** (see User Guide Supplement 3 section 2.4 pg. 37) : 1 “more disadvantaged”; 2 “neither disadvantaged nor affluent”; 3 “more affluent.”

10. Reverse coded all variables so higher numbers indicated a higher value of the construct.

11. Renamed ITSEX to L1GND and Dummy coded L1GND (0 “girl,” 1 “boy”).

12. Renamed BCBG05B to L2LOC dummy code L2LOC (0 “rural,” 1 “urban”).

13. Rename BSMMAT01-05 to L1MACH1-5 and BSSCI01- 05 to L1SACH1

14. Renamed BCDGSBC to L2SES.

15. Examined correlations/covariance of Raw data (including all individual

items) with SPSS ANALYZE – CORRELATE – BIVARIATE.

16. Conducted Principal Component Analysis (PCA) and Alpha Cronbach's reliability analysis in SPSS (TIMSS 2019 Methods and Procedures Technical Report CH.16: Creating Contextual questionnaires scales, pg. 16.168).

17. Conducted Confirmatory Factor Analysis (CFA) in R to compare with PCA results.

18. COMPOSITED L1 and L2 items into single construct scales.

19. AGGREGATED all L1 variables to L2 and L3.

20. Examined DESCRIPTIVE STATISTICS and HISTOGRAMS of all Final variables.

21. Identified univariate outliers $z < 3$.

22. Identified multivariate outliers using Mahalanobis distance.

Statistical Analysis Overview

TIMSS 2019 reported hierarchically structured data from students clustered within schools that were also clustered within countries. This structure constituted a dependance among observations within the cluster, which upon the application of single-level analysis such as ANOVA or OLS would violate the assumption of independence and contribute to statistically spurious conclusions based on underestimated standard errors from the lack of consideration for within group variability (Heck et al., 2014, p. 7). Accordingly, hierarchical linear modeling (HLM) does not require the independence of errors and produces unbiased estimates by allowing intercepts (mean) and slopes (relationships between predictor and outcome) to vary between higher-level units of the nested structure (Tabachnick & Fidell, 2013, p. 787). However, when applying HLM with an insufficient sample size, there is a greater likelihood of incorrectly rejecting a true null hypothesis (inflated type one error rate) and a reduction in statistical power or type two error rate (the probability of accepting a untrue null hypothesis) (Bickel, 2007).

Therefore, this study applied HLM analysis to investigate the associations between a single outcome variable and multiple predictors at three levels with an exceptionally large sample size to increase the statistical power and appropriately represent “structural relations and residual variability occurring at each level” (Raudenbush et al., 2019, p. 11). Nonetheless, with such a large sample size as in this study, power was indeed increased, but small interaction effects could have appeared significant yet practically negligible, so results were reported for $p < 0.001$ and $p < 0.05$ levels of significance (Seaton, 2010).

On the other hand, as previously mentioned, plausible values were multiply imputed scores for each student derived from the distribution of available observed data points. Consequently, analysis of plausible values required that parameter estimates were provided for each value separately then averaged across all plausible values to produce unbiased results (Rutkowski et al., 2010). Unfortunately, SPSS 27 could not handle these models as processing time exceeded 2 hours for one model, so that software was only utilized for data preparations. However, even though HLM8 processing time was instantaneous, the software was not capable of internally averaging plausible values when specified at two levels simultaneously as it was in these models (i.e., L1MACH and L2MACH for BFLPE models). Therefore, all models were run five times in HLM8 software then all results were transferred to Microsoft Excel where they were averaged to determine final estimates of fixed effects and variance components. Also, same level interactions for moderation analyses could not be specified in HLM8 either, so they were created as a single variable in SPSS then specified accordingly in HLM8.

Additionally, Rutkowski et al. (2010) recommended applying TOTWGT when investigating student-level outcomes as it reflects the actual student sample size in each country, so this study applied TOTWGT in all three-level HLM analyses across countries. Additionally, *full maximum likelihood (ML)* was applied in HLM8 to "estimate random and fixed components by maximizing their joint likelihood" rather than "estimating the random effects averaged over all possible fixed components as in restricted maximum likelihood (REML)" (Tabachnick & Fidell, 2013, p. 837). Generally, ML is a superior choice to REML when comparing model improvement by way of deviance reduction (Bickel, 2007). Also, due to the extreme number of iterations that were required to execute these models in HLM8, the iteration settings were changed to "continue iterating" if convergence was not achieved at 100. Moreover, internal hypothesis testing settings were filled with "deviance and numbers of parameters" for corresponding null hypothesis for each subject.

Overall, HLM8 software applied three-level HLM models to first examine the presence of school- and country-level BFLPE across countries (RQ1) (Marsh et al., 2008, p. 200, 2019, 2020; Marsh & Hau, 2003; Seaton et al., 2009). Then, three-level HLM models discretely examined how associations of student-, school- and country-level predictors effected student self-concept in math and science across countries (RQ2) (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Last, three-level HLM models investigated moderation effects of school- and country-level BFLPE in math and science across countries (Seaton et al., 2010). Distinctly, the null model was first analyzed for each subject to partition variability in the outcome that was attributed to each distinct level of analysis. Followed by fixed effects models then random effects models were analyzed to address each research question, while additional interaction effects models were finally analyzed to examine moderation effects. Explicitly, this study first addressed models with only respective math variables then addressed the same models again with only respective science variables.

Analytical Procedures.

1. Null models were analyzed for both subjects: **Model 0a** null model for math specified L1MSC was specified as the outcome; random intercepts were specified at L1, L2, L3 with no predictors included. **Model 0b** (science null model) was specified as the outcome; random intercepts were specified at L1, L2, L3 with no predictors included.

2. Research question 1 "Does BFLPE exist at the school- and country-level across all countries in math and science?" was addressed for **L2BFLPE** by applying fixed effects **model 1a** with uncentered predictors then random coefficient **model 1b** with grand mean centered predictors (Marsh et al., 2019; Marsh et al., 2020). **L3BFLPE** was addressed by applying fixed effects **model 2a** with uncentered predictors then random coefficient **model 2b** with grand mean centered predictors (see Table 8 for model equations). All models were repeated five times replacing each model with corresponding, consecutive plausible values. L2BFLPE and L3BFLPE models were analysed discretely for math then science variables.

L2BFLPE Specifications.

- **RQ1 Model 1a** math specifications included L1MSC as the outcome, random intercepts at L1, L2, L3, L1MACH1-5 and L2MACH1-5 as *fixed* predictors
- **RQ1 Model 1a** science specifications included L1SSC as the outcome, random intercepts at L1, L2, L3, L1SACH1-5 and L2SACH1-5 as *fixed* predictors.
- **RQ1 Model 1b** math specifications included L1MSC as the outcome, random intercepts at L1, L2, L3, L1MACH1-5 was specified as *randomly varying* at L2 and L3. L2MACH1-5 was specified as randomly varying at L3.
- **RQ1 Model 1b s**cience specifications included L1SSC as the outcome, random intercepts at L1, L2, L3. L1SACH1-5 was specified as *randomly varying* at L2 and L3 and L2MACH1-5 was specified as randomly varying at L3.

L3BFLPE Specifications.

- **RQ1 Model 2a** Math specifications included L1MSC as the outcome, random intercepts at L1, L2, L3, L1MACH1-5 and L3MACH1-5 as *fixed* predictors.
- **RQ1 Model 2a** Science Specifications included L1SSC as the outcome, random intercepts at L1, L2, L3, L1SACH1-5 and L3SACH1-5 as *fixed* predictors.

- **RQ1 Model 2b** Math Specifications included L1MSC as the outcome, random intercepts at L1, L2, L3, L1MACH1-5 was specified as *randomly* varying at L2 and L3. L3MACH1-5 was specified as fixed.
- **RQ1 Model 2b** Science Specifications included L1SSC as the outcome, random intercepts at L1, L2, L3, L1SACH1-5 was specified as *randomly* varying at L2 and L3 and L3SACH1-5 was specified as fixed.

3. Research question 2 “Is student-level math and science self-concept significantly associated with student-level achievement, gender, self-concepts, socioeconomic status, valuing and attitudes toward math and science, school-level achievement, socioeconomic status, location, climate, academic self-concept, valuing and attitudes toward math and science or country-level achievement, income per capita, classification of individualism, tracking practices, self-concepts, valuing and attitudes toward math and science across countries?” was addressed with fixed effects **Model 1a** and for random coefficient **Model 1b** with uncentered, student-level predictors (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Fixed effects **Model 2a** and for random coefficient **Model 2b** with uncentered, school-level predictors and fixed effects **Model 3a** only for with uncentered, country-level predictors as L3 predictors cannot be specified as random. Notably, all models were run discretely for each individual predictor. Also, all models were repeated five times replacing each model with corresponding, consecutive plausible values and all models were analysed discreetly for effects in math then effects in science.

Student-Level Predictor Specifications.

- **RQ2 Model 1a** Math specifications included L1MSC with random intercepts at L1, L2, L3 and *fixed* predictors L1VOM, L1ATM, L1SES, L1GND, L1MACH1-5.
- **RQ2 Model 1a** Science specifications included L1SSC as the outcome with random intercepts at L1, L2, L3 and *fixed* predictors L1VOS, L1ATS, L1SES, L1SACH1-5, L1GND.
- **RQ2 Model 1b** Math specifications included L1MSC with random intercepts at L1, L2, L3 and predictors L1VOM, L1ATM, L1SES, L1GND, L1MACH1-5 specified as *random* at L2 and L3.
- **RQ2 Model 1b** Science specifications included L1SSC as the outcome with random intercepts at L1, L2, L3 and predictors L1VOS, L1ATS, L1SES, L1SACH1-5, L1GND specified as *random* at L2 and L3.

School-Level Predictor Specifications.

- **RQ2 Model 2a** Math specifications included L1MSC with random intercepts at L1, L2, L3 and fixed predictors L2VOM, L2ATM, L2SES, L2MACH1-5; L2LOC.
- **RQ2 Model 2a** Science specifications included L1SSC as the outcome with random intercepts at L1, L2, L3 and fixed predictors L2VOS, L2ATS, L2SSC, L2SES, L2CLM, L2SACH1-5, L2LOC.
- **RQ2 Model 2b** Math specifications included L1MSC with random intercepts at L1, L2, L3 and predictors L2VOM, L2ATM, L2SES, L2MACH1-5; L2LOC specified as random at L3.

- **RQ2 Model 2b** Science specifications included L1SSC as the outcome with random intercepts at L1, L2, L3 and predictors L2VOS, L2ATS, L2SSC, L2SES, L2CLM, L2SACH1-5, L2LOC specified as random at L3.

Country-Level Predictor Specifications.

- **RQ2 Model 3a** Math specifications included L1MSC with random intercepts at L1, L2, L3 and fixed predictors L3VOM, L3ATM, L3MSC, L3TRK, L3MACH1-5, L3IDV, L3IPC
- **RQ2 Model 3a** Science specifications included L1SSC as the outcome with random intercepts at L1, L2, L3 and fixed predictors L3VOS, L3ATS, L3IPC, L3SSC, L3SACH1-5, L3TRK, L3IDV, L3IPC.

4. Same level interactions were created as a single variable in SPSS including L2SESINT = L2SES*L2MACH1-5; L3MSCINT = L3MSC * L3MACH1-5; L3VOMINT = L3VOM * L3MACH1-5; L2SSCINT = L2SSC*L2SACH1-5; L3ATSINT = L3ATS*L3SACH1-5.

5. Research question 3 "Is school- or country-level one variable BFLPE moderated across countries by student-, school-, or country-level variables found to be significantly associated with student-level self-concept in math and science?" was addressed with only the significantly associated variables from RQ2 that were applied as moderators (Seaton et al., 2010). Concerning moderation effects on **L2BFLPE**, **Model 1a** examined moderation effects of significant student-level predictors, **Model 2a** examined moderation effects of the significant school-level predictors, **Model 3a** examined moderation effects of the

significant country-level predictors. Concerning moderation effects on **L3BFLPE**, **Model 1b** examined moderation effects of significant student-level predictors, **Model 2b** examined moderation effects of the significant school-level predictors, **Model 3b** examined moderation effects of the significant country-level predictors (see Table 8 for model equations). Notably, all models were repeated five times replacing each model with corresponding, consecutive plausible values, all models were run discretely for each moderation effect, and all models were analysed discretely for effects in math then effects in science.

L2BFLPE Moderation Specifications.

- **RQ3 Model 1a -** Math specifications included L1MSC with random intercepts at L1, L2, L3, moderator L1VOM and covariate L1MACH1-5 specified as random at L2 and L3, covariate L2MACH1-5 and L2MACH1-5*L1VOM interaction specified as random at L3.
- **RQ3 Model 1a -** Science specifications included L1SSC with random intercepts at L1, L2, L3, moderator L1VOS, L1ATS, and covariate L1SACH1-5 specified as random at L2 and L3, covariate L2SACH1-5 and L2SACH1-5*L1VOS, L2SACH1-5*L1ATS interactions specified as random at L3.
- **RQ3 Model 1b -** Math specifications included L1MSC with random intercepts at L1, L2, L3, covariate L1MACH1-5 was specified as random at L2 and L3, moderator L2SES, covariate L2MACH1-5 and L2SESINT1-5 were specified as random at L3.

- **RQ3 Model 1b –** Science specifications included L1SSC with random intercepts at L1, L2, L3, covariate L1SACH1-5 specified as random at L2 and L3, covariate L2SACH1, moderator L2SSC, and interaction L2SSCINT1-5 specified as random at L3.
- **RQ3 Model 1c –** Math specifications included L1MSC with random intercepts at L1, L2, L3, covariate L1MACH1-5 specified as random at L2 and L3, covariate L2MACH1-5 specified as random at L3, moderator L3MACH1-5 and interaction L3MACH1-5*L2MACH1-5.
- **RQ3 Model 1c-** There were no country-level moderators of L2BFLPE in science.

L3BFLPE Moderation Specifications.

- **RQ3 Model 2a -** Math specifications included L1MSC with random intercepts at L1, L2, L3, moderator L1ATM, L1VOM, L1MACH1-5 and covariate L1MACH1-5 specified as random at L2 and L3, covariate L3MACH1-5 and L3MACH1-5*L1VOM, L3MACH1-5*L1ATM, L3MACH1-5*L1MACH.
- **RQ3 Model 2a -** Science specifications included L1SSC with random intercepts at L1, L2, L3, moderator L1SES, L1ATS, L1SACH1-5 and covariate L1SACH1-5 specified as random at L2 and L3, covariate L3SACH1-5 and interactions L3SACH1-5*L1SES, L3SACH1-5*L1ATS, L2SACH1-5*L1SACH1 interactions specified as random at L3.
- **RQ3 Model 2b –** Math specifications included L1MSC with random intercepts at L1, L2, L3, covariate L1MACH1-5 was specified as random

at L2 and L3, moderators L2ATM, L2MSC, L2VOM were specified as random at L3, including covariate L3MACH1-5 and L2ATM*L3MACH1-5, L2MSC*L3MACH1-5, L2VOM*L3MACH1-5.

- **RQ3 Model 2b -** Science specifications included L1SSC with random intercepts at L1, L2, L3, covariate L1SACH1-5 was specified as random at L2 and L3, moderators L2CLM, L2SES, L2VOS were specified as random at L3, including covariate L3SACH1-5 and L2CLM*L3SACH1-5, L2SES*L3SACH1-5, L2VOS*L3SACH1-5.
- **RQ3 Model 2c -** Math specifications included L1SSC with random intercepts at L1, L2, L3, covariate L1MACH1-5 was specified as random at L2 and L3, including moderators L3MSC and L3VOM, covariate L3MACH1-5, L3MSCINT1-5 and L3VOMINT1-5 interactions.
- **RQ3 Model 1c-** Science specifications included L1SSC with random intercepts at L1, L2, L3, covariate L1SACH1-5 was specified as random at L2 and L3, including moderator L3ATS, covariate L3SACH1-5 and L3ATSINT1-5 interaction.

Chapter 4: Results

Introduction

Chapter 4 begins with details of sample demographics. Next, results of raw data descriptive statistics and percent of missing values (MVA) for raw data items that were downloaded directly from TIMSS 2019 IDB Analyzer was reported. Missing values of raw data items were then imputed using estimation maximization (EM) in SPSS28 and corresponding descriptive statistics were reported to compare imputed items with raw items statistics. After, imputed items were reverse coded, renamed, derived, composited and/or aggregated to form the final variables that were analyzed for HLM assumptions and later included in the main HLM analyses. Thereafter, results of descriptive statistics, tests of normality of distributions, tests of univariate and multivariate outliers, multicollinearity, linearity, and homoscedasticity in SPSS28 were reported for analyses of HLM assumptions.

Thereafter, results of analyses of the measurement model were reported. First, results of correlation matrices of final variables with imputed data in SPSS28 were reported. Next, results for Principal Component Analysis (PCA) in SPSS28 of imputed items were reported. Then, results of Confirmatory Factor Analysis (CFA) in R for final variable constructs that were created from PCA results were reported. Finally, to address research questions, results of three-level, HLM analyses in HLM8.2 were reported.

Sample Demographics

Overall, 169,957 students in 5,410 schools from 26 countries that participated in both TIMSS 2019 Grade 8 math and science as a single integrated subject were analyzed in the present study (see Table 1). Of those, eight East Asian and Pacific countries, ten

Table 1

Sample Demographics

Country Code	Region	Country	# of schools	# of Students	Year Assessed	Average age	Math score	Science Score	Test format	IPC	TRK	IDV
036	EAP	Australia	284	9060	year 8	14.1	517	528	p	54910	3	90
048	MENA	Bahrain	112*	5725	grade 8	13.8	481	486	p	2210*	3	25
152	LA	Chile	164	4115	Basic 8	14.2	441	462	e	15010	2	23
158	EAP	Chinese Taipei	203	4915	grade 8	14.3	612	574	E	25501	3	17*
818	MENA	Egypt	169	7210	grade 8	13.9	413	389	p	2690	3	25
926	ECA	England	136	3365	Year 9	14.0	515	517	e	42370	3	89
344	EAP	Hong Kong SAR	136	3265*	Secondary2	14.1	578	504	e	50840	3	25
372	ECA	Ireland	149	4118	Second Year	14.4	524	523	p	62210	3	70
364	MENA	Islamic Republic of Iran	220	5980	grade 8	14.1	446	449	p	5420	3	41
376	MENA	Israel	157	3731	grade 8	14.0	519	513	e	43290	3	54
380	ECA	Italy	158	3619	Lower Secondary Grade 3	13.7	497	500	e	34460	3	76
392	EAP	Japan	142	4446	Grade 8	14.4	594	570	p	41690	3	46
400	MENA	Jordan	235	7176	grade 8	13.9	420	452	p	4300	2	30
414	MENA	Kuwait	171	4574	Grade 8	13.8	403	444	p	34290	3	25
458	EAP	Malaysia	177	7065	grade 8	14.3	461	460	e	11200	3	26
554	EAP	New Zealand	134	6051	Year 9	13.9	482	499	P	42670	2	79
578	ECA	Norway9	157	4575	grade 9	14.7	503	495	e	82500	3	69
512	MENA	Oman	228	6751	grade 8	13.9	411	457	p	15330	3	25
634	MENA	Qatar	152	3884	grade 8	14.0	443	475	e	63410	3	25
410	EAP	Republic of Korea	168	3861	Middle School grade 2	14.5	607	561	e	33720	2	18
682	MENA	Saudi Arabia	209	5680	grade 8	13.9	394	431	P	22850	2	25
702	EAP	Singapore	153	4853	Secondary 2	14.3	616**	608**	e	59590	3	20
710	SSA	South Africa 9	519	20829	grade 9	15.5	389*	370*	P	6040	2	65
792	ECA	Turkey	181	4077	grade 8	13.9	496	515	e	9610	3	37
784	MENA	United Arab Emirates	623**	22334**	grade 8	13.7	473	473	e	43470	2	25
840	NA	United States	273	8698	grade 8	14.2	515	522	e	65760**	1	91**

Note. Total participants = 169, 957 students in 5,410 schools from 26 countries. Test format p = paperTIMSS 2019, e = eTIMSS 2019. ** highest score/rank, *lowest score/rank.

Middle East and North African, five European and Central Asian, as well as one Sub-Saharan African, Latin American, and North American country were analyzed. Generally, participants represented an overall averaged income per capita estimated at 31, 823 USD including mostly middle to high income countries with Egypt as the only lower middle income country (Kelly et al., 2020; The World Bank, 2020). Notably, Singapore ranked first in math achievement with an average score of 616 and in science achievement with an average score of 608, while South Africa ranked lowest in math achievement with an average score of 389 and in science achievement with an average score of 370. Overall, students were on average 14.1 years old with the oldest averaged age at 15.5 years old in South Africa and the youngest in United Araba Emirates at 13.7 years old. Gender represented an even distribution of males and females.

Analyses of HLM Assumptions

Henceforth, *raw data* referred to itemized data that were downloaded from TIMSS 2019 IDB Analyzer, *imputed data* referred to raw data after missing values were imputed, *initial final variables* referred to imputed data items that were recoded, renamed, derived, composited and/or aggregated, and *final variables* referred to modified initial final variables that were analysed for HLM assumptions then specified in main HLM models to answer research questions.

To begin, contextual and achievement raw data in math and science that was downloaded from the TIMSS 2019 IDB Analyzer were screened for descriptive statistics and percentage of missing values using SPSS 28 (See Table 2). Missing data analyses results showed that student-level raw data was missing less than 5% with most school-level raw data missing at less than 6%. The greatest amount of missing data was found for measures of school economic disadvantaged (8.3% missing) and measures of school economic advantaged (11.9% missing). Consequently, consistent with other studies, all

Table 2

Raw Items Descriptive Statistics

Item	N	Min	Max	*M (SEM)*	*SD*	*Skew*	*Kurtosis*	Missing Count	Missing %
BSSSCI02	169957	5	873.4	478.66(0.28)	113.43	-0.29	-0.21	0	0.0
BSSSCI03	169957	11.5	858.35	479.84(0.27)	112.66	-0.29	-0.21	0	0.0
BSSSCI04	169957	5	851.31	478.14(0.28)	114.16	-0.28	-0.22	0	0.0
BSSSCI05	169957	5	865.72	479.07(0.28)	113.86	0.29	-0.22	0	0.0
BSMMAT01	169957	60.02	905.72	477.48(0.27)	109.68	0.13	-0.36	0	0.0
BSMMAT02	169957	5	902.9	477.98(0.27)	110.4	0.13	-0.35	0	0.0
BSMMAT03	169957	5	902	477.97(0.27)	111.04	0.13	-0.35	0	0.0
BSMMAT04	169957	10.62	922.7	476.93(0.27)	111.63	0.13	-0.36	0	0.0
BSMMAT05	169957	72.21	911.1	477.82(0.27)	111.17	0.13	-0.36	0	0.0
ITSEX	169898	1	2	1.5 (0.001)	0.5	0.01	-2	59	0.0
BSBG04	166820	1	5	2.52 (0.003)	1.28	0.51	-0.75	3137	1.8
BSBG05C	163411	1	2	1.3 (0.001)	0.46	0.86	-1.26	6546	3.9
BSBG05D	166582	1	2	1.15 (0.001)	0.36	1.94	1.78	3375	2.0
BSBG06A	162581	1	9	5.37 (0.006)	2.4	-0.29	-1.33	7376	4.3
BSBG06B	162065	1	9	5.49 (0.006)	2.39	-0.33	-1.31	7892	4.6
BSBS22A	165542	1	4	1.76 (0.002)	0.89	1.03	0.26	4415	2.6
BSBS22C	162944	1	4	2.93 (0.003)	1	-0.5	-0.97	7013	4.1
BSBS22E	163865	1	4	1.83 (0.002)	0.94	0.9	-0.21	6092	3.6
BSBM16A	166576	1	4	2.01 (0.002)	0.97	0.68	-0.54	3381	2.0
BSBM16C	163300	1	4	2.67 (0.003)	1.07	-0.15	-1.24	6657	3.9
BSBM16E	164142	1	4	2.13 (0.003)	1.05	0.51	-0.96	5815	3.4
BSBS24A	164262	1	4	1.86 (0.002)	0.87	0.79	-0.09	5695	3.4
BSBS24B	163912	1	4	2.8 (0.002)	1.01	-0.29	-1.06	6045	3.6
BSBS24C	162955	1	4	2.66 (0.003)	1.06	-0.13	-1.22	7002	4.1
BSBS24D	162811	1	4	1.97 (0.002)	0.92	0.58	-0.61	7146	4.2
BSBM19A	165385	1	4	2.01 (0.002)	0.93	0.64	-0.45	4572	2.7
BSBM19B	165063	1	4	2.6 (0.003)	1.04	-0.06	-1.17	4894	2.9
BSBM19C	163627	1	4	2.47 (0.003)	1.1	0.08	-1.3	6330	3.7
BSBM19D	164238	1	4	2.17 (0.002)	0.96	0.37	-0.84	5719	3.4
BSBS25A	164155	1	4	1.67 (0.002)	0.85	1.17	0.61	5802	3.4
BSBS25B	163563	1	4	1.96 (0.002)	0.95	0.6	-0.68	6394	3.8
BSBS25C	163061	1	4	1.78 (0.002)	0.94	0.94	-0.2	6896	4.1
BSBS25F	162800	1	4	1.78 (0.002)	0.91	0.94	-0.08	7157	4.2
BSBS25G	162880	1	4	1.77 (0.002)	0.91	0.97	-0.05	7077	4.2
BSBS25I	163113	1	4	1.59 (0.002)	0.83	1.33	1.07	6844	4.0
BSBM20A	165718	1	4	1.72 (0.002)	0.9	1.13	0.39	4239	2.5
BSBM20B	165227	1	4	1.92 (0.002)	0.91	0.74	-0.32	4730	2.8
BSBM20C	164714	1	4	1.61 (0.002)	0.85	1.32	0.88	5243	3.1
BSBM20F	164283	1	4	1.77 (0.002)	0.9	0.98	0.06	5674	3.3
BSBM20G	164491	1	4	1.64 (0.002)	0.84	1.2	0.77	5466	3.2
BSBM20I	164952	1	4	1.5 (0.002)	0.78	1.6	2.01	5005	2.9
BCBG03A	155863	1	4	2.49 (0.003)	1.19	0.03	-1.51	14094	8.3
BCBG03B	149767	1	4	2.19 (0.003)	1.18	0.4	-1.37	20190	11.9
BCBG05B	160145	1	5	2.43 (0.003)	1.28	0.34	-1.1	9812	5.8
BCBG14A	160873	1	5	1.82 (0.002)	0.69	0.44	-0.05	9084	5.3
BCBG14B	160500	1	5	1.97 (0.002)	0.71	0.26	-0.32	9457	5.6
BCBG14C	160492	1	5	2.05 (0.002)	0.75	0.31	-0.14	9465	5.6
BCBG14D	160522	1	5	2.12 (0.002)	0.76	0.23	-0.24	9435	5.6
BCBG14E	160658	1	5	2.95 (0.003)	1.04	0.04	-0.47	9299	5.5
BCBG14F	160713	1	5	2.85 (0.003)	1.02	0.15	-0.38	9244	5.4
BCBG14G	160528	1	5	2.26 (0.002)	0.91	0.5	0.08	9429	5.5
BCBG14H	160353	1	5	2.73 (0.002)	1	0.22	-0.3	9604	5.7
BCBG14I	160525	1	5	2.36 (0.002)	0.83	0.3	0.12	9432	5.5
BCBG14J	160252	1	5	2.47 (0.002)	0.8	0.11	0.08	9705	5.7
BCBG14K	160325	1	5	2.2 (0.002)	0.85	0.57	0.44	9632	5.7
L3TRK	169957	1	3	2.49 (0.001)	0.6	-0.68	-0.5	0	0.0
L3IDV	169957	17	91	45 (0.06)	25.16	0.65	-1.16	0	0.0

missing raw data was imputed (Marsh et al., 2019, 2020).

Specifically, expectation maximization was applied in SPSS28 to substitute missing data with conditional expected values of assumed distributions and parameter estimates of observed values followed by full maximum likelihood estimation to produce complete datasets with no missing values (Tabachnick & Fidell, 2013). Critics have contested that EM is not beneficial for ordinal data, such as that of TIMSS 2019 contextual, raw data items, because it assumes a normal distribution and does not include error in the imputations, so it underestimates standard errors (Graham, 2009, p. 561). However, descriptive statistics of raw data (see Table 2) and imputed data (see Table 3) showed negligible differences in standard errors of measurement.

Accordingly, imputed data items were reverse coded to reflect higher values as a larger amount of the construct in the main HLM analyses. As well, dichotomous items L1GND was dummy coded to reference only boy students and L2LOC to reference only rural school locations as well (Cohen et al., 2003, p. 303). Thereafter, all remaining imputed data items were renamed, derived, and/or composited to create initial final variables that were similar in construct to those of TIMSS 2019. Last, relevant student-level, initial final variables were aggregated to school-and country-levels (see Appendix B). Explicitly, analyses of HLM assumptions began with screening of initial final variables (N=169,957) for descriptive statistics and normality of distributions. Generally, a perfectly normal distribution (bell curve) has a value of zero for skewness and kurtosis with symmetrically distributed values to right and left of the mean, a balanced concentration of values closest to the mean, and asymptotic right and left tails in a histogram. However, for samples > 300, absolute skewness values > 2 and absolute kurtosis values > 7 represent substantial departures from normality (Curran et al., 1996).

Table 3

Imputed Items Descriptive Statistics

Item	n	Min	Max	*M (SEM)*	*SD*	*Skew*	*Kurtosis*
BSBG04	169957	1	5	2.52 (0.003)	1.30	0.52	-0.71
BSBG05C	169957	1	2	1.3 (0.001)	0.45	0.88	-1.20
BSBG05D	169957	1	2	1.15 (0.001)	0.36	2.00	1.90
BSBG06A	169957	1	9	5.36 (0.006)	2.40	-0.29	-1.30
BSBG06B	169957	1	9	5.49 (0.006)	2.30	-0.33	-1.30
BSBS22A	169957	1	4	1.76 (0.002)	0.88	1.05	0.34
BSBS22C	169957	1	4	2.93 (0.002)	1.00	-0.51	-0.90
BSBS22E	169957	1	4	1.83 (0.002)	0.93	0.91	-0.14
BSBM16A	169957	1	4	2.01 (0.002)	1.00	0.69	-0.50
BSBM16C	169957	1	4	2.67 (0.003)	1.05	-0.16	-1.20
BSBM16E	169957	1	4	2.13 (0.003)	1.04	0.51	-0.92
BSBS24A	169957	1	4	1.86 (0.002)	0.85	0.80	-0.01
BSBS24B	169957	1	4	2.8 (0.002)	1.00	-0.29	-1.00
BSBS24C	169957	1	4	2.66 (0.003)	1.04	-0.13	-1.16
BSBS24D	169957	1	4	1.97 (0.002)	0.90	0.59	-0.53
BSBM19A	169957	1	4	2.01 (0.002)	0.91	0.64	-0.40
BSBM19B	169957	1	4	2.6 (0.002)	1.02	-0.06	-1.12
BSBM19C	169957	1	4	2.47 (0.003)	1.08	0.08	-1.25
BSBM19D	169957	1	4	2.17 (0.002)	0.95	0.38	-0.79
BSBS25A	169957	1	4	1.67 (0.002)	0.84	1.19	0.72
BSBS25B	169957	1	4	1.96 (0.002)	0.93	0.61	-0.60
BSBS25C	169957	1	4	1.78 (0.002)	0.92	0.96	-0.10
BSBS25F	169957	1	4	1.78 (0.002)	0.89	0.95	0.03
BSBS25G	169957	1	4	1.77 (0.002)	0.90	0.98	0.05
BSBS25I	169957	1	4	1.6 (0.002)	0.81	1.36	1.21
BSBM20A	169957	1	4	1.72 (0.002)	0.86	1.14	0.46
BSBM20B	169957	1	4	1.92 (0.002)	0.90	0.75	-0.25
BSBM20C	169957	1	4	1.61 (0.002)	0.84	1.33	0.97
BSBM20F	169957	1	4	1.77 (0.002)	0.89	0.99	0.13
BSBM20G	169957	1	4	1.64 (0.002)	0.83	1.24	0.86
BSBM20I	169957	1	4	1.51 (0.002)	0.77	1.61	2.13
BCBG03A	169957	1	4	2.48 (0.003)	1.14	0.05	-1.39
BCBG03B	166957	1	4	2.17 (0.003)	1.12	0.46	-1.18
BCBG05B	169957	1	5	2.43 (0.003)	1.24	0.35	-0.99
BCBG14A	169957	1	5	1.82 (0.002)	0.67	0.45	0.11
BCBG14B	169957	1	5	1.97 (0.002)	0.69	0.26	-0.18
BCBG14C	169957	1	5	2.04 (0.002)	0.73	0.32	0.03
BCBG14D	169957	1	5	2.12 (0.002)	0.74	0.23	-0.08
BCBG14E	169957	1	5	2.95 (0.002)	1.01	0.04	-0.34
BCBG14F	169957	1	5	2.85 (0.002)	1.00	0.16	-0.23
BCBG14G	169957	1	5	2.26 (0.002)	0.89	0.51	0.26
BCBG14H	169957	1	5	2.73 (0.002)	0.96	0.22	-0.15
BCBG14I	169957	1	5	2.73 (0.002)	0.96	0.22	-0.15
BCBG14J	169957	1	5	2.36 (0.002)	0.81	0.29	0.29
BCBG14K	169957	1	5	2.47 (0.002)	0.77	0.11	0.25

Notes. Descriptive statistics displayed for only imputed items. BSSSCI01-5, BSMMAT01-5, L3TRK, L3IPC, and L3IDV were not imputed.

Results of descriptive statistics and histograms for initial final variables showed that most values of skewness and kurtosis were within the range of normality with the exception of several school-level variables (see Appendix C).

Initial final variables were then analyzed in SPSS28 for univariate and multivariate outliers. Univariate outliers represent extreme values from the sample mean of an independent variable, while multivariate outliers represent extreme values of combinations of independent variables that can potentially distort statistics (Tabachnick & Fidell, 2013, p. 72). To screen for univariate outliers in SPSS28, initial final variable's values were transformed to z-scores then compared in ascending order for values > ±3.29 for $p < 0.001$. Results indicated 101 cases for L1SACH2 and 46 cases for L1MACH4 as univariate outliers. Therefore, 147 cases were removed further reducing the sample size of initial final variables from 169,957 to 169,810. Additionally, significance tests of Mahalanobis Distance (M-D) were conducted in SPSS28 using equation [1-cdf.chisq(M-D, df)] where cdf.chisq is the cumulative distribution function of the chi squared distribution, M-D is the Mahalanobis distance, and df is the amount of variables included in calculation of M-D. Results of M-D test presented no multivariate outliers to consider ($p < .001$).

Last, final variables ($N = 169{,}810$) were screened for descriptive statistics. Similar to initial final variables, final variables appeared to be normally distributed for all student- and country-level continuous variables, except for L2SACH1-5 (kurtosis ≈ 7.5), L2ATM (kurtosis = 9.11), L2SSC (kurtosis = 9.31), and L2MSC (kurtosis = 15.54) that displayed leptokurtic values > 7 for kurtosis (see Table 4, Table 5, Table 6 for final variable descriptive statistics). For additional substantiation, significance tests of normality included a one-sample, Kilmogorov-Smirnov (KS) test in SPSS28, as well as manual calculations of comparisons of z scores with zero using a z distribution for values of skewness and kurtosis [$z = (S-0)/S_s$; $z = (K-0)/S_k$], where S is the absolute value of skewness, S_s is the standard error of skewness, K is the absolute value of kurtosis, and S is the standard error of kurtosis (Tabachnick & Fidell, 2013). Results of KS and z score tests were statistically significant ($p < 0.05$) suggesting a significant departure from normality for all final variable distributions. Nevertheless, Tabachnick and Fidell (2013)

Table 4
Student-Level Final Variable Descriptive Statistics

Variable Name	N	Min	Max	Mean	SE	SD	Skewness	Kurtosis
L1SACH1	169810	5	863	478.51	0.28	113.92	-0.29	-0.215
L1SACH2	169810	5	873	478.66	0.28	113.43	-0.29	-0.212
L1SACH3	169810	12	858	479.84	0.27	112.66	-0.29	-0.213
L1SACH4	169810	5	851	478.14	0.28	114.16	-0.28	-0.223
L1SACH5	169810	5	866	479.07	0.28	113.86	-0.29	-0.219
L1MACH1	169810	60	906	477.48	0.27	109.68	0.13	-0.362
L1MACH2	169810	5	903	477.98	0.27	110.40	0.13	-0.349
L1MACH3	169810	5	902	477.97	0.27	111.04	0.13	-0.354
L1MACH4	169810	11	923	476.93	0.27	111.63	0.13	-0.359
L1MACH5	169810	72	911	477.82	0.27	111.17	0.13	-0.361
L1GND	169810	0	1	0.50	0.00	0.50	-0.01	-2.000
L1SES	169810	2	12	8.22	0.00	2.00	-0.17	-0.288
L1ATS	169810	3	12	9.26	0.01	2.40	-0.60	-0.342
L1ATM	169810	3	12	8.49	0.01	2.63	-0.37	-0.719
L1SSC	169810	4	16	11.54	0.01	2.80	-0.10	-0.455
L1MSC	169810	4	16	10.84	0.01	2.97	-0.04	-0.489
L1VOS	169810	6	24	19.10	0.01	4.59	-0.80	-0.113
L1VOM	169810	6	24	19.56	0.01	4.23	-1.01	0.526

Table 5
School-Level Final Variable Descriptive Statistics

Variable Name	*N*	Minimum	Maximum	Mean	Std. err	Std. Dev.	Skewness	Kurtosis
L2SACH1	169957	230	620	478.51	0.09	36.57	-2.09	7.49
L2SACH2	169957	231	616	478.66	0.09	36.31	-2.08	7.45
L2SACH3	169957	235	619	479.84	0.09	35.86	-2.07	7.34
L2SACH4	169957	224	620	478.14	0.09	36.42	-2.07	7.50
L2SACH5	169957	221	615	479.07	0.09	36.45	-2.12	7.71
L2MACH1	169957	317	606	477.48	0.07	30.63	-1.50	4.51
L2MACH2	169957	305	609	477.98	0.07	30.79	-1.50	4.50
L2MACH3	169957	312	610	477.97	0.07	30.88	-1.46	4.31
L2MACH4	169957	310	609	476.93	0.08	30.96	-1.49	4.41
L2MACH5	169957	311	604	477.82	0.08	30.93	-1.52	4.53
L2ATS	169957	6	12	9.26	0.00	0.37	0.60	6.46

L2ATM	169957	6	12	8.49	0.00	0.40	1.88	9.11
L2SSC	169957	10	16	11.54	0.00	0.38	1.30	9.31
L2MSC	169957	9	16	10.84	0.00	0.35	1.61	15.54
L2VOM	169957	15	24	19.56	0.00	0.67	1.38	5.40
L2VOS	169957	15	24	19.10	0.00	0.70	1.43	5.44
L2SES	169957	1	3	1.89	0.002	0.79	0.20	-1.39
L2LOC	169957	0	1	0.77	0.00	0.42	-1.30	-0.30
L2CLM	169957	11	48	25.78	0.02	6.85	0.02	-0.18

Note. Results of descriptive statistics shown for final variables with univariate outliers N = 169,810; *KS test results $p < 0.05$.

contended that significance tests of normality for large samples are unreliable as "the null hypothesis of a normal distribution is likely to be rejected when there are only minor deviations from normality" (p.80).

Furthermore, for clustered data, the sampling distribution of means (distributions of means from random samples of each variable discretely) must be normally distributed, but with very large populations, the Central Limit Theorem, affirms the distribution of sample means will approach normality regardless of the distribution of the variables (pg. 78). Therefore, severely leptokurtic values were retained, and parametric analyses were applied using final variables (N=169,810) to later answer research questions using HLM analyses with robust standard errors to obtain unbiased standard errors and meet the HLM assumption of homoscedasticity.

Table 6

School-Level Final Variable Descriptive Statistics

Variable Name	*N*	Minimum	Maximum	Mean	Std. err	Std. Dev.	Skewness	Kurtosis
L3SACH1	169957	385.32	601.39	478.5102	0.13708	56.51412	-0.006	-0.536
LSACH2	169957	386.42	601.38	478.6606	0.13624	56.16650	0.000	-0.534
L3SACH3	169957	389.37	602.34	479.8401	0.13475	55.55083	0.017	-0.526
L3SACH4	169957	386.30	602.02	478.1395	0.13669	56.35117	0.014	-0.525
L3SACH5	169957	385.99	602.74	479.0653	0.13694	56.45662	-0.005	-0.524

L3MACH1	169957	401.17	609.19	477.4759	0.15215	62.72575	0.561	-0.511
L3MACH2	169957	401.48	611.67	477.9806	0.15317	63.14689	0.566	-0.506
L3MACH3	169957	400.37	611.23	477.9654	0.15388	63.43966	0.553	-0.512
L3MACH4	169957	398.79	610.89	476.9278	0.15476	63.80252	0.550	-0.519
L3MACH5	169957	399.75	611.40	477.8236	0.15443	63.66385	0.552	-0.522
L3ATS	169957	7.54	10.37	9.2587	0.00157	0.64573	-0.403	-0.551
L3ATM	169957	7.30	9.53	8.4864	0.00167	0.68794	0.163	-1.403
L3SSC	169957	9.51	12.59	11.5412	0.00182	0.75213	-0.938	0.339
L3MSC	169957	9.14	11.74	10.8387	0.00151	0.62399	-0.797	0.041
L3VOS	169957	16.13	21.03	19.0989	0.00340	1.40204	-0.603	-0.999
L3VOM	169957	16.18	21.39	19.5584	0.00283	1.16785	-0.665	0.976
L3IPC	169957	2690	82500	31822.98	53.631	22109.929	0.276	-1.024
L3TRK	169957	1	3	2.49	0.001	0.593	-0.679	-0.502
L3IDV	169957	17.00	91.00	45.0033	0.06102	25.15805	0.645	-1.159

Distinctly, there has been great debate regarding the robustness of parametric analyses with non-normal data that dates back to the initial categorization of levels of measurement and the subsequent analyses that is appropriate for each (Stevens, 1946). Earliest reports summarized by Glass et al., (1972) highlighted numerous simulation studies that “found little effect of non-normality on two tailed *t*-and *F*-tests (pg. 246).” He cited studies that suggested for fixed effects ANOVA models with large sample sizes the distribution of *t* is independent of shape. For example, Pearson (1931) reported the equivalence of Type I error probabilities for ANOVA with two groups having skewed and normally distributed samples (Glass et al., 1972, pg. 247). Similarly, (Norton, 1952, as cited in Lindquist, 1953) reported “minor discrepancies on nine different points in the F-distribution even for small sample sizes,” while Boneau (1960) compared violations of the normality assumption for the two-group *t*-test of various significance levels and sample sizes with few negligible differences reported (Glass et al., 1972, pg. 247). Additionally, Norman (2010) supported the robustness of parametric tests for their "ability to give the right answer even when statistical assumptions are violated to an extreme degree" (Sullivan & Artino Jr., 2013, p. 546).

Analyses of Measurement Model

First, Principal Component Analysis (PCA) in SPSS28 was applied to summarize the patterns of correlations among imputed items and examine construct validity of TIMSS 2019 scales (see Table 7). Results extracted similar factor structures as those specified in TIMSS 2019 context questionnaire scales including socioeconomic status (L1SES), valuing of math, (L1VOM), valuing of science (L1VOS), attitude toward math (L1ATM), attitude toward science (L1ATS), science self-concept (L1SSC), and math self-concept (L1MSC), while two school-level factors were extracted including socioeconomic status (L2SES) and climate (L2CLM) (Yin & Fishbein, 2020). Overall, factor loadings ranging from 0.66 to .90 and total variances explained ranged from 47.72% for L1SES to 77.33% for L2SES. Cronbach alpha reliability results were comparable to those reported by TIMSS 2019 ranging from the lowest reliability statistic found for L1SES at 0.42 and highest statistic for L2CLM at 0.92 (Yin & Fishbein, 2020, p. 168).

Afterwards, Confirmatory Factor Analysis in R 4.0.3 was conducted in R software to verify the factor structures suggested by PCA results (see Appendix C for Confirmatory Factor Analysis Results). WLSMV was applied to estimate model parameters and goodness-of-fit measures with no distributional assumptions. Additionally, CFI $\geq$ 0.95, RMSEA $\leq$ 0.06, SRMR $\leq$ 0.08, and TLI $\geq$ 0.95 were used as cut-off values (Hu & Bentler, 1999). Model 1 included all six student-level factors, Model 2 included the two school-level factors extracted by PCA, and Model 3 included the nested model with all student- and school-level factors. Regarding model fit, Model 1 showed acceptable fit χ^2 (356, N = 169,957) = 186148.642, $p < .001$, CFI = 0.961, TL1 = 0.955, RMSEA = 0.055 except for SRMR = 0.053. Similarly, Model 2 showed acceptable fit for all measures χ^2 (64, N = 169,957) = 61639.154, $p < 0.001$, CFI = 0.973, TLI = 0.967, RMSEA = 0.075, SRMR = 0.074. Model 3 showed acceptable fit as well for the nested model χ^2 (674, N = 169,957) = 259006.007, $p < 0.001$, CFI = 0.963, TLI = 0.960, RMSEA = 0.047, SRMR = 0.048.

Table 7

Principal Component Analysis Results

TIMSS Factor Name	Chi Square (DF)	Cumulative variance explained (%)	Cronbach Alpha Reliability	Imputed Item	Loadings
L1SES	21920.83 (3)**	47.72	0.42	BSBG04	0.72
				BSDGEUP	0.69
				BSDG05S	0.67
L1ATM	206764.20 (3) **	73.73	0.82	BSBM16A	0.89
				BSBM16C	0.79
				BSBG16E	0.90
L1ATS	180969.01 (3) **	70.02	0.78	BSBS22A	0.89
				BSBS22C	0.72
				BSBS22E	0.90
L1MSC	168705.96 (6) **	55.43	0.73	BSBM19A	0.74
				BSBM19B	0.72
				BSBM19C	0.77
				BSBM19D	0.75
L1SSC	167334.08 (6) **	53.25	0.71	BSBS24A	0.73
				BSBS24B	0.71
				BSBS24C	0.74
				BSBM19D	0.74
L1VOM	471434.03 (15)**	62.07	0.90	BSBM20A	0.77
				BSBM20B	0.74
				BSBM20C	0.77
				BSBM20F	0.83
				BSBM20G	0.82
				BSBM20I	0.80
L1VOS	597351.95 (15) **	68.09	0.91	BSBS25A	0.82
				BSBS25B	0.79
				BSBS25C	0.84
				BSBS25F	0.85
				BSBS25G	0.85
				BSBS25I	0.81

L2SES	60292.75 (1) **	77.33	-2.41	BCBG03A	0.88
				BCBG03B	-0.89
L2CLM	1200616.52 (52)**	56.29	0.92	BSBG14A	0.65
				BSBG14B	0.72
				BSBG14C	0.71
				BSBG14D	0.71
				BSBG14E	0.75
				BSBG14F	0.84
				BSBG14G	0.73
				BSBG14H	0.82
				BSBG14I	0.80
				BSBG14J	0.80
				BSBG14K	0.70
L1SES	21920.83 (3)**	47.72	0.42	BSBG04	0.72
				BSDGEUP	0.69

Last, results of correlation analysis in SPSS28 of all final variables including confirmed factors reflected no multicollinearity. However, all level-specific plausible values were Last, results of correlation analysis in SPSS28 of all final variables including confirmed factors reflected no multicollinearity. However, all level-specific plausible values were indeed highly correlated ($r_{L1MACH1\text{-}5)}$ ≈ 0.935; $r_{L2MACH1\text{-}5)}$ ≈ 0.997) and L3MACH1-5 completely redundant (r = 1.00) (see Appendix D). However, as main HLM analyses must be run discretely for each variable, all plausible values were retained.

Analysis of Hierarchical Linear Models

TIMSS 2019 reported hierarchically structured data from students nested within various schools that were also nested within various countries. Therefore, applying single-level analyses of variance (ANOVA) would violate the assumption of independence of observations, inflate the Type I error rate (likelihood of rejecting the null hypothesis when it is true), deflate standard errors of regression coefficients, and may contribute to statistically spurious conclusions from the lack of consideration for within group dependance of student-level observations (Heck et al., 2014). Even though, hierarchical linear modeling (HLM) does require independence of observations between higher level clusters (i.e., school- and classroom-level observations), unlike ANOVA, it uniquely accounts for the dependance of student-level observations within higher level clusters. As such, HLM produces unbiased estimates of variability in the outcome attributed from all levels by allowing intercepts (predicted value of outcome when predictor is at value 0) and slopes (value of outcome with 1-unit increase in predictor) to vary between higher-level units of the nested structure (Tabachnick & Fidell, 2013, p. 787).

Therefore, HLM was applied in the main analysis of this study that employed HLM 8 software to investigate fixed- and random effects for multileveled predictors of students' academic self-concept in math and science (Raudenbush et al., 2019, p. 11). Specifically, three-level HLM analyses first examined associations of covariates for L2BFLPE and L3BFLPE in math and science to confirm the presence of BFLPE across all 26 TIMSS 2019 countries in this study (see RQ1 in Table 8). Next, RQ2 examined the significance of student-, school-, and country-level predictors of science academic self-concept in math and science (see RQ2 in Table 8). Third, the significant predictors of ASC that were identified in RQ2 were then applied as moderators to examine moderation of L2BFLPE and L3BFLPE in math and science. Explicitly, fixed effects models were analyzed to determine initial significance of associations followed by random coefficient models examined to determine school-to-school and country-to-country variability in those associations (Marsh et al., 2020; Seaton et al., 2010).

Furthermore, TIMSS 2019 assessments were designed using a matrix sampling technique wherein each student only answered a portion of the total number of assessment questions, thus precise measures of students' individual proficiency were not reported. Instead, TIMSS reported plausible values as an estimation of proficiency scores. Specifically, plausible values are a set of five scores derived from the distribution of multiply imputed scores to account for the portion of assessment items not answer due to matrix sampling for each student. Subsequently, all respective analyses including plausible values required parameter estimates measured for each value separately then averaged across all plausible values to produce unbiased results (Rutkowski et al., 2010). As well correspondence with HLM8.2 associates, M. du Toit and S. Raudenbush confirmed that HLM8.2 software was incapable of internally producing averaged results for plausible values modelled simultaneously at more than one level (personal communication, January 28, 2022). Therefore, all models that specified plausible values were run five times in HLM8.2, once for each plausible value. Results were then manually transferred to Microsoft Excel where they were averaged to determine final parameter estimates.

Moreover, Rutkowski et al. (2010) and B. Fishbein, a TIMSS 2019 correspondent, recommended applying TOTWGT when investigating student-level outcomes as it sums to the student sample size in each country (personal communication, November 29, 2018). Therefore, TOTWGT was specified as the estimation setting in HLM 8.2 for all three-level analyses. Similarly, *full maximum likelihood (ML)* was applied by default in HLM8 to "estimate random and fixed components by maximizing their joint likelihood" rather than "estimating the random effects averaged over all possible fixed components as in restricted maximum likelihood (REML)" (Tabachnick & Fidell, 2013, p. 837). Generally, ML is a superior choice to REML when comparing model improvement by way of deviance reduction of all saturated models from the null model that perfectly fits the data and includes no predictors with only random intercepts at all levels (Bickel, 2007; Heck et al, 2014 , p 14). Additionally, to internally produce deviance results by means of likelihood ratio test, deviance and parameters estimates from corresponding null models in math (*-2LL* = 849877.02, *df* = 4) and science (*-2LL*= 814270.05, *df* = 4) were filled in the HLM8.2 hypothesis testing setting. Also, as some models exceeded the default 100 iteration setting in HLM8.2, iterations were set to "continue iterations" when maximum number of iterations was achieved, so that all results reported here were derived from models that converged.

Research Question #1 (RQ1)

Does school- and country-level BFLPE in math and science exist across 26 TIMSS 2019 countries (Marsh et al., 2019; Marsh et al., 2020)? To answer RQ1, three-level HLM models were analyzed in HLM8.2 to examine the presence of school-level L2BFLPE and country-level L3BFLPE in math and science. Notably BFLPE is present when there is a negative effect of school-averaged . achievement (L2BFLPE) or country-averaged achievement (L3BFLPE) on student-level self-concept, when a positive relationship between student-level academic self-concept and corresponding student-level achievement simultaneously persists.

Model 1a examined the presence of L2BFLPE in math and science by modelling school-averaged achievement in math (L2MACH1-5) then modelling school-averaged achievement in science (L2SACH1-5) as predictors of student-level self-concept in math (L1MSC) then science (L1SSC), while controlling for corresponding student-level achievement in math (L1MACH1-5) or science (L1SACH1-5) across all countries. Model 1a that examined L2BFLPE in math included L1MSC specified as the outcome with L2MACH as a fixed predictor and L1MACH as a fixed covariate with intercepts permitted to vary at the student-, school-, and country levels.

Model 1a that examined L2BFLPE in science included L1SSC specified as the outcome with L2SACH as a fixed predictor and L1SACH as a fixed covariate with intercepts permitted to vary at the student-, school-, and country levels. Notably, all models were run five times with each corresponding plausible value then results were

Table 8

HLM Model Equations

Research Question	Variable	Model	Math Equation	Science Equation
Null			L1MSC = G000+ r0 + u00 + e	L1SSC = G000+ r0 + u00 + e
RQ1	L2BFLPE fixed	Model1a	L1MSC = G000 + G010*L2MACH1 + G100*L1MACH1+ r0 + u00 + e	L1SSC = G000 + G010*L2SACH1 + G100*L1SACH1+ r0 + u00 + e
RQ1	L2BFLPE Random (GMC)	Model1b	L1MSC = G000 + G010*L2MACH3 + G100*L1MACH3+ r0 + r1*L1MACH3 + u u01*L2MACH3 + u10*L1MACH3 + e	L1SSC = G000 + G010*L2SACH3 + G100*L1SACH3+ r0+r1* L1SACH3 + u +u01*L2SACH3 + u10*L1SACH3 + e
RQ1	L3BFLPE fixed	Model2a	L1MSC = G000 + G001*L3MACH1 + G100*L1MACH1+ r0 + u00 + e	L1SSC = G000 + G001*L3SACH1 + G100*L1SACH1+ r0 + u00 + e
RQ1	L3BFLPE Random (GMC)	Model2b	L1MSC = G000 + G001*L3MACH3 + G100*L1MACH3+ r0 +r1*L1MACH3+ u00 + u10*L1MACH3 + e	L1SSC = G000 + G001*L3SACH3 + G100*L1SACH3+ r0 + r1*L1SACH3+ u u10*L1SACH3 + e
RQ2	L1ATM fixed	Model1a	L1MSC = G000 + G100*L1ATM+ r0 + u00 + e	L1SSC = G000 + G100*L1ATS+ r0 + u00 + e
RQ2	L1ATM Random	Model1b	L1MSC = G000 + G100*L1ATM + r0 + r1*L1ATM+ u00 + u10*L1ATM + e	L1SSC = G000 + G100*L1ATS + r0 + r1 u00 + u10*L1ATS + e
RQ2	L2ATM Fixed	Model2a	L1MSC = G000 + G010*L2ATM+ r0 + u00 + e	L1SSC = G000 + G010*L2ATS+ r0 + u00 + e
RQ2	L2ATM random	Model2b	L1MSC = G000 + G010*L2ATM+ r0 + u00 + u01*L2ATM + e	L1SSC = G000 + G010*L2ATS+ r0 + u00 + u01*L2ATS + e
RQ2	L3ATM fixed	Model.3a	L1MSC = G000 + G001*L3ATM+ r0 + u00 + e	L1SSC = G000 + G001*L3ATS+ r0 + u00 + e
RQ3	L2BFLPE INT L1 random	Model1a	L1MSC =G000 + G010*L2MACH1 + G100*L1MACH1 + G200*L1VOM + G21 L1VOM*L2MACH1 + r0 + r1*L1MACH1 r2*L1VOM+ u00 + u01*L2MACH1 +u10* L1MACH1+ u20*L1VOM+u21* L1VOM*L2MACH1+e	L1SSC = G000+G010*L2SACH5 +G100*L1SACH5+G200*L1ATS+ G21 L1ATS*L2SACH5 + r0 + r1*L1SACH5+ r2*L1ATS+ u00 + u01*L2SACH5+u10* L1SACH5 +u20*L1ATS+u21*L1ATS* L2SACH5 + e
RQ3	L2BFLPE INT L2 random	Model2a	L1MSC = G000 + G010*L2MACH1 + G020*L2SES + G030*L2SES1INT + G100*L1MACH1+ r0 +r1*L1MACH1+ u00 + u01*L2MACH1 + u02*L2SES + u03*L2SES1IN + u10*L1MACH1 + e	L1SSC = G000+G010*L2SACH1 +G020*L2SSC+G030* L2SSC1NT + G100*L1SACH1+ r0+ r1*L1SACH1 +u00+u01*L2SACH1+ u02*L2ATS+u03*L2ATS1IN + u10*L1SACH1 + e

RQ3	L2BFLPE INT L3 random	Model3a	L1MSC = G000 + G001*L3MACH1 + G010*L2MACH1 + G011*L2MACH1 *L3MACH1 + G100*L1MACH1 + r0 + r1*L1MACH1+ u00 + u01*L2MACH1 + u10*L1MACH1 + e	No L3 moderators of L2BFLPE in science
RQ3	L3BFLPE INT L1 random	Model1b	L1MSC = G000 + G001*L3MACH1 + G100*L1MACH1 + G200*L1ATM + G201*L1ATM*L3MACH1 + r0 + r1*L1MACH1+ r2*L1ATM+ u00 +u10*L1MACH1 +u20*L1ATM + e	L1SSC = G000+ G001*L3SACH5 + G100*L1SACH5 +G200*L1ATS+ G201*L1ATS*L3SACH5 + r0 + r1* L1SACH5+ r2*L1ATS + u00 + u10*L1SACH5 + u20*L1ATS + e
RQ3	L3BFLPE INT L2 random	Model2b	L1MSC = G000 + G001*L3MACH1 + G010*L2ATM + G011*L2ATM* L3MACH1 G100*L1MACH1+ r0 +r1*L1MACH1+ u00 + u01*L2ATM + u10*L1MACH1 + e	L1SSC = G000 + G001*L3SACH1 + G010*L2VOS + G011*L2VOS* L3SAC + G100*L1SACH1+ r0 + r1*L1SACH1+ u00 + u01*L2VOS + u10*L1SACH1 + e
RQ3	L3BFLPE INT L3 random	Model3b	L1MSC = G000 + G001*L3MSC + G002*L3MACH1 + G003*L3MSC1INT + G100*L1MACH1 + r0 + r1*L1MACH1 + u00+ u10*L1MACH1 + e	L1SSC = G000 + G001*L3ATS + G002*L3SACH3 +G003*L3ATS1INT + G100*L1SACH3 + r0 + r1*L1SACH3+ u00 + u10*L1SACH3 + e

averaged for final estimates. Equations for L2BFLPE Model 1a were similar for math and science with corresponding predictors interchanged accordingly (see Table 8 RQ1 Model 1a). For example, Equation 1:

L1MSC = G000 + G010*L2MACH1 + G100*L1MACH1+ r0 + u00 + e (1)

represented the fixed effects model for L2BFLPE in math, wherein L1MSC was modeled as a function of L1MACH covariate and L2MACH predictor with only L1MSC intercept specified as random at school and country-level. Statistically, notation *G000* reflected the grand mean or the collective school and country grouping effect on the L1MSC random intercept. G010*L2MACH1 represented the country effect on the slope value for L2MACH or predicted value of L1MSC with 1-unit increase in L2MACH when L2MACH is at value 0 (uncentered). G100*L1MACH1 represented the school and country effect on the slope value for the student-level covariate (L1MACH) interpreted as the predicted value of L1MSC with 1-unit increase in L1MACH when L1MACH is at value 0 (uncentered). Notation *e* represented the variability in L1MSC attributed to student differences, while *r0* represents the variability in L1MSC random intercept attributed to school differences, and *u00* represents variability in random intercept attributed to country differences (Bickel, 2007, p. 223).

Model 1b permitted intercepts and slopes to vary randomly between schools and countries reflecting variation in slope and intercepts attributed to school and country differences. Fixed effects for Model 1b included those from Model 1a, but L1MACH was permitted to vary across schools and countries and L2BFLPE across countries (See Table 8). All predictors were grand mean centered (GMC) for Model 1b in math and science due to singularity error in HLM8.2 when running the models uncentered. This did not change the statistical significance of results, instead only shifted the random effect estimates when the predictor was at value 0, to random effect estimates when the predictor is at the total sample's average. Therefore, Model 1b coefficient estimates can be interpreted as the change in students' self-concept with 1-unit increase in the predictor from its grand mean. Notably, all corresponding models were run five times for each plausible value then results were averaged for final estimates. Equations for L2BFLPE Model 1b were similar for math and science with corresponding subject predictors interchanged accordingly (see Table 8 RQ2 Model1b). For example, Equation 2:

$$L1MSC = G000 + G010*L2MACH3 + G100*L1MACH3 + r0 + r1*L1MACH3 + u00 + u01*L2MACH3 + u10*L1MACH3 + e \quad (2)$$

reflected random effects of L2BFLPE predictors. Notations were synonymous with those of Model 1a with the addition of r1*L1MACH3 that represented the variability in L1MACH random slope attributed to school differences, u10*L1MACH3 that represented variation in L1MACH random slope attributed to country differences, and u01*L2MACH3 that represented variability in L2MACH random slope attributed to country differences.

Additionally, Model 2a examined the presence of L3BFLPE by investigating associations between country-level achievement in math (L3MACH1-5) and science (L3SACH1-5) with student-level self-concept in math (L1MSC) and science (L1SSC), while controlling for student-level achievement in math (L1MACH1-5) and science (L1SACH1-5) across 26 TIMSS 2019 countries (see Table 8). Fixed effects for Model 2a specified L1MSC as the outcome variable, L3MACH1-5 as an uncentered predictor and L1MACH1-5 as an uncentered covariate with only intercepts at three-levels as random effects. Similarly, fixed effects Model 2a in science included L1SSC as the outcome variable, L1SACH1-5 as an uncentered covariate and L3SACH1-5 as an uncentered predictor with only intercepts at all three-levels included as random effects. Equations for L3BFLPE Model 2a were similar for math and science with corresponding subject predictors interchanged accordingly (see Table 8 RQ1 Model 2a). For example, Equation 3:

$$L1SSC = G000 + G001*L3SACH1 + G100*L1SACH1 + r0 + u00 + e \qquad (3)$$

represents fixed-effects of L3BFLPE in science. Notations generally follow those of Model 1a except for G001*L3SACH1 that represents school and country grouping effects on the slope of L3SACH.

Correspondingly, Model 2b examined country-to-country variation of L3BFLPE in math and science. Predictors were grand mean centered to address singularity issues. Fixed effects for Model 2b included those from Model 2a, while intercepts were permitted to vary across schools and countries and L1MACH slope was permitted to vary across schools and countries. Equations for L3BFLPE Model 2b are similar for math and science with corresponding subject predictors interchanged accordingly (See Table 8 RQ1 Model 2b). For example, Equation 4:

$$L1MSC=G000 + G001*L3MACH1 + G100*L1MACH1+ r0 + u00 + u10*L1MACH1 + e \quad (4)$$

reflects notations that are similar to those of previous models.

RQ1 L2BFLPE Results. Results for Model 1a in math confirmed the presence of L2BFLPE. Specifically, results reflected a significant negative effect of L2MACH on L1MSC (β = -0.006, *S*D = 0.001, p < .001) with a positive effect of L1MACH (β = 0.017, *SE* = 0.001, p < .001) across all TIMSS 2019 countries (see Table 9 Model 1a, Figure G11, and Figure G13).

Table 9

Results for L2BFLPE and L3BFLPE in Math

	Model 1a L2BFLPE		Model 1b L2BFLPE		Model 2a L3BFLPE		Model 2b L3BFLPE	
Fixed Effects	β	*SE*	β	*SE*	β	*SE*	β	*SE*
L1MACH	0.017*	0.001*	0.018*	0.001*	0.017*	0.001*	0.018*	0.001*
L2MACH L2BFLPE	-0.006*	0.001*	-0.005**	0.002**				
L3MACH L3BFLPE					-0.026*	0.003*	-0.017*	0.003*
Random Effects	β	*SD*	β	*SD*	β	*SD*	β	*SD*
L1 Res Var	6.577*	2.565*	7.039*	2.653*	2.565*	6.578*	6.330*	2.516*
L2 Res Var	0.743*	0.862*	6.332*	2.516*	0.871*	0.759*	7.159*	2.675*
L2 Res Var L1MACH			0.00002*	0.005*			0.00002*	0.005*
L3 Res Var	2.497*	1.580*	10.220*	3.197*	0.472*	0.223*	4.840*	2.200*
L3 Res Var L1MACH			0.00002*	0.005*			0.00002 *	0.005*
L2MACH			0.00001*	0.003*				
-2LL	39323.02		42901.44		39307		42906.94	

Notes. Results displayed for L2BFLPE and L3BFLPE covariates and predictors in math. Fixed and random effects for L2BFLPE are shaded in light gray and L3BFLPE are shaded in dark gray. Sample size = 169,810 students, 5,410 schools, 26 countries.
* $p < 0.001$, **$p < 0.05$.

For all student math achievement levels, as school achievement increased, math self-concept decreased, whereby lower achieving students reflected the lowest levels of math self-concept and high achieving students reflected higher math self-concepts (see Figure G14). Yet, students in schools with the lowest overall averaged math achievement reflected the highest levels of math self-concept as their math achievement improved, while highest averaged-achieving schools reflected students with the lowest math self-concepts as their math achievement improved (see Figure G12). Model 1b in math confirmed significant school-to-school ($\beta = 6.332$, $SD = 2.516$, $p < .001$) and country-to-country variation in L2BFLPE ($\beta = 10.220$, $SD = 3.197$, $p < .001$).

Table 10

Results of L2BFLPE and L3BFLPE in Science

	Model 1a L2BFLPE		Model 1b L2BFLPE		Model 2a L3BFLPE		Model 2b L3BFLPE	
Fixed Effects	β	*SE*	β	*SE*	β	*SE*	β	*SE*
L1SACH	0.012*	0.0007*	0.013*	0.003*	0.012*	0.0003*	0.013*	0.001*
L2SACH L2BFLPE	-0.004*	0.001*	-0.004**	0.001**				
L3SACH L3BFLPE					-0.024*	0.004*	-0.022*	0.005*
Random Effects	β	*SD*	β	*SD*	β	*SD*	β	*SD*
L1 Res Var	5.849*	2.418*	5.791*	2.406*	5.849*	2.419*	5.793*	2.407*
L2 Res Var	0.588*	0.767*	0.882**	0.937**	0.596*	0.772*	0.911**	0.952**
L2 Res Var L1SACH			0.0003	0.001			0.0000*	0.003*
L3 Res Var	2.149*	1.466*	4.747*	2.175*	0.653*	0.808*	3.020*	1.738*
L3 Res Var L1SACH			0.00001*	0.003*			0.00001*	0.003*
-2LL	24144.35		25344.02		24114.74		25310.97	

Notes. Results displayed for L2BFLPE and L3BFLPE covariates and predictors in . Fixed and random effects for L2BFLPE are shaded in light gray and L3BFLPE are shaded in dark gray. Sample size = 169,810 students, 5,410 schools, 26 countries.

* $p < 0.001$, **$p < 0.05$.

Similarly, results for Model 1a in science also indicated the presence of L2BFLPE, whereby there was a significant negative effect of L2SACH (β = -0.004, *SE* = 0.001, p < .001) with a significant positive effect of L1SACH (β = 0.012, *SE* = 0.0007, p < .001) on L1SSC (see Table 10 Model 1a, Figure G15, Figure G17). Results indicated that, for all students' science achievement levels, as school-averaged science achievement increased, students' science self-concept decreased, whereby lower achieving students reflected the lowest levels of science self-concept and high achieving students reflected higher science self-concepts (see Figure G18). Yet, students in schools with the lowest averaged science achievement reflected the highest levels of science self-concept as their science achievement improved, while highest averaged-achieving schools reflected students with the lowest science self-concepts as their science achievement improved (see Figure G16). Results for Model 1b in science confirmed significant school-to-school (β = 0.882, *SD* = 0.937, p < .001) and country-to-country variation in L2BFLPE (β = 4.747, *SD* = 2.175, p < .001) as well.

RQ1 L3BFLPE Results. Results for Model 2a in math confirmed the presence of L3BFLPE. Specifically, results reflected a significant negative effect of L3MACH on L1MSC (β = -0.026, *SE* = 0.003, p < .001) with a significant positive effect of L1MACH (β = 0.017, *SE* = 0.001, p < .001) across all TIMSS 2019 countries (see Table 9 Model 2a, Figure G11, Figure G19). Figure G20 shows that for all students, as country-averaged achievement increased, student's self-concept in math decreases, wherein high math achievers reflected the highest overall self-concept in math, high math achievers were impacted most by L3BFLPE (steepest rate of change in slope). Likewise, students in countries with the lowest averaged math achievement reflected the highest levels of math self-concept as their own math achievement increased, while highest averaged-achieving countries reflected students with the lowest math self-concepts as their own math achievement improved (see Figure G21). Model 2b in math confirms significant country-to-country variation in L3BFLPE (β = 4.840, *SD* = 2.200, *p* < .001) (see Table 9 Model 2b).

Results for Model 2a in science indicated the presence of L3BFLPE. Specifically, results reflected a significant negative effect of L3SACH on L1SSC (β = -0.024, *S*D = 0.004, p < .001) with a positive effect of L1SACH (β = 0.012, *SE* = 0.0003, p < .001) across all TIMSS 2019 countries (see Table 10 Model 2a, Figure G22, and Figure G15).

Figure G24 shows that for all students, as country-averaged achievement increased, student's self-concept in science decreased, wherein high science achievers reflected the highest overall self-concept in science, they were impacted most by L3BFLPE (steepest rate of change in slope). Likewise, students in countries with the lowest averaged science achievement reflected the highest levels of science self-concept as their science achievement increased, while highest averaged-achieving countries reflected students with the lowest science self-concepts as their own science achievement improved (see Figure G23). Results for Model 2b in science confirmed significant country-to-country variation in L3BFLPE ($\beta = 3.020$, *SD* = 1.738, $p < .001$) (see Table 10 Model 2b).

Research Question 2 (RQ2)

Is student-level academic self-concept in math and science significantly associated with student-level achievement, gender, self-concepts, socioeconomic status, value and attitude toward math or science, school-level achievement, socioeconomic status, location, climate, academic self-concept, value and attitude toward math or science and country-level achievement, income per capita, classification of individualism, tracking practices, self-concepts, value and attitude toward math or science across 26 TIMSS 2019 countries (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014)?

To answer RQ2, three-level, intercept-only, null models were initially examined for math and science. Specifications for the null model included only L1MSC or L1SSC as the outcome and random intercepts at three-levels with no predictors. For example, the null model for math Equation 5:

$$\text{L1MSC} = \text{G000} + \text{r0} + \text{u00} + \text{e} \quad (5)$$

reflects a random intercept only model wherein students' math self-concept is modeled as a function of the collective school and country effect on the random intercept ($G000$) or grand mean plus the residual variance of intercept attributed to country effects ($u00$), school effects ($r0$), and student effects (e) (see Table 8 Null Model). Results were applied to calculate Intraclass Coefficient (ICC) that determined variation in math or science self-concept attributed at each level as well as to calculate the baseline deviance statistic to which all proceeding model deviance was compared (-2LL). Notably, the baseline deviance statistic was included in the hypothesis testing feature of HLM8.2 prior to running subsequent math models. Therefore, likelihood ratio tests for deviance from null model was internally calculated for all models to determine goodness-of-fit. ICC equations included (Bickel, 2007):

$$\text{Intercept}_{L3null} / (\text{Intercept}_{L3null} + \text{Intercept}_{L2null} + \text{Residual}_{L1null}) = \text{country-level variability} \quad (6)$$

$$\text{Intercept}_{L2null} / (\text{Intercept}_{L3null} + \text{Intercept}_{L2null} + \text{Residual}_{L1null}) = \text{school-level variability} \quad (7)$$

$$\text{Residual}_{L1null} / (\text{Intercept}_{L3null} + \text{Intercept}_{L2null} + \text{Residual}_{L1null}) = \text{student-level variability} \quad (8)$$

wherein $Intercept_{L3null}$ is the variance in the intercept of the null model attributed to country-level effects, $Intercept_{L2null}$ is the intercept variance attributed to school-level effects, and $Residual_{L1null}$ is the intercept variance attributed to student-level effects.

Next, three-level, fixed-effects models discretely modelled student- (L1), school- (L2), and country-level (L3) predictors to investigate effects of each uncentered, level specific, math or science predictor of student-level self-concept in math and science. Specifically, all math models included L1MSC as the outcome and science models included L1SSC as the outcome. Model 1a specified each student-level (L1) predictor discretely, Model 2a specified all school-level (L2) predictor discretely, and Model 3a specified all country-level (L3) predictors discretely with random intercepts specified at all three levels. Notably, equations for in math were synonymous with those in science (See Table 8 RQ2 Model 1a, 1b, 1c).

For example, equations:

$$L1MSC = G000 + G100*L1ATM + r0 + u00 + e \tag{9}$$

$$L1MSC = G000 + G010*L2ATM + r0 + u00 + e \tag{10}$$

$$L1MSC = G000 + G001*L3ATM + r0 + u00 + e \tag{11}$$

notated (G000) to represent the grand mean or the school and country-level grouping influence on the math self-concept random intercept, (G100*L1ATM) represented the school and country effects on the L1ATM fixed slope, (G010*L2ATM) represented the country effects on the L2ATM fixed slope, and (G001*L3ATM) represented the country effects on the L3ATM fixed slope. Additionally, notation (*e*) represented variability in random intercept attributed to student differences, (*r0*) represents variability in random intercept attributed to school differences, and (*u00)* represents variability in random intercept attributed to country-level differences.

Effect size measures of predictors included -2LL deviance difference to evaluate the significance of additional contributions from each variable to the overall variability in the outcome, while pseudo- R^2 reported the percent of variance in the outcome explained by fixed effects of each level-specific predictor, more precisely the "percent of reduction in errors of prediction" (Bickel, 2007, p. 257). Pseudo R^2 was calculated using equation:

$$R_1^2 = (1\text{-}[RESIDUAL_{FIXED} + INTERCEPT2_{FIXED} + INTERCEPT3_{FIXED})/(RESIDUAL_{NULL} + INTERCEPT2_{NULL} + INTERCEPT3_{NULL}) * 100) \tag{11}$$

where $\text{RESIDUAL}_{\text{FIXED}}$ was the student-level variance, $\text{INTERCEPT2}_{\text{FIXED}}$ was the school-level variance, and $\text{INTERCEPT3}_{\text{FIXED}}$ was the country-level variance of the fixed effects predictor models, while $\text{RESIDUAL}_{\text{NULL}}$ was the student-level variance, $\text{INTERCEPT2}_{\text{NULL}}$ was the school-level variance, $\text{INTERCEPT3}_{\text{NULL}}$ was the country-level variance of the null model with no predictors.

Last, three-level, random effects models specified random intercepts at all three levels and Model 1b permitted L1 predictors to vary between schools and countries, Model 2b permitted L2 predictors to vary across countries, while Model 3b was unnecessary as L3 predictors cannot be modelled as randomly varying (see Table 8 RQ2). Notably models were similar for math and science with variables interchanged accordingly. For example, equations:

$$\text{L1MSC} = \text{G000} + \text{G100*L1ATM} + \text{r0} + \text{r1*L1ATM} + \text{u00} + \text{u10*L1ATM} + \text{e} \tag{12}$$

$$\text{L1MSC} = \text{G000} + \text{G010*L2ATM} + \text{r0} + \text{u00} + \text{u01*L2ATM} + \text{e} \tag{13}$$

modelled notations similar to those of fixed effects model 1a, 2a, 3a with the addition of r1*L1ATM that represented school grouping effect on L1ATM slope and u10*L1ATM represented country-level grouping effect on L1ATM slope, and u01*L2ATM represented country-level grouping effect on L2ATM.

RQ2 Results. Results of the null model in math indicated that most of the variation in L1MSC was attributed to student characteristics (88.85%), while negligible variation was attributed to school (5.67%) or country (5.50%) characteristics. Additionally, both the L2 Intercept ($\beta = 0.537$, *SD* = 0.533, $p < .001$) and L3 Intercept ($\beta = 0.519$, *SD* = 0.720, $p < .001$) estimates were significant confirming that although school- and country-level variation estimates were small, HLM was warranted as L2 and L3 variability did indeed exist.

Results of the null model in science, indicated that most of the variation in L1SSC was attributed to student characteristics (79.57%), while remaining variation was attributed more to country characteristics (12.46%) than to school characteristics (7.98%). Additionally, both the L2 Intercept ($\beta = 0.676$, *SD* = 0.822, $p < .001$) and L3 Intercept ($\beta = 1.056$, *SD* = 1.03, $p < .001$) estimates were significant, confirming that HLM was warranted as school- and country-level effects also existed. The baseline deviance statistic (814270.05, *df* = 4) was included in the hypothesis testing feature of HLM8.2 prior to running subsequent science models to determine goodness-of -fit (see Table 11 Model 1a). Results for *student-level* predictors in math reflected a significant association of L1MSC with L1ATM ($\beta = 0.726$, *SE* = 0.002, $p < .001$; -2LL= 79479.48), L1VOM ($\beta = 0.241$, *SE* = 0.010, $p < .001$; -2LL = 19073.76, L1SES ($\beta = 0.230$, *SE* = 0.020, $p < .001$; -2LL = 3307.79), L1GND ($\beta = -0.427$, *SE* = 0.093, $p < .001$; -2LL = 744.15), and L1MACH1-5 ($\beta = 0.017$, *SE* = 0.001, $p < .001$; -2LL= 39243.69) (see Table 11 Model 1a). Overall, L1ATM (37%) and L1VOM (12.95%) accounted for the most variation in L1MSC, while L1SES (0.41%), and L1MACH (-4.19%) made negligible contributions. Additionally, all significantly associated student-level predictors showed significant school-to-school and country-to country variability (see Table 11 model 1b).

Table 11

TIMSS 2019 Student-Level Predictors of Students' Math Self-Concept

	Model 1a L1 Fixed				Model 1b L1 Random		
FIXED EFFECTS	β	*SE*	*-2LL*	*% Var Explained*	β	*SE*	*-2LL*
L1Null ICC			849877.02	88.85%			
L1ATM	0.726*	0.002*	79479.48*	37.00%	0.735*	0.031*	82597.16 *
L1VOM	0.241*	0.010*	19073.76*	12.95%	0.249*	0.012*	20112.5*
L1SES	0.230*	0.020*	3307.79*	0.41%	0.247*	0.020*	3788.75*
L1GND	-0.427*	0.093*	744.15*	0.29%	-0.421*	0.113 *	1494.56*
L1MACH1-5	0.017*	0.001*	39243.69*	-4.19%	0.012*	0.001*	42855.75*
RANDOM EFFECTS	β	*SD*			β	*SD*	
L1 Res Var Null	8.405*	2.899*					
L1 Res Var L1ATM	5.300*	2.302*			5.135*	2.266*	
L1 Res Var L1VOM	7.549*	2.748*			7.442*	2.728*	
L1 Res Var L1SES	8.265*	2.875*			8.187*	2.861*	
L1 Res Var L1GND	8.361*	2.892*			8.255*	2.873*	
L1 Res Var L1MACH	6.577*	2.564*			6.33*	2.52*	
L2 Res Var Null	0.537*	0.733*					
L2 Res Var L1ATM	0.241*	0.490*			1.024*	1.012*	
L2 Res Var L1VOM	0.381*	0.617*			1.161*	1.078*	
L2 Res Var L1SES	0.462*	0.680*			1.672*	1.293*	
L2 Res Var L1GND	0.555*	0.745*			0.595*	0.771*	
L2 Res Var L1MACH	0.759*	0.871*			7.148*	2.673*	
L2 Res Var L1ATM slope					0.012*	0.109*	
L2 Res Var L1VOM slope					0.004*	0.065*	
L2 Res Var L1SES slope					0.020*	0.143*	
L2 Res Var L1GND slope					0.408*	0.638*	
L2 Res Var L1MACH slope					0.00002*	0.005*	
L3 Res Var Null	0.519*	0.720*					
L3 Res Var L1ATM	0.41986*	0.648*			1.414*	1.189*	
L3 Res Var L1VOM	0.305*	0.553*			0.826*	0.909*	
L3 Res Var L1SES	0.694*	0.833*			1.869*	1.367*	
L3 Res Var L1GND	0.517*	0.719*			0.335*	0.578*	
L3 Res Var L1MACH	2.52*	*1.587			10.247*	3.201*	
L3 Res Var L1ATM slope					0.013*	0.116*	
L3 Res Var L1VOM slope					0.002*	0.047*	
L3 Res Var L1SES slope					0.007*	0.081*	
L3 Res Var L1GND slope					0.175*	0.419*	
L3 Res Var L1MACH slope					2.045*	0.643	

Notes. NS = not a significant predictor of L1SSC. * $p < .001$, ** $p<.05$.

Table 12

TIMSS 2019 Student-Level Predictors of Students' Self-Concept in Science

	Model 1a Fixed				Model 1b Random		
Fixed Effects	β	*SE*	*-2LL*	*% Var*	β	*SE*	*-2LL*
L1NULL ICC			814270.05	79.57%			
L1ATS	0.734*	0.014*	79678.01*	41.60%	0.751*	0.012*	80927.98*
L1VOS	0.255*	0.008*	31363.86*	21.50%	0.260*	0.008*	32171.82*
L1SES	0.229*	0.025*	75730.97*	-1.50%	0.245*	0.020*	4454.78*
L1GND	-0.227	0.13	79423.83	0.10%	NS	NS	NS
L1SACH	0.012*	0.001*	23778.03*	-3.50%	0.013*	0.001*	165293.85*
Random Effects	β	*SD*			β	*SD*	
L1 Intercept L1ATS	4.273*	2.067*			4.180*	2.045*	
L1 Intercept L1VOS	5.650*	2.377*			5.567*	2.359*	
L1 Intercept L1SES	6.604*	2.570*			6.546*	2.558*	
L1 Intercept L1GND	6.73	2.59			NS	NS	
L1 Intercept L1SACH	5.849*	2.419*			5.113**	3.080**	
L2 Intercept L1ATS	0.241*	0.491*			1.336*	1.156*	
L2 Intercept L1VOS	0.409*	0.640*			0.970*	0.985*	
L2 Intercept L1SES	0.601*	0.776*			1.442*	1.201*	
L2 Intercept L1GND	0.69	0.831					
L2 Intercept L1SACH	0.771*	0.595*			0.910**	0.951**	
L2 Res Var L1ATS slope					0.014*	0.119*	
L2 Res Var L1VOS slope					0.003*	0.056*	
L2 Res Var L1SES slope					0.014*	0.117*	
L2 Res Var L1GND slope					NS	NS	
L2 Res Var L1SACH slope					0	0.002	
L3 Intercept NULL	1.056*	1.0278*					
L3 Intercept L1ATS	0.433*	0.658*			0.913*	0.956*	
L3 Intercept L1VOS	0.601*	0.775*			0.954*	0.977*	
L3 Intercept L1SES	1.399*	1.183*			2.622*	1.619*	
L3 Intercept L1GND	1.053	1.03			NS	NS	
L3 Intercept L1SACH	2.152*	1.467*			6.418*	2.533*	
L3 Res Var L1ATS slope					0.003*	0.059*	
L3 Res Var L1VOS slope					0.001*	0.030*	
L3 Res Var L1SES slope					0.005*	0.074*	
L3 Res Var L1GND slope					NS	NS	
L3 Res Var L1SACH slope					0.001*	0.002*	
L3 Res Var Int L1GND							
L3 Res Var Int L1SACH					0.001*	0.002*	

Notes. NS = not a significant predictor of L1SSC. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

Results for *student-level* predictors in science reflected a significant association of L1SSC with L1ATS (β = 0.734, *SE* = 0.014, p < .001; -2LL= 79678.01), L1VOS (β = 0.255, *SE* = 0.008, p < .001; -2LL = 31363.86), L1SES (β = 0.229, *SE* = 0.025, p < .001;

-2LL = 75730.97), L1GND (β = -0.227, *SE* = 0.13, p > .001; -2LL = 75730.97 and L1SACH1-5 (β = 0.012, *SE* = 0.001, p < .001; -2LL = 23778.03). L1GND was not significantly associated (β = -0.427, *SE* = 0.093, p > .001; -2LL = 79423.83) (see Table 12 Model 1a). Similar to math results, L1ATS (41.60%) and L1VOS (21.50%) contributed the most to variation in L1SSC, while L1SES (-1.5%), L1GND (0.10%), and L1SACH (-3.50%) made negligible contributions. Additionally, all significantly associated student-level variables in science showed significant school-to-school and country-to country variability (see Table 12 model 1b in science).

Results for *school-level* predictors in math reflected a significant association of L1MSC with L2ATM (β = 0.487, *SE* = 0.118, p < .001, -2LL = 105.16), L2CLM (β = -0.034, *SE* = 0.006, p < .001; -2LL = 301.67), L2SES (β = 0.226, *SE* = 0.056, p < .001; -2LL = 183.46), L2MSC (β = 0.669, *SE* = 0.072, p < .001; -2LL = 188.32), L2VOM (β = 0.241, *SE* = 0.010, p < .001; 2LL = 34.24), and L2MACH1-5 (β = 0.005, *SE* = 0.002, p < .05; -2LL = 75.60). L2LOC was not significantly associated (β = 0.005, *SE* = 0.048, p < .05; -2LL = 0.035). Generally, variance explained by each predictor is negligible. Whereas L2MSC (0.44%), L2CLM (0.43%), L2ATM (0.22%) contributed the most to variation in L1MSC, L2VOM (0.14%), L2SES (-0.40%), L2MACH1-5 (-0.38%) and L2LOC (-0.01%) made an even smaller contribution (see Table 13 Model 2a). Additionally, all significantly associated school-level variables in math showed significant school-to-school and country-to-country variability (see Table 13 Model 2b).

Table 13

TIMSS 2019 School-Level Predictors of Students' Math Self-Concept

	Model 2a				Model 2b		
Fixed Effects	β	*SE*	*-2 LL*	*% Var Explained*	β	*SE*	*-2 LL*
L2NULL ICC				5.67%			
L2ATM	0.487*	0.118*	105.16*	0.22%	0.498*	0.120*	134.93*
L2CLM	-0.034*	0.006*	301.67*	0.43%	-0.033*	0.006*	338.5*
L2LOC	0.005	0.048	0.035	-0.01%			
L2SES	0.226*	0.056*	183.46*	-0.40%	0.206*	0.056*	215.19*
L2MSC	0.669*	0.072*	188.32*	0.44%	0.645*	0.092*	195.58*
L2VOM	0.161*	0.050*	34.24*	0.14%	0.179*	0.075*	44.76*
L2MACH1-5	0.005**	0.002**	75.6*	-0.38%	0.004*	0.002*	77.74*
Random Effects	β	*SD*			β	*SD*	
L1 Res Var L2ATM	8.405*	2.899*			8.405*	2.899*	
L1 Res Var L2CLM	8.403*	2.899*			8.403*	2.899*	
L1 Res Var L2LOC	8.405	2.899					
L1 Res Var L2SES	8.404*	2.899*			8.404*	2.899*	
L1 Res Var L2MSC	8.408*	2.900*			8.409*	2.900*	
L1 Res Var L2VOM	8.405*	2.899*			8.406*	2.899*	
L1 Res Var L2MACH1-5	8.405*	2.9*			8.405*	2.899*	
L2 Res Var L2ATM	0.521*	0.723*			0.515*	0.718*	
L2 Res Var L2CLM	0.498*	0.705*			0.4918*	0.701*	
L2 Res Var L2LOC	0.537	0.733					
L2 Res Var L2SES	0.511*	0.715*			0.505*	0.710*	
L2 Res Var L2MSC	0.500*	0.707*			0.497*	0.705*	
L2 Res Var L2VOM	0.530*	0.728*			0.525*	0.725*	
L2 Res Var L2MACH1-5	0.564*	0.684*			0.524*	0.724*	
L3 Res Var L2ATM	0.513*	0.716*			3.488*	1.868*	
L3 Res Var L2CLM	0.518*	0.720*			1.115*	1.056*	
L3 Res Var L2LOC	0.519	0.721					
L3 Res Var L2SES	0.582*	0.763*			0.400*	0.633*	
L3 Res Var L2MSC	0.510*	0.714*			0.497*	0.705*	
L3 Res Var L2VOM	0.511*	0.715*			1.500*	1.225*	
L3 Res Var L2MACH1-5	0.527*	0.726*			0.389*	0.622*	
L3 Res Var L2ATM slope					0.094*	0.307*	
L3 Res Var L2CLM slope					0.0002*	0.014*	
L3 Res Var L2SES slope					0.016*	0.125*	
L3 Res Var L2MSC slope					0.030*	0.173*	
L3 Res Var L2VOM slope					0.022*	0.149*	
L3 Res Var L2MAC slope					.00000*	0.0002*	

Notes. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

Results for *school-level* predictors in science reflected a significant association of L1SSC with L2ATS (β = 0.559, *SE* = 0.091, p < .001; -2LL = 136.64), L2CLM (β = -0.030, *SE* = 0.006, p < .001; -2LL = 213.96), L2SES (β = 0.177, *SE* = 0.026, p < .001; -2LL =101.75), L2SSC (β = 0.710, *SE* = 0.078, p < .001; -2LL = 239.07), L2VOS (β = 0.263, *SE* = 0.049, p < .001; -2LL = 93.43), and L2SACH1-5 (β = 0.003, *SE* = 0.001, p < .05; -2LL = 29.52). L2LOC was not significantly associated (β = 0.066, *SE* = 0.092, p > .05; -2LL = 209.33). Overall, school-level predictors made similar contributions to the

Table 14

TIMSS 2019 School-Level Predictors of Students' Science Self-Concept

	Model 2a				Model 2b		
Fixed Effects	β	SE	-2 LL	% Var	β	SE	-2 LL
L2NULL ICC			814270.05	8.00%			
L2ATS	0.559*	0.091*	136.64*	12.70%	0.566*	0.110*	144.75*
L2CLM	-0.030*	0.006*	213.96*	10.30%	-0.032*	0.006*	247.37*
L2LOC	0.066	0.092	209.33	9.90%			
L2SES	0.177*	0.026*	101.75*	10.60%	0.167*	0.029*	112.663*
L2SSC	0.710*	0.078*	239.07*	11.10%	0.708*	0.127*	251.318*
L2VOS	0.263*	0.049*	93.43*		0.275*	0.059*	99.282*
L2SACH	0.003**	0.001**	29.52**		0.003**	0.001**	35.23**
Random Effects	β	SD			β	SD	
L1 Res Var Int	6.745*	2.597*			6.745*	2.597*	
L1 Res Var Int	6.744*	2.597*			6.744*	2.597*	
L1 Res Var Int L2SES	6.744*	2.597*			6.744*	2.597*	
L1 Res Var Int	6.744	2.597			NS	NS	
L1 Res Var Int	6.747*	2.597*			6.747*	2.598*	
L1 Res Var Int	6.745*	2.597*			6.745*	2.597*	
L1 Res Var Int	6.744*	2.597*			6.744*	2.597*	
L2 Res Var Int	0.649*	0.805*			0.643*	0.802*	
L2 Res Var Int	0.640*	0.800*			0.631*	0.794*	
L2 Res Var Int L2SES	0.657*	0.811*			0.654*	0.809*	
L2 Res Var Int	2.597	0.822					
L2 Res Var Int	0.627*	0.792*			0.619*	0.787*	
L2 Res Var Int	0.658*	0.811*			0.654*	0.809*	
L2 Res Var Int	0.669*	0.818*			0.728*	0.758*	
L3 Res Var Int	1.020*	1.041*			6.914*	2.629*	
L3 Res Var Int	1.105*	1.051*			1.755*	1.325*	
L3 Res Var Int L2SES	1.058*	1.118*			0.861*	0.928*	
L3 Res Var Int	1.064	1.031					
L3 Res Var Int	1.036*	1.018*			16.367*	4.046*	
L3 Res Var Int	1.045*	1.022*			6.772**	2.602**	
L3 Res Var L2SACH	1.063*	1.062*			0.161**	0.401**	
L3 Res Var L2ATS					0.078*	0.279*	
L3 Res Var L2CLM					0.0003*	0.016*	
L3 Res Var L2SES					0.006	0.0745	
L3 Res Var L2SSC					0.135*	0.367*	
L3 Res Var L2VOS					0.019**	0.139**	
L3 Res Var L2SACH					0.001*	0.000*	

Notes. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

explanation of variation in L1MSC. Specifically, L2ATS (12.70%) contributed the most, while L2SSC (11.10%), L2VOS (10.70%), L2SES (10.60%), L2SACH (10.40%), and L2CLM (10.30%) made comparable contributions as well (see Table 14 Model 2a). Additionally, there was evidence of significant school-to-school and country-to-country variability in all significantly associated school-level predictors in science (see Table 14 model 2b).

Results for *country-level* predictors in math reflected significant association of L1MSC with L3ATM ($\beta = 0.485$, $SE = 0.221$, $p < .05$; -2LL = 6.73), L3MSC ($\beta = 1.008$, $SE = 0.009$, $p < .001$; -2LL = 126.23), L3TRK ($\beta = -0.829$, $SE = 0.369$, $p < .05$; -2LL = 9.94), L3VOM ($\beta = 0.506$, $SE = 0.144$, $p < .05$; -2LL = 20.70), and LMACH1-5 ($\beta =0.00$, $SE = 0.00$, $p < .05$; -2LL = 19.24) (see Table 13 Model 2a.3). L3IPC ($\beta = -0.000003$, $SE = 0.00001$, $p > .05$; -2LL = 0.143) and L3IDV ($\beta = 0.0030$, $SE = 0.006$, $p > .05$; -2LL = 0.303) were not significantly associated. Additionally, L3MSC contributed the most (5.50%), followed by L3VOM (2.90%) and L3MACH (2.90%), with L3TRK (1.80%) and L3ATM (1.30%), contributing the least. L3IDV (0.01%) and L3IPC (0.00%) made more negligible contributions (see Table 15 Model 3a).

Finally, results for *country-level* predictors in science reflected a significant association of L3SSC with L3ATS ($\beta = 0.983$, $SE = 0.183$, $p < .001$; -2LL = 24.96), L3IPC ($\beta = -0.00003$, $SE = 0.00001$, $p < .05$; -2LL = 10.13), L3SSC ($\beta = 1.027$, $SE = 0.010$, $p < .001$; -2LL = 151.00), L3VOS ($\beta = 0.537$, $SE = 0.136$, $p < .001$; -2LL = 21.46), and L3SACH ($\beta = -0.012$, $SE = 0.004$, $p < .05$; -2LL = 11.92). L3IDV ($\beta = 0.0005$, $SE = 0.008$, $p > .05$; -2LL = 0.01) and L3TRK ($\beta = 0.589$, $SE = 0.314$, $p > .05$; -2LL = 0.19) were not significantly associated. However, L3SACH (22.10%) and L3SSC (21.50%)

contributed most to the explanation of variation in L3SSC, while L3ATS (17.30%) and L3VOS (16.70%) contributed comparably with L3IPC (11.10%), L3IDV (10.40%), and TRK (see Table 16 Model 3a).

Table 15

TIMSS 2019 Country-Level Predictors of Students' Math Self-Concept

	Model 3a Fixed			
Fixed Effects	**β**	***SE***	***-2 LL***	***% Var Explained***
L3 Null ICC			849877.019	5.5%
L3ATM	0.485**	0.221**	6.733*	1.3%
L3IDV	0.0030	0.006	0.303	0.1%
L3IPC	-0.000003	0.00001	0.143	0.0%
L3MSC	1.008*	0.009*	126.227*	5.5%
L3TRK	-0.829**	0.369**	9.937*	1.8%
L3VOM	0.506**	0.144**	20.696*	2.9%
L3MACH1-5	-0.009**	0.003**	19.235*	2.9%
Random Effects	β	*SD*		
L1 Res Var Int NULL	8.405*	2.899*		
L1 Res Var Int L3ATM	8.405*	2.899*		
L1 Res Var Int L3IDV	8.405*	2.900*		
L1 Res Var Int L3IPC	8.405*	2.900*		
L1 Res Var Int L3MSC	8.405*	2.899*		
L1 Res Var Int L3TRK	8.405*	2.899*		
L1 Res Var Int L3VOM	8.405*	2.899*		
L1 Res Var Int L3MACH1-	8.405*	2.899*		
L2 Res Var Int NULL	0.537*	0.733*		
L2 Res Var Int L3ATM	0.537*	0.733*		
L2 Res Var Int L3IDV	0.537*	0.733*		
L2 Res Var Int L3IPC	0.537*	0.733*		
L2 Res Var Int L3MSC	0.534*	0.731*		
L2 Res Var Int L3TRK	0.537*	0.733*		
L2 Res Var Int L3VOM	0.537*	0.733*		
L2 Res Var Int L3MACH1-	0.537*	0.733*		
L3 Res Var Int Null	0.519*	0.720*		
L3 Res Var Int L3ATM	0.400*	0.632*		
L3 Res Var Int L3IDV	0.513*	0.716*		
L3 Res Var Int L3IPC	0.516*	0.516*		
L3 Res Var Int L3MSC	0.0007*	0.027*		
L3 Res Var Int L3TRK	0.353*	0.594*		
L3 Res Var Int L3VOM	0.245*	0.495*		
L3 Res Var Int L3MACH1-	0.246*	0.496*		

Notes. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

Table 16

TIMSS 2019 Country-Level Predictors of Students' Science Self-Concept

	Model 3a			
Fixed Effects	β	SE	-2 LL	% Var Explained
L3NULL ICC			#########	12.13%
L3ATS	0.983*	0.183*	24.96**	17.30%
L3IDV	0.0005	0.008	0.01	
L3IPC	-0.00003**	0.00001**	10.13**	11.10%
L3SSC	1.027*	0.010*	151.00*	21.50%
L3TRK	0.589	0.314	0.19	
L3VOS	0.537*	0.136*	21.46*	16.70%
L3SACH	-0.012**	0.004**	11.92*	14.50%
Random Effects	β	SD		
L1 Res Var Int	6.744*	2.597*		
L1 Res Var Int	6.744*	2.597*		
L1 Res Var Int	6.744*	2.597*		
L1 Res Var Int	6.749*	2.597*		
L1 Res Var Int	6.744*	2.597*		
L1 Res Var Int	6.744*	2.597*		
L1 Res Var Int	6.744*	2.597*		
L2 Res Var Int	0.677*	0.823*		
L2 Res Var Int	0.676*	0.822*		
L2 Res Var Int	0.823*	0.677*		
L2 Res Var Int	0.674*	0.821*		
L2 Res Var Int	0.676*	0.822*		
L2 Res Var Int	0.677*	0.823*		
L2 Res Var Int	0.677*	0..823*		
L3 Res Var Int	0.402*	0.634*		
L3 Res Var Int	1.056*	1.028*		
L3 Res Var Int	0.845*	0.714*		
L3 Res Var Int	0.0002	0.014		
L3 Res Var Int	1.049*	1.024*		
L3 Res Var Int	0.460*	0.678*		
L3 Res Var Int	0.667*	0.816*		

Notes. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

Research Question 3 (RQ3)

Is school- or country-level BFLPE moderated by student-, school, or country-level variables found to be significantly associated with student-level self-concept in math and science across 26 TIMSS 2019 countries (Seaton, 2010)? To answer RQ3, three-level, random coefficient models were examined in HLM 8.2 to determine moderation effects of interactions made from significantly associated predictors of students' self-concept in math and science (see RQ2 results) with L2BFLPE and L3BFLPE predictors in math and science (see RQ1 results). Notably, it is recommended to conduct stepwise testing of hierarchical linear models to first confirm fixed effects then random effects prior to including interactions for moderation analyses (Bickel, 2007; Heck et al., 2014; Raudenbush et al., 2019; Tabachnick & Fidell, 2013).

Uniquely, this study divided the stepwise process among three research questions. To be precise, RQ1 results first confirmed the significance of fixed and random effects of predictors and covariates of L2BFLPE and L3BFLPE then RQ2 results confirmed the significance of fixed and random predictors of L1MSC or L1SSC. Thereafter, RQ3 examined random coefficient, moderation analyses in HLM8.2. Notably, all models were run five times, once for each plausible value, then all five estimates were manually transferred and averaged in Excel for report results.

Specifically, to examine moderation effects of significant student-level (L1) predictors of L2BFLPE in math or science, Model 1a specified L1MSC or L1SSC as the outcome, L1MACH1-5 or L1SACH1-5 as BFLPE covariates, L2MACH1-5 or L2SACH1-5 as BFLPE predictors, significant L1 predictors (student-level) of L1MSC or L1SSC as moderators, as well as cross-level interactions between L2MACH1-5 x L1 predictors or L2SACH1-5 x L1 predictors as moderation effects, with all intercepts, covariates, and predictors specified as random at the school- and country-levels.

Next, to examine moderation effects of significant school-level (L2) predictors as moderators of L2BFLPE in math and science, Model 2a specified L1MSC or L1SSC as the outcome, L1MACH1-5 or L1SACH1-5 as a BFLPE covariates, L2MACH1-5 or L2SACH1-5 as BFLPE predictors, significant L2 predictors of L1MSC or L1SSC as moderators, as well as same-level interactions between L2MACH1-5 x moderators in math or L2SACH1-5 x L2 moderators in science as moderation effects, with all intercepts, covariates, and predictors specified as random at the school- and country-levels.

Last, to examine moderation effects of significant country-level (L3) predictors as moderators of L2BFLPE in math and science, Model 3a specified L1MSC or L1SSC as the outcome, L1MACH1-5 or L1SACH1-5 as a BFLPE covariates, L2MACH1-5 or L2SACH1-5 as a BFLPE predictors, significant country-level (L3) predictors of L1MSC and L1SSC as moderators, as well as cross-level interactions between L2MACH1-5 x L3 moderators in math or L2SACH1-5 x L3 moderators in science as moderation effects, with only intercepts and covariates specified as random at the school- and country-levels as L3 predictors cannot be specified as random in a three-level model.

For instance, equation:

L1MSC = G000 + G010*L2MACH1 + G100*L1MACH1 + G200*L1ATM + G210*L1ATM*L2MACH1 + r0 + r1*L1MACH1+ r2*L1ATM+ u00 + u01*L2MACH1l + u10*L1MACH1 + u20*L1ATM + u21*L1ATM*L2MACH1 + e (14)

Notations, represented the common value of the random L1MSC intercept for all students across all schools and countries, also known as the grand mean of L1MSC (G000).

Additionally, the model specified the country effect on the L2MACH random slope ($G010*L2MACH1$), as well as the school and country effect on the L1MACH random slope ($G100*L1MACH1$), L1ATM moderator ($G200*L1ATM$), and same-level interaction of L1ATM x L2MACH ($G210*L1ATM*L2MACH1$). Also, the country-level effect on random intercept ($u00$), the L1MACH slope ($u10*L1MACH1$), L1ATM slope ($u20*L1ATM$), and L2MACH slope ($u01*L2MACH1$). Moreover, notations represented the school-level grouping effect on random intercept ($r0$), L1MACH random slope ($r1*L1MACH1$) and L1ATM random slope ($r2*L1ATM$), while student-level effects on random intercept were notated as well (e) (see Table 8 RQ3).

Furthermore, analyses for L3BFLPE moderation effects in math and science followed similar specifications as L2BFLPE models. Specifically, Model 1b specified L3BFLPE covariate and predictor, L1 moderators, and cross-level interactions of L1 moderators x L3MACH1-5 or L3SACH1-5. Model 2b specified L3BFLPE covariate and predictor, L2 moderators, and cross-level interactions of L3MACH x L2 moderators in math or L3SACH x L2 moderators in science and Model 3b specified L3BFLPE covariate and predictor, L3 moderators, and same-level interactions of L3 math moderators x L3MACH1-5 or L3 science moderators x L3SACH1-5 with L1SSC intercept specified as random across schools and countries. For example, equation:

$$L1SSC = G000 + G001*L3SACH1 + G100*L1SACH1 + G200*L1ATS + G201*L1ATS*L3SACH1 + r0 + r1*L1SACH1 + r2*L1ATS + u00 + u10*L1SACH1 + u20*L1ATS + e \tag{15}$$

for L3BFLPE random coefficient, moderation effects presented similar notations as those for L2BFLPE equation, wherein ($G000$) represented the collective school and country grouping effects on L1SSC random intercept or grand mean, ($G001*L3SACH$) represented effects of L3SACH country-level predictor, ($_{G201*L1ATS*L3SACH}$) represented effects of L1ATSxL3SACH cross-level interaction, ($_{G100*L1SACH}$) represented school and country grouping effects on L1SACH slope and L1ATS random slope ($_{G200*L1ATS}$) with country grouping effects on L1SSC random intercept ($_{u00}$), L1SACH random slope ($_{u10*L1SACH}$) and L1ATS random slope ($_{u20*L1ATS}$) specified as well.

Additionally, school effects on L1SSC random intercept ($_{r0}$), L1SACH random slope ($_{r1*L1SACH}$), and L1ATM random slope ($_{r2*L1ATS}$) were represented as was ($_{e}$) student-level effect on random intercept was included as well. Notably, for the sake of brevity, only fixed and random estimates of significant moderation effects of L2BFLPE in math and science (see Table 17 and Table 18) and significant moderation effects of L3BFLPE in math and science (see Table 21 and Table 22) were reported. However, numerical results of non-significant moderation analyses in math and science for L2BFLPE are available in Appendix E as well as Appendix F for L3BFLPE.

Next, effect sizes of significant moderation effects were calculated (Δ). Uniquely, Tymms (2004) described effect sizes for continuous, unstandardized predictors in multilevel models as the "distance between the two residuals of the dependent variable that correspond to the observed scores of the predictor, located one standard deviation above (+1 SD) and below (-1 SD) the predictor's mean divided by the standard deviation of the student-level residual variance" (p.62). Explicitly, effect size was calculated by Tymms (2004) equation:

$$\Delta = 2 * \beta_{\text{interaction}} * SD_{\text{moderator}} / \sigma_e \qquad (16)$$

wherein 2 represents the distance between the residuals of the outcome that corresponds to one SD above and one standard deviation below the mean of moderator, $\beta_{interaction}$ represents the unstandardized regression coefficient of the interaction effect at each corresponding level, $SD_{moderator}$ is the standard deviation of the moderator in the interaction (p. 62), and σ^e is the SD at student-level or L1 intercept of the model (Seaton et al., 2010b; Trautwein et al., 2008; Tymms, 2004). Effect sizes were comparable to Cohen's *d* whereby 0.20 was considered small, 0.50 was considered medium, and 0.80 was considered large. Furthermore, the magnitude of change in BFLPE attributed to significant moderation effects in math (see Table 20) and science (see table 21) was calculated with equation:

$$\beta\ \mathrm{DIFF} = \beta_{interaction} - \beta_{RQ1} \tag{17}$$

wherein $\beta_{interaction}$ is the unstandardized coefficient of the interaction effect at each corresponding level and β_{RQ1} is the baseline unstandardized coefficient of either L2BFLPE in math (β_{L2MACH} = -0.007, *df* = 25, p < 0.001, see Table 9), L2BFLPE in science (β_{L2SACH} = -0.004, *df* =25, p < 0.001, see Table 10), L3BFLPE in math (β_{L3MACH} = -0.026, *df* =24, p < 0.001, see Table 9), or L3BFLPE in science (β_{L3SACH} = -0.024, *df* = 24, p < 0.001, see Table 10). Finally, the corresponding statistical significance of β_{DIFF} was calculated using a 2-tailed T-test (see Table 21 and Table 22) to compare baseline unstandardized coefficients of L2BFLPE and L3BFLPE predictors (i.e. see Table 9 for L2MACH and L3MACH baseline coefficients; see Table 10 for L3MACH and L3SACH baseline coefficients) with unstandardized coefficients of L1, L2, and L3 interaction effects (see Table 17, 18, 21, and 22 for results of L2BFLPE and L3 BFLPE moderation

Table 17

Significant L2BFLPE Moderation Effects in Math

	Model 1a L1 INT		Model 2a L2 INT		Model 3a L3 INT	
Fixed Effects	β	*SE*	β	*SE*	β	*SE*
L1VOM moderator	0.182*	0.011*				
L1MACH	0.016*	0.001*				
L2MACH (L2BFLPE)	-0.005*	0.001*				
L1VOM1xL2MACH	0.0005**	0.0002**				
L2SES moderator			1.223*	0.247*		
L1MACH			0.018*	0.001*		
L2MACH (L2BFLPE)			0.001	0.002		
L2SESxL2MACH			-0.003*	0.0005*		
L3MACH moderator					-0.026*	0.003*
L1MACH					0.018*	0.001*
L2MACH (L2BFLPE)					-0.005**	0.002**
L3MACHxL2MACH					0.00003**	0.00001**
-2LL						
L1NULL	849877.02					
L1VOM x L2MACH	56419.47					
L2SES x L2MACH			43170.97			
L3MACH x L2MACH					42959.10	
Random Effects	β	*SD*	β	*SD*	β	*SD*
L1 Res Var Int L1VOM	5.820	2.410				
L2 Res Var Int L1VOM	0.582*	0.763*				
L2 Res Var L1MACH slope	0.00002*	0.004*				
L2 Res Var L1VOM slope	0.005*	0.068*				
L3 Res Var Int L1VOM	1.805*	1.343*				
L3 Res Var L1MACH slope	0.00002*	0.004*				
L3 Res Var L1VOM slope	0.002*	0.040*				
L3 Res Var L2MACH slope	0.00000	0.002				
L3 Res Var L1VOMxL2MACH slope	0.00000	0.0003				
L1 Res Var Int L2SES			6.33	2.52		
L2 Res Var Int L2SES			0.674*	0.821*		
L2 Res Var L1MACH slope			0.00002*	0.005*		
L3 Res Var Int L2SES			2.391*	1.546*		
L3 Res Var L1MACH slope			0.00002*	0.005*		
L3 Res Var L2SES slope			0.076	0.272		
L3 Res Var L2MACH slope			0.00000	0.002		
L3 Res Var L2SESxL2MACH slope			0.00000	0.0004		
L1 Res Var Int L3MACH					6.331	2.516
L2 Res Var Int L3MACH					0.769*	0.877*
L2 Res Var L1MACH slope					0.00002*	0.005*
L3 Res Var Int L3MACH					0.280*	0.529*
L3 Res Var L1MACH slope					0.00002*	0.005*
L3 Res Var L2MACH slope					0.00001*	0.004*

Notes. Results displayed for only significant moderation effects of L2BFLPE in math. See Appendix E for results of all L2BFLPE moderation effects in math. Res Var = residual variance, Res Var Int = residual variance intercept. Sample size = 169,810 students; 5,410 schools; 26 countries. *p < 0.001, **p <0.05.

effects in math and science). Precisely, calculations for statistical significance of BFLPE change attributed to the moderator required a 5-step process in Excel (Currell, 2015):

1. β DIFF = $\beta_{\text{interaction}} - \beta_{\text{RQ1}}$
2. SE DIFF = $\sqrt{SE_{RQ3}^2} + SE_{RQ1}^2$
3. T statistic = β DIFF/ SE DIF
4. Degrees of Freedom = ($df_{\text{RQ3}} + df_{\text{RQ1}}$) / 4
5. P-value = [Tdistr2T (T-stat, *df*)]

RQ3 L2BFLPE Math Results. Results for moderation analyses of L2BFLPE in math reflected one statistically significant moderation effect at each level (see Table 17 for significant moderation results, Table 19 for effect size results, and Appendix E for all L2BFLPE moderation results in math). Specifically, L1VOM x L2MACH interaction was statistically significant ($\beta = 0.0005$, $SE = 0.0002$, $p < 0.05$, -2LL = 56419.47) and varied significantly across countries ($\beta_{\text{L3Intercept}} = 1.805$, $SD = 1.343$, $p < 0.001$). Though, the effect size of L1VOMxL2MACH interaction was negligible ($\Delta = 0.00175$) per Cohen's *d* (Seaton et al., 2010), L1VOM significantly decreased the negative effects of L2BFLPE (β DIFF = 0.007, SE DIFF = 0.002, $p < 0.05$). Furthermore, whereas students' self-concept in math decreased as school-averaged math achievement increased (L2BFLPE) for all levels of L1VOM, students that value math the least reflected the lowest levels of overall math self-concept as school-averaged achievement in math increased and a slightly greater rate of decrease (steeper slope) was found for students that valued math more (see Figure G25).

Additionally, L2SES x L2MACH reflected a statistically significant moderation effect ($\beta = -0.003$, $SE = 0.0005$, $p < 0.001$, $-2LL = 43170.97$) and varied significantly

Table 18

Significant L2BFLPE Moderation Effects in Science

	MODEL 1a L1 INT		MODEL 2a L2 INT	
FIXED EFFECTS	β	*SE*	β	*SE*
L1ATS1xL2SACH	0.0007**	0.0003**		
L1ATS moderator	0.678*	0.012*		
L1SACH1 slope (L1ATS)	0.008*	0.0006*		
L2SACH1 slope (L1ATS) L2BFLPE	-0.004**	0.001**		
L1VOS1xL2SACH	0.0001*	0.0005*		
L1VOS moderator	0.218*	0.008*		
L1SACH1 slope (L1VOS)	0.0008*	0.012*		
L2SACH1 slope (L1VOS) L2BFLPE	0.001**	-0.003**		
L2SSC x L2SACH			0.00004**	0.00002**
L2SSC moderator			0.520*	0.076*
L1SACH slope (L2SSC)			0.013*	0.0008*
L2SACH slope (L2SSC) L2BFLPE			-0.005*	0.001*
-2LL				
L1NULL	814270.05			
L1ATS interaction	95649.41			
L1VOS interaction	50739.98			
L2SSC interaction			25514.39	
RANDOM EFFECTS	β	*SD*	β	*SD*
L1 Intercept L1ATS	3.839	1.959		
L2 Intercept L1ATS	0.153*	0.391*		
L2 Res Var L1SACH slope (L1ATS)	0.0000*	0.002*		
L2 Res Var L1ATS slope	0.014*	0.120*		
L3 Intercept L1ATS	0.795*	0.892*		
L3 Res Var L1SACH slope (L1ATS)	0.0000*	0.027*		
L3 Res Var L1ATS slope	0.002*	0.048*		
L3 Res Var L2SACH slope (L1ATS)	0.00001*	0.003*		
L3 Res Var L1ATSxL2SACH slope	0.0000**	0.0008**		
L1 Intercept L1VOS	4.977	2.231		
L2 Intercept L1VOS	0.003*	0.055*		
L2 Res Var L1SACH slope (L1VOS)	0.0000*	0.001*		
L2 Res Var L1VOS slope	0.339*	0.582*		
L3 Intercept L1VOS	1.394*	1.181*		
L3 Res Var L1SACH slope (L1VOS)	0.00001*	0.002*		
L3 Res Var L1VOS slope	0.0007*	0.026*		
L3 Res Var L2MACH slope (L1VOS)	0.00001**	0.003**		
L3 Res Var L1VOSxL2MACH slope	0.00000	0.0002		
L1 Intercept L2SSC			5.79	2.41
L2 Intercept L2SSC			0.524*	0.724*
L2 Res Var L1SACH slope (L2SSC)			0.00000*	0.002*
L3 Intercept L2SSC			2.41*	1.55*
L3 Res Var L1SACH slope (L2SSC)			0.00001*	0.003*
L3 Res Var L2SACH slope (L2SSC)			0.00001**	0.003**
L3 Res Var L2SSC slope			0.035	0.187
L3 Res Var L2SSCxL2SACH slope			0.00000	0.00002

across countries ($\beta_{L3Intercept}$ = 2.391, *SD* = 1.546, $p < 0.001$). Though, the effect size of L2SES xL2MACH interaction was negligible (Δ = -0.003), L2SES significantly decreased the negative effects of L2BFLPE (β DIFF = 0.009, SE DIFF = 0.001, $p < 0.05$). Furthermore, whereas students' self-concept in math decreased as school-averaged math achievement increased (L2BFLPE) at a similar rate of change across all levels of schools' socio-economic status (L2SES), students from more disadvantaged schools reflected overall higher math self-concepts with increased school-averaged math achievement when compared to their more advantaged counterparts (see Figure G26). As well, L3MACH x L2MACH reflected a statistically significant interaction effect (β = 0.00003, SE = 0.00001, $p < 0.05$, -2LL = 42959.10) and varied significantly across countries ($\beta_{L3Intercept}$ = 0.280, *SD* = 0.529, $p < 0.001$). Though the effect size of L3MACH was negligible (Δ = 0.0151), L3MACH significantly decreased the negative effects of L2BFLPE (β DIFF = 0.006, SE DIFF = 0.001, $p < 0.05$). However, whereas students' math self-concept decreased as school-averaged math achievement increased (L2BFLPE) for all levels of L3MACH, students in countries with lower-averaged math reflected overall higher self-concept in math as school-level math achievement increased and a slightly greater rate of decrease (steeper slope) in math self- concept as school-averaged achievement increased was evident for students in higher achieving countries (see Figure G27).

RQ3 L2BFLPE Science Results. Results for moderation analyses of L2BFLPE in science reflected two statistically significant student-level moderation effects (L1ATS and L1VOS), one school-level moderation effect (L2SSC), and no country-level moderation effect (see Table 18 for significant moderation results in science, Table 20 for effect size results in science, and Appendix E for all L2BFLPE moderation results in science). Specifically, L1ATS x L2SACH interaction was statistically significant (β = 0.0007, *SE* = 0.0003, $p < 0.05$, -2LL = 95649.41) and varied significantly across countries ($\beta_{L3Intercept}$ = 0.795, *SD* = 0.892, $p < 0.001$). Yet, the effect size of L1ATSxL2SACH interaction were negligible (Δ = 0.001712) and though L1ATS decreased the negative effects of L2BFLPE (β DIFF = 0.003, *SE* DIFF = 0.002, $p > 0.05$), the change was not statistically significant.

Additionally, whereas students' science self-concept decreased as school-averaged science achievement increased (L2BFLPE) for all levels of L1ATS, students with a less favorable attitude toward science reflected the lowest levels of science self-concept with increased school achievement at a greater rate of change (steeper slope) than those with those with more favorable attitude toward science (See Figure G28). Correspondingly, L1VOS x L2SACH interaction was statistically significant (β = 0.0001, *SE* = .0005, $p < 0.001$, -2LL = 50739.98) and varied significantly across countries ($\beta_{L3Intercept}$ = 1.394, *SD* = 1.181, $p < 0.001$).

However, the effect size of L1VOSxL2SACH interaction was negligible (Δ =0.00411) and though L1VOS decreased the negative effects of L2BFLPE, the change was not statistically significant (β DIFF = 0.004, *SE* DIFF = 0.002, $p > 0.05$). As well, whereas students' science self-concept decreased as school-averaged science achievement increased (L2BFLPE) at all levels of L1VOS, students that valued science less reflected the lowest overall science self-concept as school science achievement increased as well as slightly steeper rate of decline in self-concept when compared to those that valued science most (See Figure G29). Also, L2SSC x L2SACH interaction was statistically significant (β = 0.00004 , *SE* = 0.00002, p < 0.05, -2LL = 25514.39), and varied across countries ($\beta_{L3Intercept}$ = 2.41, *SD* = 1.55, p < 0.001).

Moreover, the effect size of L2SSCxL2SACH interaction was negligible (Δ = 0.0000127) and though L2SSC decreased the negative effects of L2BFLPE, the change was not statistically significant (β DIFF = 0.00404, *SE* DIFF = 0.002, p > 0.05). Thus, students' science self-concept decreased as school-averaged science achievement increased (L2BFLPE) at a similar rate of decline for all levels of L2SSC, though schools with the lowest averaged self-concept in science reflected a slightly lower overall science self-concepts when compared to schools with higher averaged self-concepts as students' science achievement increased (See Figure G30).

RQ3 L3BFLPE Math Results. Results for moderation effects of L3BFLPE in math reflected three statistically significant student-level interactions (L1ATM, L1VOM, and L1MACH), three significant school-level interactions (L2ATM, L2MSC, and L2VOM), and two significant country-level interactions (L3MSC and L3VOM) (see Table 21 for significant moderation results for L3BFLPE in math, Table 19 for moderation effect size results in math, and Appendix F for all L3BFLPE moderation results in math). Specifically, L1ATM x L3MACH interaction was statistically significant (β = -0.005, *SE* = 0.0004, p < 0.001, -2LL = 108289.71) and varied significantly across countries ($\beta_{L3Intercept}$ = 0.386, *SD* = 0.621, p < 0.001.)

Though, the effect size of L1ATMxL3MACH interaction was negligible (Δ = -0.020), L1ATM significantly decreased the negative effects of L3BFLPE (β DIFF = 0.021, *SE* DIFF = 0.005, p < 0.05). Additionally, whereas students' math self-concept decreased at a similar rate of change (slope) for all levels of students' attitudes toward

Table 19

Effect Size of Significant Moderation Effects of L2BFLPE and L3BFLPE in Math

BFLPE Level	RQ3 Interaction Variable	RQ3 Moderator EFFECT SIZE (Δ)	RQ3 Interaction β (SE) (*df*)	RQ1 Baseline β (SE) (*df*)	β DIFF (SE DIFF)	Moderation Effect t Statistic (*df*)
L2BFLPE	L1VOMxL2MACH	0.00175	0.0005 (0.002)** (*25*)	-0.006 (0.001)** (*25*)	0.007 (0.002)	3.5 (*12.5*)**
L2BFLPE	L2SESxL2MACH	-0.003	-0.003 (0.0005)* (25)	-0.006 (0.001)** (*25*)	0.009 (0.001)	9.0 (12.5)**
L2BFLPE	L3MACHxL2MACH	0.00151	0.00003 (0.00001)** (*24)*	-0.006 (0.001)** (*25*)	0.006 (0.001)	6.0 (*12.25*)**
L3BFLPE	L1ATMxL3MACH	-0.020	-0.005 (0.004)* (*24)*	-0.026 (0.003)* (*24)*	0.021 (0.005)	4.2 (*12*)**
L3BFLPE	L1VOMxL3MACH	0.001754	-0.0005 (0.0001)* (*25*)	-0.026 (0.003)* (*24*)	0.026 (0.003)	8.67 (*12.25*)**
L3BFLPE	L1MACHxL3MACH	0.352	0.004 (0.00001)** (*24*)	-0.026 (0.003)* (*24*)	0.030 (0.003)	10.0 (*12*)**
L3BFLPE	L2ATMxL3MACH	-0.034	-0.005 (0.001)* (*24*)	-0.026 (0.003)* (*24*)	0.021 (0.003)	7.0 (*12*)**
L3BFLPE	L2MSCxL3MACH	-0.043	-0.005 (0.002)** (*24*)	-0.026 (0.003)* (*24*)	0.021 (0.004)*	5.25 (*12*)**
L3BFLPE	L2VOMxL3MACH	-0.031	-0.002 (0.001)** (*24*)	-0.026 (0.003)* (*24*)	0.024 (0.003)*	8.0 (*12*)**
L3BFLPE	L3MSCxL3MACH	0.001	0.003 (0.007)* (*22*)	-0.026 (0.003)* (*24*)	0.029 (0.008)*	3.63 (*11.5*)**
L3BFLPE	L3VOMxL3MACH	0.003	0.003 (0.002)** (*22*)	-0.026 (0.003)* (*24*)	0.029 (0.004)*	7.25 (*11.5*)**

Notes. RQ1 Baseline refers to L2BFLPE model 1b and L3BFLPE Model 2b in math (see Table 9). See Equation 14 for (Δ) . See Equation 15 for β DIFF and pg. 38 for significance of change process.

substantial effect *SD* = 0.535, $p < 0.001$). Notably, L1MACH moderator had the only substantial effect size of all else ($\Delta = 0.352$), L1MACH significantly decreased the negative effects of L3BFLPE (β DIFF = 0.03, *SE* DIFF = 0.003, $p < 0.05$). Additionally, whereas students' math self-concept decreased as country-level math achievement increased at a similar rate of change for all levels L1MACH, students with lower math achievement reflected lower levels of math self-concept as country-level achievement in math (see Figure 30). Similarly, L1MACHxL3MACH interaction was statistically significant ($\beta = 0.004$, *SE* = 0.00001, $p < 0.05$, -2LL = 42913.43) and varied significantly across countries ($\beta = 0.286$, *SD* = 0.535, $p < 0.001$).

Concerning school-level moderation effects of L3BFLPE in math, L2ATM x L3MACH interaction was statistically significant ($\beta = -0.005$, *SE* = 0.001, $p < 0.05$, -2LL = 177458.96) and varied significantly across countries ($\beta = 0.300$, *SD* = 0.546, $p < 0.001$). Though, L2ATMxL3MACH interaction had a small effect size ($\Delta = -0.034$), it significantly decreased the negative effects of L3BFLPE (β DIFF = 0.021, *SE* DIFF = 0.003, $p < 0.05$). Also, math self-concept decreased at a similar rate for all levels of L1ATM as country-level math achievement increased (L3BFLPE) for all levels of L2ATM, such that students in schools with a collectively less favorable attitude toward math, reflected lower overall self-concept in math as country-averaged achievement in math increased. Yet, those in schools with collectively more favorable attitudes toward math suffered a slightly greater rate of decline (steeper slope) in math self-concept as country-averaged achievement increased (see Figure G34).

As well, L2MSCxL3MACH interaction was statistically significant ($\beta = -0.005$, *SE* = 0.002, $p < 0.05$, -2LL = 42980.22) and varies across countries ($\beta = 0.284$, *SD* =

0.822, $p < 0.001$). Though, the effect size of L2MSCxL3MACH interaction was negligible ($\Delta = -0.043$), L2MSC significantly decreased the negative effects of L3BFLPE (β DIFF = 0.021, *SE* DIFF = 0.004, $p < 0.05$). Additionally, whereas students' math self-

Table 20

Effect Size of Significant Moderation Effects of L2BFLPE and L3BFLPE in Science

BFLPE Level	Moderation Variable	RQ3 Moderator EFFECT SIZE (Δ)	RQ3 Interaction β (SE) (df)	RQ1 Baseline β (SE) (df)	β DIFF (SE DIFF)	t-Statistic (df)
L2BFLPE	L1ATS x L2SACH	0.001712	0.0007 (0.0003) ** (25)	-0.004 (0.002)** (25)	0.003 (0.002)	1.65 (12.5)
L2BFLPE	L1VOS x L2SACH	0.00411	0.0001 (0.0005)* (25)	-0.004 (0.002)** (25)	0.0041 (0.002)	2.00 (12.5)
L2BFLPE	L2SSC x L2SACH	0.0000127	0.00004 (0.00002)** (25)	-0.004 (0.002)** (25)	0.00404 (0.002)	2.0 (12.5)
L3BFLPE	L1ATS1xL3SACH	0.01223	0.0005 (0.0001)* (24)	-0.022 (0.005)* (25)	0.023 (0.005)	4.5** (12.25)
L3BFLPE	L1SESxL3SACH	0.00067	0.0004 (0.0002)** (24)	-0.022 (0.005)* (25)	0.0224 (0.005)	4.48** (12.25)
L3BFLPE	L1SACH1xL3SACH	0.0283	0.0003 (0.000008)* (24)	-0.022 (0.005)* (25)	0.0223 (0.005)	4.46** (12.25)
L3BFLPE	L2CLMxL3SACH	0.001705	0.0003 (0.00009)* (24)	-0.022 (0.005)* (25)	0.022 (0.005)	4.46** (12.25)
L3BFLPE	L2SESx L3SACH	-0.00179	-0.002 (0.0008)* (24)	-0.022 (0.005)* (25)	0.020 (0.009)	2.22** (12.25)
L3BFLPE	L2VOS x L3SACH	-0.000584	-0.001 (0.0004)** (24)	-0.022 (0.005)* (25)	0.021 (0.005)	4.2** (12.25)
L3BFLPE	L3ATSxL3SACH	0.008055	0.015 (0.006)** (22)	-0.022 (0.005)* (25)	0.037 (0.008)	4.625** (11.75)

Notes. RQ1 Baseline refers to results of RQ1 L2BFLPE Model 1b in science (see Table 10) and RQ3 Moderation refers to results of RQ3 L2BFLPE Model 1a, 1b, 1c in science (see Table 17). See Equation 14 for (Δ). See Equation 15 for β DIFF equation. * p < .001, ** p < .05.

concept decreased as country-level math achievement increased (L3BFLPE) for all levels of L2MSC, students in schools with a collectively lower math self-concept, reflected lower overall self-concept in math as country- average achievement in math increased. Yet, students in schools with a collectively higher self-concept in math suffered a slightly greater rate of decline in math self-concept as country-averaged achievement increased (see Figure G35).

Additionally, L2VOMxL3MACH interaction was statistically significant ($\beta = -0.002$, *SE* = 0.001, $p < 0.05$, -2LL = 42940.24) and varies across countries ($\beta = 0.286$, *SD* = 0.535, $p < 0.001$). Though L2VOM moderation effect size was negligible ($\Delta = -0.031$), L2VOM significantly decreased the negative effects of L3BFLPE (β DIFF = 0.024, *SE* DIFF = 0.003, $p < 0.05$). Moreover, students' self-concept in math decreased as country-level math achievement increased (L3BFLPE) at a similar rate of change for all levels of L2VOM. However, students in schools that collectively valued math less reflected overall lower math self-concepts than students in schools that collectively valued math more (see Figure G36).

Correspondingly, L3MSCxL3MACH interaction was statistically significant ($\beta = 0.003$, *SE* = 0.0007, $p < 0.001$, -2LL = 42984.19) and varied across countries ($\beta = 0.010$, *SD* = 0.102, $p < 0.001$). Though, the effect size of L3MSCxL3MACH interaction was negligible ($\Delta = 0.001$), L3MSC significantly decreased the negative effects of L3BFLPE (β DIFF = 0.029, *SE* DIFF = 0.008, $p < 0.05$). Moreover, whereas students' math self-concept decreased as country-level achievement in math increased (L3BFLPE) at all levels of L3MSC, students in countries with lower averaged self-concepts in math reflected overall lower self-concepts in math as country averaged math achievement increased. Yet, a slightly greater rate of decline in self-concept (steeper slope) was evident for countries with higher averaged self-concepts in math (see Figure G37).

Finally, L3VOMxL3MACH interaction was statistically significant ($\beta = 0.003$, *SE* = 0.002, $p < 0.05$, -2LL = 42915.31) and varied across countries ($\beta = 0.209$,

Table 21

Significant Moderation Effects of L3BFLPE in Math

	Model 1a		Model 2a		Model 3a	
	L1 Int		L2 Int		L3 Int	
Fixed Effects	β	*SE*	β	*SE*	β	*SE*
L1ATM moderator	0.620*	0.022*				
L1VOM moderator	0.188*	0.009*				
L1MACH moderator	0.018	0.001				
L1MACH (L1ATM)	0.011*	0.0005*				
L1MACH (L1VOM)	0.016*	0.001*				
L3MACH (L1ATM)	-0.016*	0.003*				
L3MACH (L1VOM)	-0.022*	0.003*				
L3MACH (L1MACH)	-0.030*	0.003*				
L1ATM1xL3MACH	-0.005*	0.0004*				
L1VOM1xL3MACH	-0.0005*	0.0001*				
L1MACH1xL3MACH	0.004**	0.00001**				
L2ATM moderator			0.263*	0.070*		
L2MSC moderator			0.422**	0.128**		
L2VOM moderator			0.091	0.093		
L1MACH (L2ATM)			0.018*	0.001*		
L1MACH (L2MSC)			0.018*	0.001*		
L1MACH (L2VOM)			0.018*	0.001*		
L3MACH (L2ATM)			-0.026*	0.003*		
L3MACH (L2MSC)			-0.026*	0.003*		
L3MACH (L2VOM)			-0.026*	0.003*		
L2ATMxL3MACH			-0.005*	0.001*		
L2MSCxL3MACH			-0.005**	0.002**		
L2VOMxL3MACH			-0.002**	0.001**		
L1MACH (L3MSC)					0.018*	0.001*
L1MACH (L3VOM)					0.018*	0.001*
L3MSC moderator					-0.552	0.425
L3VOM moderator					-1.501	0.759
L3MSCxL3MACH					0.003*	0.0007*
L3VOMxL3MACH					0.003**	0.002**
-2LL						
L1NULL	849877.02					
L1ATM x L3MACH	108289.71					
L1VOM x L3MACH	56371.53					
L1MACH x L3MACH	42913.43					
L2ATM x L3MACH			177458.96			
L2MSC x L3MACH			42980.22			
L2VOM x L3MACH			42940.24			
L3MSC x L3MACH					42984.19	
L3VOM x L3MACH					42915.31	
Random Effects	β	*SD*	β	*SD*	β	*SD*
L1 Res Var Int L1ATM	4.355	2.087				
L1 Res Var Int L1VOM	5.817	2.412				
L1 Res Var Int L1MACH	6.33	2.516				
L2 Res Var Int L1ATM	0.201*	0.448*				
L2 Res Var Int L1VOM	0.594*	0.771*				
L2 Res Var Int L1MACH	0.783*	0.885*				
L2 Res Var L1MACH slope (L1ATM)	0.00001	0.003				
L2 Res Var L1MACH slope (L1VOM)	0.00002	0.004				
L2 Res Var L1ATM slope	0.012*	0.108*				
L2 Res Var L1VOM slope	0.005*	0.068*				
L2 Res Var L1MACH slope	0.00002*	0.005*				
L3 Res Var Int L1ATM	0.386*	0.621*				
L3 Res Var Int L1VOM	0.286*	0.535*				
L3 Res Var Int L1MACH	0.286*	0.535*				
L3 Res Var L1MACH slope (L1ATM)	0.00001*	0.002*				
L3 Res Var L1MACH slope (L1VOM)	0.00002*	0.004*				
L3 Res Var L1MACH slope (L1MACH)	0.00002*	0.004*				
L3 Res Var L1ATM slope	0.007*	0.080*				
L3 Res Var L1VOM slope	0.001*	0.038*				
L3 Res Var L1MACH slope	0.286*	0.535*				
L1 Res Var Int L2ATM			6.331	2.516		
L1 Res Var Int L2SES			6.33	2.516		
L1 Res Var Int L2MSC			6.330	2.516		
L1 Res Var Int L2VOM			6.329	2.516		
L2 Res Var Int L2ATM			0.765*	0.875*		
L2 Res Var Int L2SES			0.783*	0.885*		
L2 Res Var Int L2MSC			0.685*	0.827*		
L2 Res Var Int L2VOM			0.773*	0.879*		
L2 Res Var L1MACH slope (L2ATM)			0.00002*	0.005*		
L2 Res Var L1MACH slope (L2SES)			0.00002*	0.005*		
L2 Res Var L1MACH slope (L2MSC)			0.00002*	0.005*		
L2 Res Var L1MACH slope (L2VOM)			0.00002*	0.005*		
L3 Res Var Int L2ATM			0.300*	0.546*		
L3 Res Var Int L2SES			0.245*	0.494*		
L3 Res Var Int L2MSC			0.284*	0.822*		
L3 Res Var Int L2VOM			0.286*	0.535*		
L3 Res Var L1MACH slope (L2ATM)			0.0000*	0.005*		
L3 Res Var L1MACH slope (L2SES)			0.00002*	0.005*		
L3 Res Var L1MACH slope (L2MSC)			0.00002*	0.005*		
L3 Res Var L1MACH slope (L2VOM)			0.00002*	0.005*		
L3 Res Var L2ATM slope			0.006	0.072		
L3 Res Var L2SES slope			0.018*	0.135*		
L3 Res Var L2MSC slope			0.061**	0.246**		
L3 Res Var L2VOM slope			0.029**	0.171**		
L1 Res Var Int L3MSC					6.330	2.516
L1 Res Var Int L3VOM					6.330	2.516
L2 Res Var Int L3MSC					0.782*	0.884*
L2 Res Var Int L3VOM					0.784*	0.885*
L2 Res Var L1MACH slope (L3MSC)					0.00002*	0.005*
L2 Res Var L1MACH slope (L3VOM)					0.00002*	0.005*
L3 Res Var Int L3MSC					0.010*	0.102*
L3 Res Var Int L3VOM					0.209*	0.457*
L3 Res Var L1MACH slope (L3MSC)					0.00002*	0.0045*
L3 Res Var L1MACH slope (L3VOM)					0.00002*	0.0045*

Notes. Results displayed for only significant moderation effects of L3BFLPE in math. See Appendix F for results of all L3BFLPE moderation effects in math. Res Var = residual variance, Res Var Int = residual variance intercept. Sample size = 169,810 students; 5,410 schools; 26 countries. *p < 0.001, **p <0.05.

SD = 0.457, $p < 0.001$). Though, the effect size of L3VOM interaction was negligible ($\Delta = 0.003$), L3VOM significantly decreased the negative effects of L3BFLPE (β DIFF = 0.029, *SE* DIFF = 0.004, $p < 0.05$). Moreover, whereas students' math self-concept decreased as country-level achievement in math increased (L3BFLPE) at a similar rate of change for all levels of L3VOM, students in countries that collectively valued math less, reflected the lowest overall self-concepts in math as country averaged math achievement increased (see Figure G38).

RQ3 L3BFLPE Science Results. Results of three-level random coefficient, moderation analyses for L3BFLPE in science revealed three significant student-level moderation effects (L1ATS, L1SES, L1SACH), three significant school-level moderation effects (L2CLM, L2SES, L2VOS), and one significant country-level moderation effect (L3ATS) (see Table 22 for significant moderation results in science, Table 20 for effect size results in science, and Appendix F for all L3BFLPE moderation results in science). Specifically, L1ATS x L3SACH interaction was statistically significant ($\beta = 0.0005$, *SE* = 0.0001, $p < 0.001$, -2LL = 95531.90) and varied significantly across countries ($\beta_{L3Intercept} = 0.441$, *SD* = 0.664, $p < 0.001$).

However, the effect size of L1ATSxL2SACH interaction was negligible ($\Delta = 0.01223$), yet L1ATS significantly decreased the negative effects of L3BFLPE (β DIFF = 0.023, *SE* DIFF = 0.005, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at all levels of L1ATS, students with a less favorable attitude toward science reflected the lowest overall science self-concepts. Yet, a slightly greater rate of decline (steeper slope) in students' math self-concept as country-level math achievement increased was evident for students with the most favorable attitude toward science (See Figure G39).

Likewise, L1SESxL3SACH interaction was statistically significant ($\beta = 0.0004$, *SE* = 0.0002, $p < 0.05$, -2LL = 26150.94) and varied significantly across countries ($\beta_{L3Intercept} = 0.879$, *SD* = 0.938, $p < 0.001$). However, the effect size of L1SESxL2SACH interaction was negligible ($\Delta = 0.00067$). Though, L1SES significantly decreased the negative effects of L3BFLPE (β DIFF = 0.0224, *SE* DIFF = 0.005, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at a similar rate of change for all levels of L1SES, students from more disadvantaged backgrounds reflected the lowest overall science self-concept as country-level science achievement increased (See Figure G40).

As well, L1SACHxL3SACH interaction was statistically significant ($\beta = 0.00003$, *SE* = 0.000008, $p < 0.001$, -2LL = 25323.81) and varied significantly across countries ($\beta_{L3Intercept} = 0.850$, *SD* = 0.922, $p < 0.001$). However, the effect size of L1SACHxL3SACH interaction was small ($\Delta = 0.03$), yet L1SACH significantly decreased the negative effects of L3BFLPE (β DIFF = 0.0223, *SE* DIFF = 0.005, p < 05). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at all levels of L1SACH, students with lower achievement reflected the lowest overall science self-concept as country-level science achievement increased (See Figure G41).

Table 22

Significant Moderation Effects of L3BFLPE in Science

	MODEL 1b L1 INT		MODEL 2b L2 INT		MODEL 3b L3 INT	
FIXED EFFECTS	β	*SE*	β	*SE*	β	*SE*
L1ATSxL3SACH	0.0005*	0.0001*				
L1SESxL3SACH	0.0004**	0.0002**				
L1SACH1xL3SACH	0.00003*	0.000008*				
L1ATS moderator	0.676*	0.011*				
L1SES moderator	0.085*	0.016*				
L1SACH moderator	0.012*	0.0007*				
L1SACH slope (L1ATS)	0.008*	0.0006*				
L1SACH1 slope (L1SES)	0.012*	0.0008*				
L3SACH slope (L1SACH)	-0.025*	0.006*				
L3SACH slope (L1ATS)	-0.012**	0.004**				
L3SACH slope (L1SES)	-0.023*	0.005*				
L3SACH slope (L1SACH)	-0.025*	0.006*				
L2CLMxL3SACH			0.0003*	0.00009*		
L2SESxL3SACH			-0.002*	0.0008*		
L2VOS x L3SACH			-0.001**	0.0004**		
L2CLM moderator			0.013**	0.005**		
L2SES moderator			-0.193*	0.034*		
L2VOS moderator			0.145	0.02		
L1SACH slope (L2CLM)			0.013*	0.0007*		
L1SACH slope (L2SES)			0.013*	0.0007*		
L1SACH slope (L2VOS)			0.013*	0.0008*		
L3SACH slope (L2CLM)			-0.022*	0.005*		
L3SACH slope (L2SES)			-0.022*	0.005*		
L3SACH slope (L2VOS)			-0.022*	0.005*		
L3ATSxL3SACH					0.015**	0.006**
L3ATS moderator					-6.714**	3.031**
L1SACH slope (L3ATS)					0.0127*	0.0008*
L3SACH slope (L3ATS)					-0.154**	0.057**
-2LL						
L1NULL	814270.05					
L1ATS interaction	95531.90					
L1SES interaction	26150.94					
L1SACH interaction	25323.81					
L2CLM interaction			25478.43			
L2SES interaction			25521.62			
L2VOS interaction			25364.23			
L3ATS interaction					25333.39	
RANDOM EFFECTS	β	*SD*	β	*SD*	β	*SD*
L1 Intercept L1ATS	3.837	1.959				
L1 Intercept L1SES	5.728	2.393				
L1 Intercept L1SACH	5.791	2.406				
L2 Intercept L1ATS	0.156*	0.395*				
L2 Intercept L1SES	0.566*	0.753*				
L2 Intercept L1SACH	0.565*	0.752*				
L2 Res Var L1SACH slope	0.00000*	0.002*				
L2 Res Var L1SACH slope	0.00000*	0.002*				
L2 Res Var L1SACH slope	0.00000*	0.002*				
L2 Res Var L1ATS slope	0.015*	0.121*				
L2 Res Var L1SES slope	0.012*	0.111*				
L3 Intercept L1ATS	0.441*	0.664*				
L3 Intercept L1SES	0.879*	0.938*				
L3 Intercept L1SACH	0.850*	0.922*				
L3 Res Var L1ATS slope	0.002*	0.040*				
L3 Res Var L1SES slope	0.002*	0.042*				
L3 Res Var L1SACH slope	0.00000	0.002				
L3 Res Var L1SACH slope	0.00001	0.003				
L3 Res Var L1SACH slope	0.850	0.922				
L1 Intercept L2CLM			5.79	2.41		
L1 Intercept L2SES			5.79	2.41		
L1 Intercept L2VOS			5.79	2.41		
L2 Intercept L2CLM			0.547*	0.739*		
L2 Intercept L2SES			0.544*	0.738*		
L2 Intercept L2VOS			0.556*	0.745*		
L2 Res Var L1SACH slope			0.00000*	0.002*		
L2 Res Var L1SACH slope			0.00000*	0.002*		
L2 Res Var L1SACH slope			0.00000*	0.002*		
L3 Intercept L2CLM			0.848*	0.921*		
L3 Intercept L2SES			0.817*	0.904*		
L3 Intercept L2VOS			0.870*	0.933*		
L3 Res Var L2CLM slope			0.0002*	0.015*		
L3 Res Var L2SES slope			0.015*	0.120*		
L3 Res Var L2VOS slope			0.0003	0.016		
L3 Res Var L1SACH slope			0.00001*	0.003*		
L3 Res Var L1SACH slope			0.00001*	0.003*		
L3 Res Var L1SACH slope			0.00001*	0.003*		
L1 Intercept L3ATS					5.790	2.406
L2 Intercept L3ATS					0.565*	0.752*
L2 Res Var L1SACH slope					0.00000*	0.002*
L3 Intercept L3ATS					0.351*	0.592*
L3 Res Var L1SACH slope					0.00001*	0.003*

Notes . Results displayed for significant moderation effects of L3BFLPE in science. See Appendix F for results of all L3BFLPE moderation effects in science. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p<.05.

Similarly, L2CLMxL3SACH interaction was statistically significant ($\beta = 0.0003$, *SE* = 0.00009, $p < 0.001$, -2LL = 25478.43) and varied significantly across countries $_{3Intercept}$ = 0.848, *SD* = 0.921, $p < 0.001$). Though the effect size of L2CLMxL3SACH interaction was negligible ($\Delta = 0.0017$), L2CLM significantly decreased the negative effects of L3BFLPE (β DIFF = 0.0223, *SE* DIFF = 0.005, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at a similar rate of change for all levels of L2CLM, students with lower climate levels reflected the lowest overall science self-concept as country-level science achievement increased (See Figure G42).

Furthermore, L2SESxL3SACH interaction was statistically significant ($\beta = -0.002$, *SE* = 0.0008, $p < 0.001$, -2LL = 25521.62) and varied significantly across countries ($\beta_{L3Intercept} = 0.817$, *SD* = 0.904, $p < 0.001$). However, the effect size of L2SESxL3SACH interaction was negligible ($\Delta = -0.00179$), yet L2SES significantly decreased the negative effects of L3BFLPE (β DIFF = 0.020, *SE* DIFF = 0.009, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at a similar rate for all levels of L2SES, students in more disadvantaged schools showed the highest overall science self-concept as country-level science achievement increased when compared to their more advantaged counterparts (See Figure G43).

Correspondingly, L2VOSxL3SACH interaction was statistically significant ($\beta = -0.001$, *SE* = 0.0004, $p < 0.05$, -2LL = 25364.23) and varied significantly across countries ($\beta_{L3Intercept} = 0.870$, *SD* = 0.933, $p < 0.001$). Though, the effect size of L2VOSxL3SACH interaction was negligible ($\Delta = -0.0006$), L2VOS significantly decreased the negative effects of L3BFLPE (β DIFF = 0.021, *SE* DIFF = 0.005, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at all levels of L2VOS, students in schools that valued math less reflected lower overall science self-concept as country-level science achievement increased. Yet, a slightly greater rate of decline in science self-concept was evident for students in schools that valued science most (see Figure G44).

Finally, L3ATSxL3SACH interaction was statistically significant ($\beta = 0.015$, *SE* = 0.0006, $p < 0.05$, -2LL = 25333.39) and varied significantly across countries ($\beta_{L3Intercept}$ = 0.351, *SD* = 0.529, $p < 0.001$). Though, the effect size of L3ATSxL3SACH interaction was negligible ($\Delta = 0.008$), L3ATS significantly decreased the negative effects of L3BFLPE (β DIFF = 0.037, *SE* DIFF = 0.008, $p < 0.05$). Additionally, whereas students' science self-concept decreased as country-averaged science achievement increased (L3BFLPE) at all levels of L3ATS, students in countries that have the least favorable attitude toward science reflected lower overall science self-concept as country-level science achievement increased. Yet, a slightly greater rate of decline in science self-concept was evident for students in countries that had a more favorable attitude toward science (see Figure G45).

Summary of Results

In review, results for RQ1, indicated that L2BFLPE and L3BFLPE indeed exists at the school-level and country-level in math and science across 26 countries. More specifically, there was a negative relationship between school-averaged achievement as well as country-averaged achievement with student self-concept in math and science while a positive relationship between student achievement and student self- concept in math and science remained. Additionally, results for RQ2 in math showed significant associations between students' math self-concept and students' attitude toward math, valuing of math, socio-economic status, gender, and math achievement, school-averaged attitude toward math, math self-concept, valuing of math, climate and socio-economic status, as well as country-averaged attitude toward math, math self-concept, valuing of math, math achievement, and tracking practices. School location, country-averaged individualism, and country income per capita were not significantly associated with students' math self-concept. Similarly, results for RQ2 in science showed significant associations between students' science self-concept and students' attitude toward science, valuing of science, socio-economic status, and science achievement, school-averaged school- attitude toward science, science self-concept, valuing of science, climate and socio-economic status, as well as country-averaged attitude toward science, science self-concept, valuing of science, science achievement, and income per capita tracking practices. Students' gender, school location, country level of individualism, and country tracking practices were not significantly associated with science self-concept.

Finally, results for RQ3 in math showed that students' valuing of math, schools' socio-economic status, and country-averaged math achievement significantly moderated the negative relationship between school-averaged math achievement and students' math self-concept (L2BFLPE). Also, students' attitude toward math, valuing of math, and math achievement, school-averaged attitude toward math, math-self-concept, and attitude toward math, as well as country-averaged attitude toward math and math-self-concept significantly moderated the negative relationship between country-averaged math achievement and students' math self-concept (L3BFLPE). Additionally, results for RQ3 in science showed that students' attitude toward science and valuing of science, as well as school-averaged science self-concept significantly moderated the negative relationship between school-averaged science achievement and students' science self-concept (L2BFLPE). Similarly, students' attitude toward science, socio-economic status, and science achievement, school-averaged valuing of science, attitude toward science, climate, and socio-economic status significantly moderated the negative relationship between country-averaged science achievement and students' science self-concept (L3BFLPE).

Chapter 5: Discussion

Introduction

Self-concept, defined as one's perception of their own ability, is an important construct across a variety of disciplines as a vital element of well-being and a critical facilitator for achieving one's own greatest human potential (Areepattamannil et al., 2011; Bandura, 1994; Chiu & Klassen, 2010, 2010; Marsh et al., 2015, 2019; Marsh & Craven, 2006, p. 134; Marsh & Hau, 2003; Marsh & Martin, 2011; Moller et al., 2011; Primavera et al., 1974; Rosenberg, 1989; Scheirer & Kraut, 1979; West & Fish, 1973; Wylie, 1979). Noticeably, an abundance of research in educational psychology has reported on the reciprocal and mutually beneficial impact of a positive academic self-concept (ASC) on achievement in math and science (Areepattamannil et al., 2011; Arens et al., 2017; Chiu & Klassen, 2010; Hooper et al., 2013; Jansen et al., 2014, 2015; Lui & Meng, 2010; Marsh, 1986; Marsh & Martin, 2011; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Rogers, 1951; Tucker-Drob et al., 2014; Wilkins, 2004; Zheng et al., 2019).

Accordingly, Herbert Marsh and his colleagues have spent a good part of four decades refining his Big-Fish-Little-Pond Effect (L2BFLPE) that has consistently reported a negative effect of school-averaged achievement (L2BFLPE) on corresponding subject-specific ASC, whereby students' perception of their academic ability declines as school-averaged achievement increases, while a positive effect of students' own achievement on their corresponding ASC simultaneously persists (Marsh, 2019, 2020). (Marsh, 1991; Marsh et al., 2008a, 2008b, 2014, 2015; Marsh & Hau, 2003b; Marsh & Parker, 1984; Nagengast & Marsh, 2012). Additionally, as an extension of L2BFLPE, cross-cultural, paradoxical BFLPE theory has consistently reported a negative effect of country-level achievement on corresponding, subject-specific ASC, whereby students' academic self-concept decreases as country-averaged achievement increases (L3BFLPE), while the positive effect of students' achievement on their corresponding subject-specific ASC simultaneously persists as well (Marsh, 2019, 2020).

Nevertheless, critics of BFLPE theory have contended that a great deal of current BFLPE research has analyzed outdated datasets, applied limited statistical designs, been narrow in scope of subject domains, and analyzed few moderation effects from a limited range of multileveled influences (Dai & Rinn, 2008; Huguet et al., 2009; Marsh, Abduljabbar, et al., 2014; Marsh et al., 2000, 2007; Marsh et al., 2008; Marsh & O'Mara, 2009; Marsh & Parker, 1984; Nagengast & Marsh, 2012; Pekrun et al., 2019; Seaton et al., 2009, 2010; Wang, 2015). Therefore, this study has headed the suggestion of previous research to advance current BFLPE theory and extend research in educational psychology concerning academic self-concept in STEM subjects.

Precisely, this study partially replicated cutting-edge, multilevel methodology of leading BFLPE research studies to analyze comparable results from a similar, but more current, large-scale, international assessment of math and science achievement (Marsh et al., 2019, 2020; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Seaton et al., 2010). Specifically, three main research questions were addressed by applying three-level, hierarchical linear modelling to analyze TIMSS 2019, large-scale, international assessment results of math and science achievement as well as complimentary contextual survey results across a sample of 169,810 eighth grade students within 5,410 schools from 26 countries. Notably, appropriate as it accounted for violations of HLM assumption of independence from the nesting of students within schools and schools within countries, wherein student-level observations are dependent within higher level clusters.

To begin, this study examined the presence of L2BFLPE and L3BFLPE across all countries for math and science. Next, discrete effects of student-, school-, and country-level predictors of ASC in math and science were examined across all countries. Last, significant student, school, and country-level predictors were applied as interactions with corresponding school-averaged math achievement or school-averaged science achievement to determine discrete moderation effects on L2BFLPE. Additionally, the same significant predictors were applied as interactions with corresponding country-averaged math achievement or country-averaged science achievement to determine discrete moderation effects on L3BFLPE (Marsh et al., 2019, 2020b; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Seaton et al., 2010b).

Summary and Context of Findings

Research Question 1

Does L2BFLPE and L3BFLPE Exist Across All Countries in Math and Science? Overall results confirmed that L2BFLPE and L3BFLPE in math and science indeed existed across all 26 countries in this sample. Explicitly, L2BFLPE was present as there was a negative effect of school-averaged math achievement (β = -0.006) on students' MSC and a negative effect of science achievement (β = -0.004) on students' SSC, while a positive effect of student' math achievement (β = 0.018) and science achievement (β = 0.013) on corresponding ASCs persisted. Likewise, L3BFLPE was present as there was a negative effect of country-averaged math achievement (β = -0.017) and a negative effect of science achievement (β = -0.022) on corresponding ASC in math and science, while a similar positive effect of students' math (β = 0.018) and science achievement (β = 0.013) remained. In other words, students' MSC and SSC increased as their corresponding achievement in math or science increased, but their MSC and SSC declined as their school-averaged and country-averaged achievement in the corresponding subject increased. Though this was the first of its kind to examine the negative effects of country-averaged achievement on SSC, referred to here as L3BFLPE in science, these results were consistent with those of Marsh et al., (2019, 2020) partially replicated studies and with results of numerous other BFLPE studies too (Marsh et al., 1995, 2001, 2007, 2015; Marsh & Hau, 2003; Marsh & O'Mara, 2009; Nagengast & Marsh, 2012; Pekrun et al., 2019; Seaton, 2007; Seaton et al., 2009, 2010).

Even though it was beyond the scope of this study to confirm or deny the statistical significance of a rationale, these results compliment BFLPE theory that has persistently attributed the negative effects of BFLPEs to Festinger (1954) social comparison theory (SCT), whereby students' form the perception of their ability based on external comparisons of their achievement scores to those of classmates as well as on external comparisons of their achievement scores to those of national ability reports too (Chiu, 2012; Dai & Rinn, 2008; Huguet et al., 2009; Jonkmann et al., 2012; Marsh et al., 2008, 2015; Marsh et al., 2000, p. 200; Marsh & O'Mara, 2010; McFarland & Buehler, 1995; Nagengast & Marsh, 2012; Plieninger & Dickhäuser, 2015; Schwabe et al., 2019; Seaton et al., 2009, 2010; Wouters et al., 2012). Additionally, these results compliment BFLPE theory that has attributed the positive effects of individual achievements on corresponding ASC to internal comparisons such as those in Marsh (1986) basic I/E model that reports positive effects on students' math and verbal ASC when internal comparisons are made to their achievement in the same domain, but negative when comparisons are made to achievement in the opposite domain (see Figure 5).

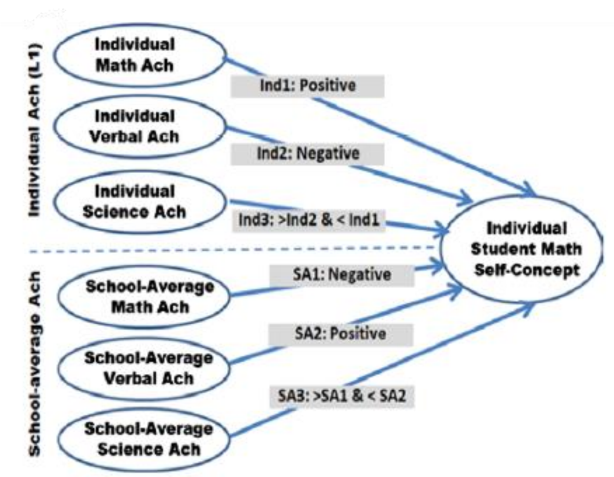
Specifically, like Marsh (2020) the persistently positive effects of students' achievement on their corresponding academic self-concept found here were slightly greater in math (0.018) than in science (0.013). Marsh (2020) accredited this effect to Moller and Marsh (2013) Dimensional Comparison Theory (DCT) that suggested similar internal comparisons as those of the basic I/E model. However, DCT theory extended I/E model such that subject domains were examined along a continuum with math and verbal places farthest apart at polar opposites and similar domains placed closer together (see Figure 6). DCT theory reported that effects on ASC reflect the same continuum as well, whereby internal comparisons with math achievement had a negative effect on verbal ASC and a stronger positive effect on math ASC with lesser positive effects on science ASC as science is closer to math than verbal on the subject domain continuum. Therefore, according to the DCT rationale, "if students perform poorly and doubt their ability in one subject domain, they may otherwise view their ability in other subject's domains more positively owning to changes in the internal ranking (perception) of their ability in different domains" (Marsh et al. 2020, p. 187).

Additionally, complimenting Marsh (2020) results challenge of the Bright Student Hypothesis that suggested high achieving or bright students would be immune to negative effects of BFLPE, post hoc data visualizations similarly showed that L2BFLPE and L3BFLPE indeed negatively affected all ability levels and students such that high achieving schools reflected overall lower self-concepts than students from lower achieving schools. However, consistent with Huguet et al., (2009) alternative post hoc data visualizations also revealed that the ASC's of students in the lowest ability range displayed a very slightly greater negative impact of school- and country averaged BFLPE then higher and averaged ability students in math and science. Ultimately these results support Cheng et al., (2014) notion that there were many inconsistencies concerning the effects of BFLPE on various student abilities.

Moreover, empirically, it was found here that the negative effect of L3BFLPE in math (-0.017) was greater than the negative effect of L2BFLPE in math (-0.006). Uniquely as well, it was found that the negative effects of L3BFLPE in math (-0.017) and science (-0.022) were almost double the size of the negative effect of L2BFLPE in math (-0.006) and science (-0.004) that showed similarly lesser effects. This suggests that the impact of external comparisons with country-averaged achievement negatively impacts ASC more so than comparisons with school-averaged achievement school-averaged comparisons. Though beyond the scope of this study to confirm or deny, these results vaguely resemble those represented by Marsh (2020) BFLPE-CE compensatory effect model. BFLPE-CE model theorized similar effects as those described in DCT model whereby effects on MSC from internal comparisons with one's own achievements were arranged along a subject domain continuum such that comparisons with achievement in a domain near to MSC would be positive and decrease for achievement domains that were farther away (see Figure 8). Thus, the BFLPE-CE model demonstrates that comparisons with math achievement reflected a greater positive effect on MSC than effects of science that were negative or effects of verbal that were even more negative. However, BFLPE-CE was unique in that the model combined DCT and BFLPE by including the L2BFLPE effects of social comparisons from school-averaged achievements on MSC.

Figure 8

Dimensional Comparison Theory Compensatory Effect (BFLPE-CE)



Note. From *"Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: Pan-Human Cross-Cultural Generalizability over 68 Countries" by* Marsh, 2020, European Journal of Personality, 34, p. 180-202.

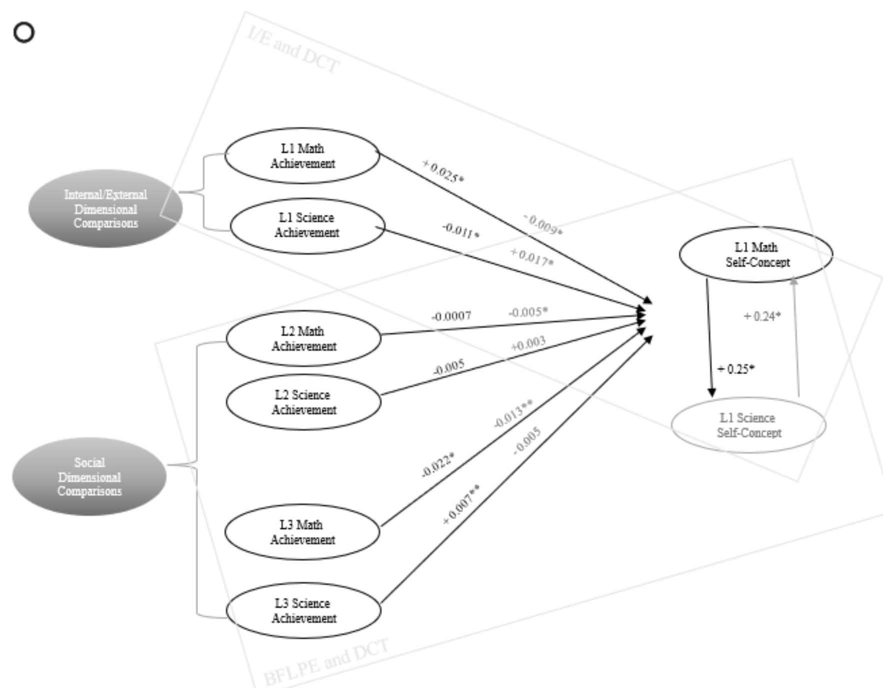
Similarly, results of social comparisons were reflected along a domain continuum with effects opposite those of internal comparisons such that effects of comparisons to school- averaged achievement on corresponding ASC in near subject domains reflected greatest negative results while comparisons to school-averaged achievement in farther subject domains reflected lesser negative or positive effects. However, Marsh (2020) BFLPE-CE model only examined effects of individual and school-averaged math, science, and verbal achievements on MSC, the model did not examine those effects on SSC, nor did it include L3BFLPE external comparisons with country-averaged achievements.

Consequently, this study performed ex-post facto analyses to contribute a fully integrated school-level and cross-cultural BFLPE-CE model to represent the simultaneous effects of individual-, school- and country- averaged math and science achievement on MSC and SSC. These analyses followed the same HLM specifications as those previously reported for discrete L2BFLPE and L3BFLPE results, but additionally analyzed cross comparisons of simultaneous effects, first on MSC then on SSC as the outcome (see Figure 9). Generally, these results supported Marsh (2020) BFLPE-CE model at the individual level, whereby students' math achievement showed a greater positive effect on MSC then the effect of science achievement and students' science achievement showed a greater positive effect on SSC than the effect of students' math achievement.

However, empirically, this study extended the BFLPE-CE one step further to also include the effects of L3BFLPE with the assumption that results of social comparisons with country-averaged achievement would reflect those similar for school-averaged achievement in BFLPE-CE model. Nonetheless, that was not the case here. Actually, only the effects of country-averaged achievement followed BFLPE-CE assumptions, such that country-averaged math achievement ($\beta = -0.022$) showed a more negative effect on MSC than country-averaged science achievement ($\beta = -0.013$). Yet, effects of school-averaged achievement more resembled those at the individual level such that external comparisons to school-averaged math achievement ($\beta = -0.0007$) showed a lesser negative effect on MSC than did school-averaged science achievement ($\beta = -0.005$). Similarly, concerning SSC, the effect of external comparisons with school-averaged science achievement (β = 0.003) was positive when compared to the effect of school-averaged math achievement on SSC was significantly related, likely due to suppression effects of a combined model.

Figure 9

Integrated School-Level and Cross-Cultural BFLPE-CE Model



Notes. Adapted From Model G *"Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: Pan-Human Cross-Cultural Generalizability over 68 Countries" by* Marsh, 2020, European Journal of Personality, 34, p. 184.

Equally as interesting, the collective student-level effects were over and above those of collective country-level effects which were over and above those of school-level effects. This is in direct dispute of the standard BFLPE theory that suggested the negative effects of "external comparisons at the school-level were over and above those of student-level internal comparisons (p. 185)." However, these results do complement initial results of this study that determined discrete L3BFLPE effects in math and science were almost double those of discrete effects of L2BFLPE results in math and science. Respectively, Marsh (2020) rationalizes the varying effects of external comparisons by Bronfenbrenner's (1979) Ecological Model of Human Development suggesting that students' self-concepts are influenced by micro-level processes found in the proximal environment such as at home and with immediate family, meso-level processes such as influences at school, with friends, extended family, social media, or mass media, as well as macro-level processes such as cultural norms, government regulation, social classes, economic systems, and ancestral patterns.

Specifically, he proposed that distal processes did not directly influence the individual, rather indirectly through micro-contexts (p. 195). Thus, these results of discrete analyses of L2BFLPE and L3BFLPE as well as the extended fully integrated BFLPE-CE model suggest a potential hierarchy of influence on the individuals perception of their ability, whereas influences from individual contexts such as internal comparisons with own achievement displayed the greatest size of overall effect on ASC in math and science, while influences from macro-level contexts such as external comparisons with country-averaged achievement demonstrated more effect on ASC than meso-level contexts of external comparisons with school-averaged achievement in math and science.

Research Question 2

Is student-level math and science self-concept significantly associated with student-level achievement, gender, self-concepts, socioeconomic status, valuing and attitudes toward math and science, school-level achievement, socioeconomic status, location, climate, academic self-concept, valuing and attitudes toward math and science or country-level achievement, income per capita, classification of individualism, tracking practices, self-concepts, valuing and attitudes toward math and science across countries (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014)? Notably, individual/micro-level influences explained the greatest percentage of variability in students' math and science self-concepts followed by macro- then meso-level influences that accounted for substantially less of the variations. Respectively, corresponding math predictors at the student-, school-, and country-level accounted for approximately 79%, 8%, and 13% of variability in MSC, while corresponding science predictors accounted for 89%, 5.7% and 5.5% of variability in SSC. In lieu of REM, these results were consistent with those of replicated studies and previous research that reported similar effects on achievement that occurred primarily at the student-level with greater variation accredited to country-level predictors than school-level predictors as well (Areepattamannil et al., 2011; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014).

Overall, students' SES , students' attitude toward math and science, students' valuing of math and science, and students' math and science achievement positively predicted students' corresponding math or science ASC across all countries, whereby as each of these predictors increased, students' ASC also increased in corresponding subject domains. Distinctly, in consecutive order, students' attitude towards math and science (β

= 0.734), value of math (β = 0.241) and science (β = 0.255), and SES in math and science (β = 0.230) were found here to be the strongest positive predictors of their corresponding ASC, while students' math achievement (β = 0.017) and science achievement (β = 0.012) were the weakest positive predictor. Additionally, though the significant negative effect of *gender* (β = -0.043) on MSC reflected higher math self-concepts for females, gender was not a significant predictor of SSC suggesting that there was no significant difference in SSC between males and females.

Generally, these results were consistent with those of previous research that similarly reported that students' SES (Areepattamannil et al., 2011; Bachman & O'Malley, 1977; Chiu & Klassen, 2010; Marsh & Parker, 1984; Strein & Grossman, 2010; Yang, 2003) and affective qualities such as value of math and science (Arens et al., 2019; Guay et al., 2010; Marsh & O'Mara, 2009; Valentine et al., 2004), attitudes toward math and science (Chen et al., 2018; Hacieminoglu, 2016; Osborne et al., 2003; Zimmerman et al., 1992) and students' self-concept positively predicted students' math and science achievement. Likewise, these results also corresponded to those of partially replicated studies that examined similar effects of the same student-level predictors on math and science achievement for a comparable sample of TIMSS data, yet found that students' SES was a stronger positive predictor than students' attitude toward math and students' valuing of math and science (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014).

However, contrastingly, Mohammadpour et al., (2015) found that students' attitude toward science was not a significant predictor of science achievement and only found gender to be a significant predictor of science achievement in half of the countries examined, those of which showed males to have higher science achievement. Also, contrastingly, Mohammadpour et al. (2014) found that gender was not a significant predictor of math achievement, while countless other studies confirmed higher male self-concepts in math (Arens et al., 2017; Helmke & van Aken, 1995; Lee & Kung, 2018; Mohammadpour & Ghafar, 2014; Nagy et al., 2006) and science (Jansen et al., 2014; Marsh et al., 2015; Ruschenpöhler & Markic, 2019; Schroeders & Jansen, 2020; Wilkins, 2004). As well, students' ASC was the strongest positive predictor of math and science achievement in replicated studies and similar previous studies (Areepattamannil et al., 2011; Chiu & Klassen, 2010; Lui & Meng, 2010; Mohammadpour, 2012; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Wilkins, 2004).

Furthermore, school SES in math ($\beta = 0.226$) and science ($\beta = 0.177$), school-averaged attitudes toward math ($\beta = 0.487$) and science ($\beta = 0.559$), school-averaged self-concepts in math ($\beta = 0.669$) and science ($\beta = 0.71$), school-average value of math ($\beta = 0.161$) and science ($\beta = 0.263$), and school-averaged math achievement ($\beta = 0.005$) and science achievement ($\beta = 0.003$) positively predicted students' corresponding ASC in math or science across all countries. On the other hand, school climate negatively predicted students' ASC in math and science ($\beta = 0.-0.034$), but school location did not significantly predict ASC in either subject, suggesting that there is no difference in ASC between urban and rural schools. Distinctly, school-averaged math self-concept and school-averaged attitude toward math were the strongest positive predictors of students' ASC in math and science, while school climate and corresponding school-averaged achievement in math and science predicted students' ASC the least.

Largely, these results were consistent with results of replicated studies and prior research that also reported higher school-averaged SES predicted higher student achievement in math, science and reading for similar samples of older TIMSS or PISA data (Armor et al., 2018b; Caponera & Losito, 2016; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Willms, 2010). However, conversely, replicated studies reported that school location was not only significant, but the strongest predictor of math and science achievement that favored urban schools followed by the effects of a positive school climate, while school-averaged SES was reported to have a lesser negative effect on achievement in math and science (Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). Specifically, it was rationalized that students from smaller and rural schools received an education inferior to that of students from larger urban or suburban schools due to shortages of resources" and quality instruction" (Areepattamannil et al., 2011; Coleman, 1975; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Triandis, 1989; Young, 1998, p. 387) as well as differences in cultural practices (Triandis, 1989; Young, 1998; Zhang et al., 2016).

Concerning school climate, these results were not consistent with previous and replicated studies that reported school climates positively influenced student outcomes (Berkowitz et al., 2017; Caponera & Losito, 2016; Loeb et al., 2019; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014). However, though beyond the scope of this study to confirm or deny, upon closer examination of TIMSS 2019 itemized measurement of "school's emphasis on success" as measured by the school principal, it is possible that a bias exists among the itemized factors of the construct. Specifically, this contextual questionnaire included teacher, parent and student indicators such as emphasis on achievement, parental involvement and respect for high achievers those of which are seemingly biased toward high SES and high achievement standards, thus potentially indirectly promote comparison, competition and judgement that could as well potentially trigger an inferiority complex in some students.

However, conversely Seaton (2010) reported on effects of students that preferred cooperative climates that exacerbated negative impact of social comparisons on perception, yet those results could instead imply a detachment from co-dependent nature rather than simply an anti-competitive nature. Thereafter, Ludtke (2015) found positive effects on ASC concerning the extent to which teachers adapted an individualized frame of reference that focuses on individual students when providing feedback rather than subjectively comparing their performance to that of mean class averages.

Finally, country-averaged self-concept in math ($\beta = 1.008$) and science ($\beta = 1.027$), country-averaged value of math ($\beta = 0.506$) and science ($\beta = 0.537$), and country-averaged attitude toward math ($\beta = 0.485$) and science ($\beta = 0.983$) positively predicted students' corresponding ASC in math or science across all countries, while country-averaged achievement in math ($\beta = -0.009$) and science ($\beta = -0.012$) negatively predicted ASC in math and science, country tracking practices ($\beta = -0.829$) negatively predicted ASC in only math, and country income per capita ($\beta = -0.00003$) negatively predicted ASC in only science. However, country individualism did not significantly predict ASC in either subject, neither did income per capita predict ASC in math nor tracking practices predict ASC in science. Distinctly, country-averaged self-concept, attitude towards, and value of math and science were the strongest positive predictors of MSC, while country-averaged achievement in math and science were the weakest predictors of MSC and SSC as was income per capita for SSC.

Consistent with these results, Hattie (2002) reported little positive effects of tracking with minimal benefits even for the most advantaged groups, while Arens et al., (2017) explained that any benefits of tracking are attributed to improved instructional support for higher tracks and less support in lower tracks. However, whereas these results no difference in ASC for collectivism, previous research reported lower ASC in collectivist cultures with higher ASC in individualistic cultures (Chiu & Klassen, 2010; Chiu & Xihua, 2008; Kashima et al., 1995; Rhee et al., 1996; Triandis, 1989). Conversely as well, Chiu & Klassen (2010) found that IPC had a greater influence on the relationship between ASC and achievement in math, whereas Tucker-Drob et al., (2014) reported a stronger link between science interest and achievement in countries with higher GDP than those in countries with lower GDP.

Research Question 3

School- or country-level BFLPE moderated by student-, school, or country-level variables found to be significantly associated with student-level self-concept in math and science across 26 TIMSS 2019 countries (Seaton, 2010)? Remarkably, all significant student-, school-, and country-level predictors that were applied as moderators either reduced or reversed the negative effects of school-averaged achievement (L2BFLPE) or country-averaged achievement (L3BFLPE) on students' corresponding MSC and SSC. Comparatively, more significant moderation effects were found for L3BFLPE than L2BFLPE in math and science. Additionally, magnitudes of change in BFLPEs attributed to each moderator were slightly greater for L3BFLPE than L2BFLPE in math and science as well, while magnitudes of change were similar for L3BFLPE in both math and science, magnitudes were slightly greater for L2BFLPE in math than science.

Nonetheless, like the few prior BFLPE moderation studies, effect sizes were negligible and the magnitudes of change in BFLPEs were trivial as well. For instance, Seaton et al., (2010) replicated study of the moderation effects of several individual characteristics on L2BFLPE in math using PISA data for a similar size international sample of 15-year-olds also reported many effects as significant, but "not substantial enough to be practically important based on the large sample size and Tymms (2004) effect size standards" (p. 409). Thus, it was in the opinion of this research that a measure revealing the precise changes in BFLPEs attributed to each moderator was of more practical importance rather than simply disregarding significant, but trivial moderation effects as determined by Tymms (2004) measurement of correlational changes (Seaton et al., 2010; Trautwein et al., 2008; Tymms, 2004; Marsh, 2019, Marsh, 2020). Moreover, the calculations for magnitudes of change (β *Diff*) in this study were empirically designed as a supplemental measure to supplement alternative effect size measures.

Notably, concerning student-level moderators, increases in students' science achievement ($\beta = 0.223$) reversed the negative effects of L3BFLPE in science, while increases in students' math achievement ($\beta = 0.030$) also reversed the negative effects of L3BFLPE in math, but to the greatest magnitude of all multilevel moderators of BFLPEs in math and was the only moderator in math and science to register large enough to be of practical importance according to Tymms (2004) effect sizes measures (Marsh et al., 2019, 2020; Seaton et al., 2010). Precisely, these results did not entirely reject the notion of the bright student hypothesis that high achieving students are immune to the negative effects of BFLPE, nor does it grossly accept BFLPE theory that reports all levels of ability suffer from BFLPE. Instead, this moderation effect of achievement suggests that lower ability students would suffer more from the BFLPE in high ability classrooms than higher achieving students that would reflect higher MSC in high achieving math classrooms (Marsh, 2020; Hughet 2009). Yet, as BFLPE theory suggests, this effect size was close to zero implying that increases in MSC for high ability students in high ability schools was not extremely substantial (Marsh, 2020). Likewise, Seaton (2010) reported that ability increased L2BFLPE for more intelligent students in math, though Marsh (2020) found that increases in students' math achievement significantly reduced L2BFLPE in math and notoriously contests the advantages of ability grouping, tracking and streaming practices.

Additionally, increases in students' value of math trivially reversed the negative effects of school-averaged math achievement ($\beta_{Diff} = 0.007$) and country-averaged math achievement ($\beta_{Diff} = 0.026$) on students' ASC in math, while increases in students' value of science only reduced the negative effects of school-averaged achievement ($\beta_{Diff} = 0.0041$) on students' ASC in science not math. Moreover, increases in students' attitude toward math only reduced L3BFLPE in math ($\beta_{Diff} = 0.021$), but increases in students' attitude toward science reversed both L2BFLPE ($\beta_{Diff} = 0.003$) and L3BFLPE ($\beta_{Diff} = 0.023$) in science. Also, increases in students' SES reversed L3BFLPE In science ($\beta_{Diff} = 0.224$). Notably, due to the empirical nature of these BFLPE moderation analyses, few past studies are available to compare with these results, though Seaton (2010) similarly found that higher SES as measured by students' home possessions significantly moderated L2BFLPE in math and science.

Furthermore, concerning significant school-level moderators, increases in school-averaged value of math reduced L3BFLE in math ($\beta_{Diff} = 0.024$), while increases in school-averaged value of science only reduced L3BFLPE in science ($\beta_{Diff} = 0.021$). Additionally, increases in school-averaged attitude toward math reduced only L3BFLPE in math ($\beta_{Diff} = 0.021$), while increases in school-averaged SES reduced L2BFLPE in math ($\beta_{Diff} = 0.009$) and reversed L3BFLPE in science ($\beta_{Diff} = 0.020$). Also, increases in school-averaged math self-concept reduced L3BFLPE in math ($\beta_{Diff} = 0.021$) and increases in school-averaged science self-concept reversed L2BFLPE in science ($\beta_{Diff} = 0.004$). Likewise, increases in school climate reversed L3BFLPE in science ($\beta_{Diff} = 0.022$). Even though Seaton (2010) did not directly measure school climate, results were reported that negative effects of school-averaged achievement (L2BFLPE) were exacerbated for students that preferred a cooperative climate, but justified the "possibility that attending a high-ability school, promotes more competition that fosters yearnings to work more cooperatively for some student" (p. 420).

Finally, concerning country-level moderators, these results confirmed that increases in country-averaged achievement in math reversed L2BFLPE in math ($\beta_{Diff} = 0.006$), while increase in country-averaged value of math ($\beta_{Diff} = 0.029$) and increases in country-averaged academic self-concept in math ($\beta_{Diff} = 0.029$) reversed the negative effects of L3BFLPE in math. Notably, as well, increases in country-averaged attitude toward science ($\beta_{Diff} = 0.037$) reversed the negative effects of L3BFLPE in science reflecting the greatest magnitude of change of all L2BFLPE and L3BFLPE moderators in math and science, though the effect size was unsubstantial according to Tymms (2004).

In short, these results predicted that the negative effects on students' math self-concept from external comparisons to school-averaged level of math ability (L2BFLPE in math) would be significantly ameliorated if students valued math more, if their school population was composed of more financially advantaged students, and if the averaged ability of their home country was higher. Similarly, the negative effects on students' math self-concept from external comparisons to country-averaged math achievement (L3BFLPE in math) would be significantly ameliorated if students valued math more, if students reflected elevated attitudes toward math, if students demonstrated greater math ability, if their school collectively valued math more, collectively demonstrated elevated attitudes toward math and collectively reflected higher math-self-concepts, as well as if their home country collectively valued math more and demonstrated collectively higher math self-concepts.

Correspondingly, these results predicted that the negative effects on students' science self-concept from external comparisons with school-averaged science achievement (L2BFLPE in science) would be significantly ameliorated if students valued science more, if students reflected elevated attitudes toward science, and if their school collectively demonstrated higher science self-concepts. Likewise, the negative effects on students' science self-concept from external comparisons to country-averaged science achievement (L3BFLPE) would be significantly ameliorated if students demonstrated an elevated attitude toward science, if students come from a more advantaged home environment, if students demonstrate overall improved science ability, as well as if the school was composed of more financially advantaged students and exhibited greater emphasis on success, as well as if their school and home country reflected collectively elevated attitudes toward science.

Interpretation and Implications of Findings

BFLPE research is indeed a valuable avenue to gain insight into the processes and influences on students' perception of their own academic ability, a construct that is mutually beneficial and reciprocally related to academic achievement. Ultimately, the design and results of this study have extended current BFLPE theory and advanced research in educational psychology with implications to evolve current policy and practice in STEM education. By and large, regarding BFLPE theory, thus study indeed illustrated that the Big-Fish-Little-Pond Effect (BFLPE) is a pan-human, universal phenomena that affects students of all abilities worldwide. Generally, students' perception of ability was positively affected by internal comparisons of ability in the corresponding subject domain, but less so for internal comparisons of achievement in the more distant domain. In practice with respect to the Dimensional Comparison Theory, these models can distinguish unforeseen subject preferences in either math or science domains, "such that students can be aware of the relative strengths of their abilities, channel their development of unique abilities effectively, and best use their abilities in different domains creatively for different contexts" (Chiu, 2012; Huguet et al., 2009; Marsh et al., 2020; Marsh et al., 2014).

On the other hand, as these results were consistent with most generalizable BFLPE findings, implications concerning the detrimental effects of social comparisons on student perceptions of their ability and reciprocally their achievement as well, are particularly more alarming than ever in today's era of social media and remote learning.

Even so, critics often challenge ability grouping, streaming, and tracking practices suggesting that placement of high ability students in high ability tracks exacerbate the negative effects of social comparisons compared to equally able students placed in average or lower ability tracks (Dicke et al., 2018b; Marsh et al., 2008, 2014, 2015; Salchegger, 2016; Trautwein et al., 2008).

Distinctly, visualizations here illustrated that high ability students in highest achieving schools and countries had the lowest academic self-concerts in math and science. However, patterns in these results revealed that affective, cognitive, financial, environmental, and cultural aspects positively influenced perceptions of ability and positively moderated the negative effects of BFLPEs. Collectively, regarding ability grouping, these findings suggest that ability alone is not the only source of influence on ASC, thus it could be a flaw in the design of basing grouping, streaming, and tracking practices exclusively on ability that is prompting detrimental effects of social comparisons. Therefore, regardless of their minimal effect sizes, these patterns present viable indicators to consider when evaluating or revising grouping, streaming and tracking practices in STEM subjects.

Precisely, Bloom et al. (1956) Domains of Learning provided a framework for a more holistic approach to education which demonstrates the most relevant application to the results of this study as the cognitive and affective domains he described coincide with the indicators reported here as well. Taken together, designs for explicit school wide tracking practices as well as implicit grouping and streaming practices within the classroom can be modified and improved by adapting the empirical framework of the ex-post facto extension of Marsh et al. (2020) BFLPE-CE model presented in this study (see

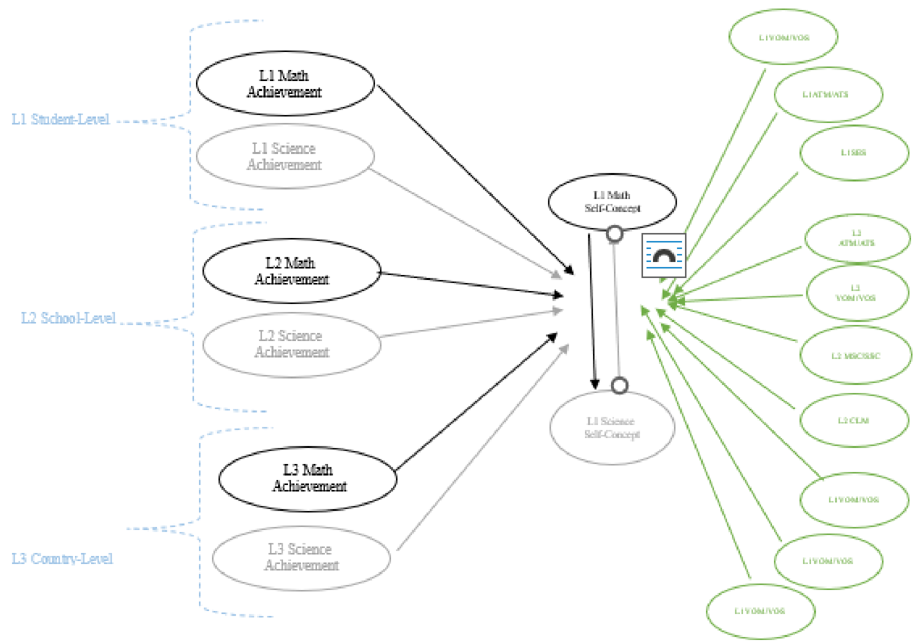
Figure 8). This framework could also be adapted across multiple subject domains to include amicable consideration for not only cognitive indicators such as achievement in, but also affective and financial indicators. Concisely, the combined format, hypothetically referred to here as the “education indicator sequence,” is an adaptation of Marsh et al. (2020) Fully Integrated Social and Dimensional Comparison Model (see Figure 10) that when generated for each individual student, could be applied as a reference to identify subject domains in which students excel (BFLPE-CE) and also be cross-referenced to compatibly group students according to a wider range of indicators rather than exclusively on ability alone, similar to a protein sequence, like DNA or a compatibility algorithm like those used for online dating apps.

Furthermore, equally as important as identifying indicators and compatibly grouping students is understanding the underlying processes that impact students’ overall academic self-concepts as well. Essentially, these results revealed that perception is influenced most by individual and micro-level aspects, but this does not come as a surprise considering that academic self-concept is a uniquely individual construct. However, it was quite a revelation to discover that perception was influenced more by country-level aspects than school-level aspects. Indeed, a viable justification attributes this phenomenon to that of Bronfenbrenner (1979) Ecological Model of Learning such that meso- and macro-level aspects influence individuals through more local contexts (Marsh et al., 2020).

However, it is in the opinion of this research that the greater effects of country-level comparisons and well as the fact that discrete country-level influences accounted for more variation in ASC, implies a hierarchy of influence on students’ self-concepts.

Figure 10

Educational Indicator Sequence



Notes. Concept map adapted From Model H *"Psychological Comparison Processes and Self-Concept in Relation to Five Distinct Frame-of-Reference Effects: Pan-Human Cross-Cultural Generalizability over 68 Countries" by* Marsh, 2020, European Journal of Personality, 34, p. 180-202.

Alternatively, Jung (1921) Structure of the Psyche justifies this phenomena. In general, the self is described as being composed of the ego, the personal unconscious, and the collective unconscious. To be precise, analogies can link Bronfenbrenner (1979) and Jung (1921) theories to rationalize these results in metaphysical terms, though beyond the scope of this study to confirm interpretations. Hypothetically, the ego can be characterized as the conscious reactions to reality and represented here as the measurable decline in ASC, the personal unconscious can be characterized as unmeasurable effects of micro-, meso-, and macro-level influences, and the collective unconscious can be characterized as the preprogrammed, neural conditioning, we are born with that has been passed from previous generations. Thus, the hierarchy of influence shown in these results can be rationalized as a series of unconscious filters that form our perceptions and corresponding behavior, whereby the collective unconscious that supremely influences the personal unconscious and subsequent ego.

Alternatively, this can be visualized by imagining that everyone is born wearing sunglasses with uniquely colored lenses that represent our culturally preprogrammed, neural conditioning or collective unconscious. Thus, the way we view life experiences through our own eyes or personal unconscious is filtered through the lens of our sunglasses. It is these filters that persuade our physical reactions and behavior. Comparatively, not everyone is wearing the same sunglasses, so some may have brighter lenses that easily "reflect the glory" of more successful or counterparts Cialdini & Richardson (1980) while other sunglasses may be a darker shade that " relatively deprive" Davis (1966) those wearing them from seeing the bright side of life experiences and respond with negative, defensive or inferior reactions.

Remarkably, perception of ability (ASC) is most influenced at the student-level indicating that students are able to override the impact of the unconscious lenses that filter perceptions. Therefore, in terms of educational policy and practice, designing curriculum that involves student-centered, individualized ancestorial research such as lineage studies with complimentary genetic testing to identify collective unconscious patterns would be beneficial. For instance, classroom practices such as collective guidance counseling sessions or peer mentorship to identify and address self-defeating patterns of thought and enhance positive self-talk, designing classroom environments that promote positive emphasis on success as uniquely defined by the range of possibility unique to each student, implement group bonding and peer encouragement activities geared toward diverse combinations of learners through in class role play or outdoor adventure retreats.

Overall, in order to address the negative repercussions of social comparisons on academic self-concepts and corresponding achievement, the results of this study suggest that education policy and practice 1. Implement individualized perception profiles to extract hidden capabilities of students in STEM subjects 2. Compatibly groups of students according to adaptations of the multilevel educational indicator sequence to form classroom climates that minimize potential superiority or inferiority complexes 3. Implement curriculum that harnesses a variety of student-centered, individualized instructional techniques within diverse group settings and inspires students to identify, accept, and appreciate their unique attributes, yet move beyond the limitation of their cultural conditioning to evolve into the best version of themselves and maximize their greatest academic and human potential.

Strengths, Limitations, and Future Research

Overall, this study has commendably contributed to advancing current BFLPE research as well as research in educational psychology as a whole. Specifically, these results extended BFLPE research in four major ways. First, the most current version of

HLM 8.2 software (release date 2019) was employed here to analyze the most current results of TIMSS 2019 large-scale, international assessment in math and science that measured grade-specific, curricular knowledge, whereas replicated studies employed older versions of HLM or MLwin software (release date 2002) to analyze primarily only outdated results from TIMSS or PISA large scale, international assessment that measured age-appropriate, general knowledge in math and science for 15 year olds (Marsh et al., 2015, 2020; Mohammadpour et al., 2015; Mohammadpour & Ghafar, 2014; Seaton et al., 2010). Second, this study was the first of its kind to simultaneously examine multileveled moderation effects of both L2BFLPE and L3BFLPE in math and science, whereas replicated studies investigated effects of BFLPE only in math with far fewer moderators.

Additionally, this study has provided an empirical, supplementary effect size measure of change in L2BFLPE and L2BFLPE attributed to each moderator accompanied by its corresponding statistical significance value to discretely determine the precise change in L2BFLPE and L2BFLPE attributed to each statistically significant moderator, whereas replicated studies only reported Tymms (2004) effect size measures a general difference in correlation coefficients corresponding to one degree above and one degree below the moderators mean. Third, this is the first study of its kind to extend research in educational psychology by offering insight into significant micro-, meso-, and macro-level influences on students' academic self-concept in math and science. Fourth, this study offers an empirical "educational indicator sequence" framework as a reference to not only improve ability grouping, but also alleviate the negative repercussions inevitable social comparisons as well as a complimentary metaphysical rationale in contribution to a greater understanding of the underlying processes of ASC.

However, there are several limitations to consider as well. Fundamentally, many of these effects on ASC were measured discreetly when in reality they coexist. Thus, it would be beneficial to replicate models with combined specifications to determine a more realistic perspective. Too, as achievement is of great concern in education, it was assumed here that the reciprocal effects model (REM) held true for this study. However, future research could apply multilevel, structural equation modeling to not only examine the reciprocal effects of ASC and corresponding achievement, but also to examine if the indicators of BFLPE mediate or moderate that reciprocal relationship. Nonetheless, this study was principally investigated using cross-sectional data that was collected at only one time point limits, so future research could additionally examine longitudinal effects as well as

Moreover, similar to past BFLPE research, the justification for internal and external comparisons as a rationale for the negative effects of BFLPE were only implied and not directly measured here (Dai & Rinn, 2008; Huguet et al., 2009; Jonkmann et al., 2012). Therefore, future research could apply mixed methods replications of this study, so supplemental qualitative measures would contribute a more personal perspective concerning quantitative results. As well, the sample consisted of only developed countries with Egypt being the only developing country represented including an underrepresentation of the sub-Saharan and Latin American regions, so future research should include a more diverse range of countries. Also, generalizability of these results only applies to the specifications of this particular study and sample, but future research could consider replications such that ordinally measured tracking (L3TRK) and continuously measured individualism (L3IDV) and Income Per Capita (L3IPC) were instead dichotomously dummy coded.

Finally, drawing from the literature review, future BFLPE research could model outcomes with other closely related constructs such as self-efficacy, self-esteem, or emotional intelligence to determine if BFLPE is indeed only an effect on ASC or does it hold true for similar constructs as well. Furthermore, experimental designs would advance BFLPE research in the future. For instance, examine effects on BFLPE from implementing the "educational indicator sequence" to compatibly group students versus ungrouped students. As well, future research could examine the effects of BFLPE for treatment group that participates in an accredited social and emotional learning (SEL) program versus a control group that does not. As well longitudinal experimental designs could substantiate REM and extend BFLPE by comparing interventions based on Calsyn and Kenny (1977) skill development model versus interventions based on self enhancement model to determine effects BFLPE. Furthermore, future studies should apply longitudinal and mixed methods designs to investigate effects of BFLPE for diverse populations discretely based on ethnicity, LGBTQ, SES, and ability to really one in on which students are affected most by BFLPE and social comparisons.

Conclusion

For over two decades, BFLPE results have confirmed the global generalizability of negative effects of school- and country- averaged achievement on students' academic self-concept based on social comparisons with implications that generally dispute ability grouping and tracking at both ends of the ability spectrum. However, those investigations have often been narrow in scope, applied a limited statistical designs, and reported on only a few multileveled variables of influence that could ameliorate those negative effects. Accordingly, this study was the first of its kind to apply a complex hierarchical linear modeling design with state-of-the-art software to simultaneously examine student, school, and country-level moderation effects of both L2BFLPE and L3BFLPE in math and science using results from TIMSS 2019 International large-scale assessment for a sample of 169,810 eight grade students in 5,410 school in 26 countries. Overall, these results advanced current BFLPE theory by revealing specific affective, cognitive, environmental, and financial factors that diminished the negative effects of school-and country-level BFLPE and offered the "educational indicator sequence" as a complimentary standardized framework by which students could be more compatibly group, stream or tracked. Furthermore, this study has contributed to research in educational psychology with implications suggesting a hierarchical structure of the social comparison process whereby individuals have the greatest overall impact on their perceptions, but macro-level, unconscious cultural preprogramming is the overarching influence through which perceptions are filtered.

Nonetheless, with the magnitude of global recognition and prioritization STEM education, the long debate concerning benefits of segregating the most able students as a means of improving achievement continues. Likewise, today's society of technology and social media has inevitably intensified the negative impact of social comparisons on self-concept in the classroom. Therefore, more BFLPE research is necessary to persuade policy and practice alike to design curriculums and climates that improve achievement, while addressing the underlying psychological processes that encourage students to move beyond the limitation of their cultural conditioning so they may become the best version of themselves, maximize their greatest academic potential and improve their overall well-being. *"With realization of one's own potential and self-confidence in one's own ability, one can build a better world"*

-Dalai Lama

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*, 485–503.

Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Influence of motivation, self-beliefs, and instructional practices on science achievement of adolescents in Canada. *Social Psychology of Education*, *14*(2), 233–259. https://doi.org/10.1007/.s11218-010-9144-9

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & von Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*(5), 621–634. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/edu0000163

Arens, A. K., Schmidt, I., & Preckel, F. (2019). Longitudinal relations among self-concept, intrinsic value, and attainment value across secondary school years in three academic domains. *Journal of Educational Psychology*, *111*(4), 663–684.

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018a). The impact of school SES on student achievement: Evidence from U.S. statewide achievement data. *Educational Evaluation and Policy Analysis*, *40*(4), 613–630. https://doi.org/ 10.3102/0162373718787917

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018b). The impact of school SES on student achievement: Evidence from U.S. statewide achievement data. *Educational Evaluation and Policy Analysis*, *40*(4), 613–630. https://doi.org/10.3102/0162373718787917

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of*

Personality and Social Psychology, *35*, 365–380. https://doi.org/10.1037//0022-3514.35.6.365

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). Academic Press.

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, *87*(2), 425–469. https://doi.org/10.3102/0034654316669821

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression*. Guilford Publication, Inc.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive Domains*. David McKay Company.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, *57*, 49–64.

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of Educational Psychology*, *90*, 102–110.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/1040-726X/03/0300-0001/0

Bronfenbrenner, U. (1979). *Ecology of human development: Experiments by nature and design.* Harvard University Press.

Brookover, W. B., Thomas, S., & Paterson, A. (1964). Self-concept of ability and school achievement. *Sociology of Education*, *37*(3), 271. https://doi.org/10.2307/2111958

Burns, R. A., Crisp, D. A., & Burns, R. B. (2020). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the cross-lagged panel and random-intercept cross-lagged panel frameworks. *British Journal of Educational Psychology*, *90*(1), 77–91. https://doi.org /10. 1111 /bjep.12265

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*, *55*(3), 31.

Byrne, B. M. (1988). Adolescent self-concept, ability grouping, and social comparison: Reexamining academic track differences in high school. *Youth and Society*, *20*(1), 46–67.

Byrne, B. M. (1998, April). *Testing for causal predominance between academic self-concept and academic achievement: A developmental perspective.* Annual Meeting of the American Educational Research Association, San Diego, CA.

Byrne, B. M., & Gavin, D. A. W. (1996). The Shavelson Model revisited: Testing for the structure of academic self-concept across pre-, early, and late adolescents. *Journal of Educational Psychology*, *88*(2), 215–228. http://dx.doi.org.ezproxylocal. library.nova.edu/10.1037/0022-0663.88.2.215

Byrne, B. M., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology*, *78*(6), 474–481. http://dx.doi.org. ezproxylocal.library.nova.edu/10.1037/0022-0663.78.6.474

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, *69*(2), 136–145. http://dx.doi.org.ezproxylocal. library.nova.edu/ 10.1037/0022-0663.69.2.136

Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-Scale Assessments in Education*, *4*(1), 12. https://doi.org/10.1186/s40536-016-0030-6

Centurino, V. A. S., & Jones, L. R. (2017). Chapter 2: TIMSS 2019 science frameworks. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 27–55). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/science-framework/

Chang, Y.-L. (2012). A study of fifth graders' mathematics self-efficacy and mathematical achievement. *Asia-Pacific Education Researcher*, *21*(3), 519–525.

Chen, L., Bae, S. R., Battista, C., Qin, S., Chen, T., Evans, T. M., & Menon, V. (2018). Positive attitude toward math supports early academic success: Behavioral evidence and neurocognitive mechanisms. *Psychological Science*, *29*(3), 390–402. https://doi.org/10.1177/0956797617735528

Cheng, R. W., McInerney, D. M., & Mok, M. (2014). Does big-fish–little-pond effect always exist? Investigation of goal orientations as moderators in the Hong Kong context. *Educational Psychology*, *34*(5), 561–580. https://doi.org/ 10.1080/ 01443410.2014.898740

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2–17. https://doi.org/Cal

Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction*, *18*(4), 321–336. https://doi.org/10.1016/j.learninstruc.2007.06.003

Chiu, M.-S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology*, *104*(1), 87–107. https://doi.org/10.1037/a0025734

Chui, M. M., & Klassen, R. M. (2008). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*, 2–17.

Cialdini, R. B., & Richardson, K. D. (1980). Two indirect tactics of image management: Basking and blasting. *Journal of Personality and Social Psychology*, *39*(3), 406–415. https://doi.org/10.1037/0022-3514.39.3.406

Cohen, J., Cohen, P., West, S. A., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* (3rd ed.). Lawrence Eribaum Associates, Inc.

Coleman, J. S. (1975). Methods and Results in the IEA Studies of Effects of School on Learning. *Review of Educational Research*, *45*(3), 33.

Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, A. M., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of education for all*. US Government Printing Office.

Cooley, C. H. (1902). *Human nature and the social order*. Charles Scribner's Sons.

Coopersmith, S. (1959). A method for determining types of self-esteem. *The Journal of Abnormal and Social Psychology*, *59*(1), 87–94. https://doi.org/10.1037/h0048001

Coopersmith, S. (1967). *The antecedents of self-esteem.* Freeman.

Cotter, K. E., Centurino, V. A. S., & Mullis, I. V. S. (2020). Chapter 1: Developing the TIMSS 2019 mathematics and science achievement instruments. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (p. 1.1-1.36). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-1.html

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson Education, Inc.

Cronbach, L. J. (1972). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). American Council on Education.

Crowne, D. P., & Stephens, M. W. (1961). Self-acceptance and self-evaluative behavior: A critique of methodology. *Psychological Bulletin*, *58*(2), 104–121. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/h0047015

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29.

Currell, D. (2015). *Scientific data analysis: Difference in slope*. Oxford University Press. https://www.youtube.com/watch?v=myL_qzuLHTQ

Dai, D. Y., & Rinn, A. N. (2008a). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review*, *20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x

Dai, D. Y., & Rinn, A. N. (2008b). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review*, *20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x

Dai, Y. D., & Rinn, A. N. (2008). The big-fish-little-pond-effect: What do we know and where do we go from here? *Educational Psychology Review; New York*, *20*, 283–317.

Davis, J. A. (1959). A formal interpretation of the theory of relative deprivation. *Sociometry*, *22*(4), 280–296. https://doi.org/10.2307/2786046

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *The American Journal of Sociology*, *72*(1), 17–31.

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

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018b). 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

Eklöf, H. (2007). Self-concept and valuing of mathematics in TIMSS 2003: Scale structure and relation to performance in a Swedish setting. *Scandinavian Journal of Educational Research*, *51*(3), 297–313. https://doi.org/ 10.1080/ 00313830701 356141

Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, *28*(5), 404–416. http://dx.doi.org.ezproxylocal.library. nova.edu /10.1037/h0034679

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Fishbein, B., Foy, P., & Tyack, L. (2020). Chapter 10: Reviewing the TIMSS 2019 achievement item statistics. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (p. 10.1-10.70). TIMSS & PIRLS International Study Center, Boston College. https:// timssandpirls.bc.edu/timss2019/methods/chapter-10.html

Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 user guide for the international database* (2021930389). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/international-database/

Foy, P. (2017). *TIMSS 2015 User guide for the international database*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UserGuide.pdf

Gao, S., Long, H., Li, D., & Yang, L. (2020). The mediation effect of student self-efficacy between teaching approaches and science achievement: Findings from 2011 TIMSS US data. *Social Psychology of Education*, *23*, 385–410. https://doi.org/10.1007/s11218-019-09534-1

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–717. https://doi.org/10.1037/a0026659

Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Green, S. B., & Salkind, N. J. (2017). *Using SPSS for windows and macintosh: Analyzing and understanding the data* (8th ed.). Pearson Education, Inc.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/0022-0663.95.1.124

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*(6), 644–653. https://doi.org/10.1016/j.lindif.2010.08.001

Hacieminoglu, E. (2016). Elementary school students' attitude toward science and related variables. *The International Journal of Environmental and Science Education*, *11*(2), 35–52. https://doi.org/10.12973/ijese.2016.288a

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, *52*(1), 123–142. https://doi.org/10.2307/1170275

Hart, S. C., DiPerna, J. C., Lei, P.-W., & Cheng, W. (2020). Nothing lost, something gained? Impact of a universal social-emotional learning program on future state test performance. *Educational Researcher*, *49*(1), 5–19. https://doi.org/ 10.3102 /0013189X19898721

Hattie, J. (1992). *Self-concept*. Lawrence Eribaum Associates, Inc.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*(5), 449–481. https://doi.org/10.1016/S0883-0355(03)00015-6

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). Routledge.

Helmke, A., & van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*(4), 624–637. http://dx.doi.org. ezproxylocal.library.nova.edu/10.1037/0022-0663.87.4.624

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). SAGE.

Hooper, M. (2016). Chapter 2: Developing the TIMSS 2015 Context questionnaires. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (p. 2.1-2.8). TIMSS and PIRLS International Study Center, Boston College and International Association of the Evaluation of Educational Achievement. http://timss.bc.edu/publications/timss/2015-methods/chapter-2.html

Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013a). Chapter 3: TIMSS 2015 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (p. 23). TIMSS and PIRLS International Study Center, Boston College and International Association of the Evaluation of Educational Achievement.

Hooper, M., Mullis, I. V. S., & Martin, M. O. (2013b). TIMSS 2015 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 61–82). Boston College, TIMSS & PIRLS International Study Center.

Hooper, M., Mullis, I. V. S., Martin, M. O., & Fishbein, B. (2017). Chapter 3: TIMSS 2019 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 57–78). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/context-questionnaire-framework/

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Huguet. (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.

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: An integrative study. *Journal of Personality and Social Psychology*, *97*(1), 156–170. https://doi.org/10.1037/a0015558

IEA. (2016). *About TIMSS 2015*. https://timss2015.org/download-center

James, W. (1890). *The principles of psychology* (Vol. 1). Peter Smith.

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. https://doi.org/10.1016 /j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, *30*, 11–21. https://doi.org/ 10.1016/ j. lindif. 2013.12.003

Johansone, L. (2020). Chapter 6: Survey operations for TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019*

Technical Report (p. 6.1-6.28). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-6.html

Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2012). Personality traits moderate the big-fish–little-pond effect of academic self-concept. *Learning and Individual Differences*, *22*(6), 736–746. https://doi.org/10.1016/j.lindif.2012.07.020

Jung, C. G. (1921). *Psychological types. The collected works of CG Jung.* (Vol. 6). Bollingen Series XX.

Jun-Li Chen, J. (2005). Relation of academic support from parents, teachers, and peers to Hong Kong adolescents' academic achievement: The mediating role of academic engagement. *Genetic, Social, and General Psychology Monographs*, *131*(2), 77–127.

Kaplan, H. B., & Pokorny. (1970). Ageing and self-attitude: A conditional relationship. *Ageing and Human Development*, *1*, 241–250.

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S.-C., & Gelfand, M. J. (1995). *Culture, Gender, and Self: A Perspective From Individualism-Collectivism Research*. *69*(5), 13.

Kelly, D. L., Centurino, V. A. S., Martin, M. O., & Mullis, I. V. S. (2020). *TIMSS 2019 encyclopedia: Education policy and curriculum in mathematics and science* (2020922703). TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/encyclopedia/

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Rand McNally.

Labenne, W. D., & Greene, B. (1969). *Educational implications of self-concept theory*. Goodyear Publishing Company, Inc.

LaRoche, S., Joncas, M., & Foy, P. (2020). Ch. 3: Sample design in TIMSS 2019. In *Methods and procedures: TIMSS 2019 technical report* (p. 3.1-3.33). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc. edu/ timss2019/methods/chapter-3.html

Lecky, K. (1945). *Self-consistency: A theory of personality*. Island Press.

Lee, C.-Y., & Kung, H.-Y. (2018). Math self-concept and mathematics achievement: Examining gender variation and reciprocal relations among junior high school students in Taiwan. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(4), 1239–1252. https://doi.org/10.29333/EJMSTE/82535

Lindquist, E. F. (1953). *Design and analysis of experiments in education and psychology*. Houghton Mifflin Company.

Lindquist, M., Philpot, R., Mullis, I. V. S., & Cotter, K. E. (2017). Chapter 1: TIMSS 2019 mathematics frameworks. In *TIMSS 2019 assessment frameworks* (pp. 11–25). TIMSS & PIRLS International Study Center, Boston College. https:// timssandpirls. bc.edu/timss2019/frameworks/framework-chapters/ mathematics-framework/

Loeb, S., Christian, M. S., Hough, H., Meyer, R. H., Rice, A. B., & West, M. R. (2019). School differences in social–emotional learning gains: Findings from the first large-scale panel survey of students. *Journal of Educational and Behavioral Statistics*, *44*(5), 507–542. https://doi.org/10.3102/1076998619845162

Lui, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics. *Educational Psychology*, *30*(6), 699–712.

Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review. *Educational Psychology Review*, *9*(2), 89–120.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. http://dx.doi.org. ezproxylocal.library. nova.edu/10.1037/0022-0663.79.3.280

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*, 623–636. https://doi.org/10.1037/ 0022-0663.82.4.623

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal*, *28*(2), 445–480.

Marsh, H. W. (2006). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. British Psychological Society.

Marsh, H. W. (2005). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology* [Vernon-Wall Keynote Address].

Educational Psychology section of the British Psychological Society, Durham University.

Marsh, H. W., Abduljabbar, A. S., Morin, A. J. S., Parker, P., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*(1), 258–271. https://doi.org/10.1037/a0037485

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of U.S. and Saudi Arabian TIMSS responses. *Journal of Cross-Cultural Psychology*, *45*(5), 777–804. https://doi.org/ 10.1177/002202 2113519858

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., Möller, J., & Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: Age-cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202. https://doi.org/10.3102/0002831214549453

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and Its relation to academic achievement. *Journal of Educational Psychology*, *80*, 366–380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and... *Educational Psychologist*, *34*(3), 155–167. https://doi.org/10.1207/s15326985ep3403_2

Marsh, H. W., & Craven, R. (2002). The pivotal role of frames of reference in academic self-concept formation: The "big fish-little pond" effect. In F. Pajaras & T. C. Urdan (Eds.), *Academic Motivation of Adolescents* (pp. 82–132). Information Age Pub.

Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child Development*, *69*(4), 1030–1053. https://doi.org/10.1111/j.1467-8624.1998.tb06159.x

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology*, *83*(3), 377–392. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/0022-0663.83.3.377

Marsh, H. W., & Hattie, J. A. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). Wiley.

Marsh, H. W., & Hau, K. T. (2003a). Big-fish-little-pond effect on academic self-concept. *American Psychologist*, *58*(5), 364–376.

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*, 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Marsh, H. W., & Hau, K.-T. (2003b). Big-Fish--Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. https://doi.org/ 10. 1037/0003-066X.58.5.364

Marsh, H. W., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and west German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology*, *29*(3), 264–282. https://doi.org/10.1016/S0361-476X(03)00034-1

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000a). Longitudinal multilevel models of big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology*, *93*(3), 543–553. https://doi.org/10.3102/0002831214549453

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000b). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037/0022-3514.78.2.337

Marsh, H. W., Kuyper, H., Morin, A. J. S., 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., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., & O'Mara, A. (2009). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of big-fish-little-pond effect. *German Journal of Educational Psychology*, *24*(1), 51–72. https://doi.org/10.1024/1010-0652/a000004

Marsh, H. W., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie*, *24*(1), 51–72. https://doi.org/10.1024/1010-0652/a000004

Marsh, H. W., & O'Neill, R. (1984). Self-description questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, *21*, 153–174. https://doi.org/10.1111/j.1745-3984. 1984.tb00227.x

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. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/0022-3514.47.1.213

Marsh, H. W., Parker, J. W., & Smith, I. D. (1983). Pre- adolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. *British Journal of Educational Psychology*, *53*, 60–78. https://doi.org/10.1111/j.2044-8279.1983.tb02536.x

Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020a). 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., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020b). 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., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist*, *24*(3), 231–242. https://doi.org/10.1027/1016-9040/a000332

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self--concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and Social Psychology*, *45*(1), 173–187. https://doi.org/10.1037/0022-3514.45.1.173

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. https://doi.org/ 10.1207 /s15326985ep2003_1

Marsh, H. W., Trautwein, U., Ludtke, O., Baumert, J., & Koller, O. (2007). The big-fish-little- pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*(3), 631–669.

Marsh, H. W., Trautwein, U., Ludtke, O., & Koller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*, 510–524.

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, *89*(1), 41–54. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1037/0022-0663.89.1.41

Martin, M. O., Mullis, I. V. S., & Foy, P. (2013). Chapter 4: TIMSS 2015 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (pp. 85–98). TIMSS and PIRLS International Study Center, Boston College and International Association of the Evaluation of Educational Achievement.

Martin, M. O., Mullis, I. V. S., & Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (p. 8.1-9.1). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/ assessment-design/

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in science*. TIMSS & PIRLS International Study Center, Boston College and The International Association for the Evaluation of Educational Achievement. http://timssandpirls.bc.edu/timss2015/international-results/

Marx, R. W., & Winne, P. H. (1978). Construct interpretations of three self-concept inventories. *American Educational Research Journal*, *15*(1), 99–109.

McFarland, C., & Buehler, R. (1995). Collective self-esteem as a moderator of the frog-pond effect in reactions to performance feedback. *Journal of Personality and Social Psychology*, *68*(6), 1055–1070.

Mead, G. (1934). *Mind, self and society from the standpoint of a social behaviorist*. University of Chicago Press. http://livros01.livrosgratis.com.br/bu000001.pdf

Mohammadpour, E. (2012). *Factors accounting for mathematics achievement of Singaporean eighth-graders*. *21*(3), 507–518.

Mohammadpour, E., & Ghafar, M. N. A. (2014). Mathematics achievement as a function of within- and between-school differences. *Scandinavian Journal of Educational Research*, *58*(2), 189–221. https://doi.org/10.1080/00313831.2012.725097

Mohammadpour, E., Shekarchizadeh, A., & Kalantarrashidi, S. A. (2015). Multilevel modeling of science achievement in the TIMSS participating countries. *The Journal of Educational Research*, *108*(6), 449–464. https://doi.org /10. 1080/ 00220671.2014.917254

Moller, J., & Koller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, *93*, 826–835.

Moller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*, 544–560. https://doi.org/10.1037/a0032459

Möller, J., & Pohlmann, B. (2010). Achievement differences and self-concept differences: Stronger associations for above or below average students? *British Journal of*

Educational Psychology, *80*(3), 435–450. https://doi.org/10.1348/000709909X485234

Moller, J., Pohlmann, B., Koller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and aca- demic self-concept. *Review of Educational Research*, *79*, 1129–1167.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Moller, J., Retelsdorf, J., Koller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, *48*(6), 1315–1346. https://doi.org/10.3102/0002831211419649

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scales, F. (2014). Doubly latent multilevel analyses of classroom climate: An illustration. *Journal of Experimental Education*, *82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412

Mullis, I. V. S. (2017). Introduction. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 1–10). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/ timss2019/frameworks/ framework-chapters/introduction/

Mullis, I. V. S., Cotter, K. E., Centurino, V. A. S., Fishbein, B., & Liu, J. (2016). Chapter 14: Using scale anchoring to interpret the TIMSS 2915 achievement scale. In M.

O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (p. 14.1-14.47). TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Cotter, K. E., Fishbein, B. G., & Centurino, V. A. S. (2016). Chapter 1: Developing the TIMSS 2015 achievement items. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (p. 1.1-1.22). TIMSS and PIRLS International Study Center, Boston College and International Association of the Evaluation of Educational Achievement. http://timssandpirls.bc.edu/publications/timss/2015-methods.html

Mullis, I. V. S., & Fishbein, B. (2020a). Chapter 2: Updating the TIMSS 2019 instruments for describing the contexts for students learning. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (p. 2.1-2.9). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-2.html

Mullis, I. V. S., & Fishbein, B. (2020b). Chapter 15: Using scale anchoring to interpret the TIMSS 2019 achievement scales. In M. O. Martin, M. von Davier, & I. V. S Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (p. 15.1-15.60). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-15.html

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the eighth and fourth grades*. TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. TIMSS & PIRLS International Study Center.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: The big-fish-little-pond-effect for PISA 2006. *Educational Psychology*, *31*(5), 629–656.

Nagengast, B., & Marsh, H. W. (2012). Big fish little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, *12*(4), 323–345. https://doi.org/10.1080/13803610600765687

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. http://dx.doi.org/10.1007/s10459-010-9222-y

O'Mara, A. J., Green, J., & Marsh, H. W. (2006). Administering self-concept interventions in schools: No training necessary? A meta-analysis. *International Education Journal*, *7*(4), 524–533.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Pajaras, F., & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, *86*, 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Papanastasiou, C., & Papanastasiou, E. C. (2004). Major influences on attitudes toward science. *Educational Research and Evaluation*, *10*(3), 239–257. https://doi.org /10.1076/edre.10.3.239.30267

Parker, P. D., Marsh, H. W., Ciarrochi, J. M., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievment outcomes. *Educational Psychology*, *34*, 29–48.

Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, *23*, 114–133.

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: Personality Processes and Individual Differences*, *117*(1), 166–185. http://dx.doi.org/10.1037/pspp0000230

Plieninger, H., & Dickhäuser, O. (2015). The female fish is more responsive: Gender moderates the BFLPE in the domain of science. *Educational Psychology*, *35*(2), 213–227. https://doi.org/10.1080/01443410.2013.814197

Primavera, L. H., Simon, W. E., & Primavera, A. M. (1974). The relationship between self-esteem and academic achievement: An investigation of sex differences. *Psychology in the Schools*, *11*(2), 213–216. https://doi.org/10.1002/1520-6807

Purkey, W. W. (1970). *Self-concept and school achievement*. Prentice-Hall, Inc.

R Core Team. (2017). *R: A language and environment for statistical computing* (3.60) [Windows]. R Foundation for Statistical Computing.

Raimy, V. C. (1948). Self-reference in counseling interviews. *Journal of Consulting Psychology*, *12*(3), 153–163. https://doi.org/10.1037/h0056600

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2019). *HLM 8 for Windows* (8.0) [Windows]. Scientific Software International, Inc. https://www.hearne.software/getattachment/a104646a-d78e-4ccd-88cf-cf04a78b47f2/HLM8-Manual.aspx

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2019). *HLM 8: Linear and Nonlinear Modeling*. Scientific Software International, Inc. https://www.hearne.software/getattachment/a104646a-d78e-4ccd-88cf-cf04a78b47f2/HLM8-Manual.aspx

Reyes, L. H. (1984). Affective variables and mathematics education. *The Elementary School Journal*, *84*(5), 558–581. https://doi.org/10.1086/461384

Rhee, E., Uleman, J. S., & Lee, H. K. (1996). Variations in collectivism and individualism by ingroup and culture: Confirmatory factor analyses. *Journal of Personality and Social Psychology*, *71*(5), 1037–1054.

Rogers, C. R. (1951). *Client-centered therapy*. Houghton Mifflin Company.

Rosenberg, M. (1989). Self-concept research: A historical overview. *Social Forces*, *68*(1), 34–44. http://dx.doi.org.ezproxylocal.library.nova.edu/10.1093/sf/68.1.34

Rosenberg, M., & Simmons, R. G. (1972). *Black and white self-esteem: The urban school child*. American Sociological Association, Washington, DC.

Ruschenpöhler, L., & Markic, S. (2019). Self-concept research in science and technology education – theoretical foundation, measurement instruments, and main findings. *Studies in Science Education*, *55*(1), 37–68. https://doi.org/10 .1080/03057267. 2019.1645533

Rutkowski, L., Gonzalez, M. J., & Davier, M. V. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish-–little-pond effect across cultures. *Journal of Educational Psychology*, *108*(3), 405–423. https://doi.org/ 10.1037/edu0000063

Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self-concept change. *Review of Educational Research*, *49*(1), 131–150.

Schroeders, U., & Jansen, M. (2020). Science self-concept – More than the sum of its parts? *The Journal of Experimental Education*, 1–17. https://doi.org/ 10.1080 /00220973.2020.1740967

Schwabe, F., Korthals, R., & Schils, T. (2019). Positive social relationships with peers and teachers as moderators of the big-fish-little-pond effect. *Learning and Individual Differences*, *70*, 21–29. https://doi.org/10.1016/j.lindif.2018.12.006

Sears, P. S., & Sherman, V. (1964). *In pursuit of self-esteem: Case studies of eight elementary school children*. Wadesworth Publishing Company.

Seaton, M. (2007). *The big fish little pond effect under the grill: Tests of universality, a search for moderators and the role of social comparison*. University of Western Sydney.

Seaton, M., Marsh, H. W., & Craven, R. (2009a). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 1 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*(2), 403–419.

Seaton, M., Marsh, H. W., & Craven, R. (2010a). Big-fish-little-pond effect: Generalizability and moderation-Two sides of the same coin. *American Educational Research Journal*, *47*(2), 390–433.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009b). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*(2), 403–419. https://doi.org/10.1037/a0013838

Seaton, M., Marsh, H. W., & Craven, R. G. (2010b). Big-fish-little-pond effect: Generalizability and moderation-Two sides of the same coin. *American Educational Research Journal*, *47*(2), 390–433. https://doi.org/ 10.3102 /0002831209350493

Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: Juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology*, *34*(1), 49–72. https://doi.org/ 10.1080/01443410.2013.825232

Shavelson, R. J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology*, *74*(1), 3–17. http://dx.doi.org.ezproxylocal .library.nova.edu/10.1037/0022-0663.74.1.3

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct Interpretations. *Review of Educational Research*, *46*(3), 407–441.

Shavelson, R. J., & Stuart, K. R. (1981). Application of causal modeling methods to the validation of self-concept interpretations of test scores. In M. D. Lynch, K. Gergen, & A. A. Norem-Hebelson (Eds.), *Self-concept: Advances in theory and research* (pp. 223–235). Ballinger Press.

Skaalvik, E. M., & Rankin, R. J. (1992). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology*, *82*, 546–554. https:// doi.org/10.1037/0022-0663.82.3.546

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education; Washington D.C., Wash.*, *67*(2), 135–149. https://doi.org/10.1080/00220979909598349

Snygg, D., & Combs, A. W. (1949). *Individual behavior*. Harper & Row, Publishers.

Soares, A. T., & Soares, L. M. (1969). Self-perceptions of culturally disadvantaged children. *American Educational Research Journal*, *6*, 31–45.

Soares, L. M., & Soares, A. T. (1983). *Components of students' self-related cognitions*. Annual meeting of the American Educational Research Association, Montreal, Canada.

Spearman, C. (1904). "General intelligence," Objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–293. https://doi.org/ 10.2307 /1412107

Stevens, S. S. (1946). On the theory of scales and measurement. *American Association for the Advancement of Science*, *103*(2684), 677–680.

Strein, W., & Grossman, J. (2010). *Big fish and other school effects on academic self-concept* [Poster Session]. Annual Convention of the American Psychological Association, San Diego CA. https://doi.org/10.1037/e630882010-001

Sullivan, G., & Artino Jr., A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education, Inc.

Taijfel, H., & Turner, J. C. (2004). *The social identity theory of intergroup behavior* (J., Ed.). Psychology Press.

The World Bank. (2020). *World Bank Country and Lending Groups FY21*. Data. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Trautwein, U., Gerlach, E., & Ludtke, O. (2008). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of refence effects. *Journal of Educational Psychology*, *100*(4), 988–1001.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506–520.

Trowbridge, N. (1972). Self-concept and socio-economic status in elementary school children. *American Educational Research Journal*, *9*, 525–537. https://doi.org /10.2307/1162274

Tucker-Drob, E. M., Cheung, A. K., & Briley, D. A. (2014). Gross domestic product, science interest, and science achievement: A person × nation interaction. *Psychological Science*, *25*(11), 2047–2057. https://doi.org /10.1177/ 0956797614548726

Turner, J. C., & Reynolds, K. J. (2012). Self-categorization theory. In P. A. Van Lange, A. W. Krunglanski, & T. Higgins (Eds.), *Handbook of Theories of Social Psychology*. SAGE Publications Ltd.

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). National Foundation for the educational research.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

von Davier, M. (2020). Chapter 11: Scaling methodology, item response theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (p. 11.1-11.25). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html

Wang, J. (2007). A trend study of self-concept and mathematics achievement in a cross-cultural context. *Mathematics Education Research Journal*, *19*(3), 33–47. https://doi.org/10.1007/BF03217461

Wang, Z. (2015a). Examining big-fish-little-pond-effects across 49 countries: A multilevel latent variable modelling approach. *Educational Psychology*, *35*(2), 228–251.

Wang, Z. (2015b). Examining big-fish-little-pond-effects across 49 countries: A multilevel latent variable modelling approach. *Educational Psychology*, *35*(2), 228–251. https://doi.org/10.1080/01443410.2013.827155

West, C. K., & Fish, J. A. (1973). Relationships between self-concept and school achievement: A survey of empirical investigations. *ERIC Clearinghouse on Early Childhood Education*, *ED092239*, 1–42.

West, C. K., Fish, J. A., & Stevens, R. J. (1980). General self-concept, self-concept of academic ability and school achievement: Implications for “causes” of self-concept. *The Australian Journal of Education*, *24*(2), 194–213. https://doi.org/10.1177/000494418002400207

Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education*, *72*(4), 331–346. https://doi.org/10.3200/JEXE.72.4.331-346

Willms, D. J. (2010). School composition and contextual effects on student outcomes. *Teacher College Record*, *112*(4), 1008–1037.

Winne, P. H., Marx, R. W., & Taylor, T. D. (1977). A multitrait-multimethod study of three self-concept inventories. *Child Development*, *48*(3), 893–901.

Wouters, S., Colpin, H., Van Damme, J., & Verschueren, K. (2015). Endorsing achievement goals exacerbates the big-fish-little-pond effect on academic self-concept. *Educational Psychology*, *35*(2), 252–270. https://doi.org/ 10.1080/01443410.2013.822963

Wouters, S., De Fraine, B., Colpin, H., Van Damme, J., & Verschueren, K. (2012). The effect of track changes on the development of academic self-concept in high school: A dynamic test of the big-fish–little-pond effect. *Journal of Educational Psychology*, *104*(3), 793–805. https://doi.org/10.1037/a0027732

Wylie, R. C. (1974). *The self-concept*. University of Nebraska Press.

Wylie, R. C. (1979). *The self-concept: Theory and research on selected topics* (Vol. 2). University of Nebraska Press.

Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Educational Research*, *47*(1), 21–41. https://doi.org/10.1080/00313830308609

Yin, L., & Fishbein, B. (2020). Chapter 16: Creating and interpreting 2019 context questionnaire scales. In *Methods and procedures: TIMSS 2019 technical report* (p. 16.1-16.331). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-16.html

Young, D. J. (1998). Rural and urban differences in student achievement in science and mathematics: A multilevel analysis. *School Effectiveness and School Improvement*, *9*(4), 386–418. https://doi.org/10.1080/0924345980090403

Zhang, X., Pomerantz, E., Setoh, P., Qu, Y., & Wang, M. (2016). The role of affect in the positive self: Two longitudinal investigations of young adolescents in the United States and China. *Journal of Personality and Social Psychology*, *111*(1), 83–97.

Zheng, A., Tucker-Drob, E. M., & Briley, D. A. (2019). National gross domestic product, science interest, and science achievement: A direct replication and extension of the Tucker-Drob, Cheung, and Briley (2014) study. *Psychological Science*, *30*(5), 776–788. https://doi.org/10.1177/0956797619835768

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting.

American Educational Research Journal, *29*(3), 663–676. https://doi.org/10.3102/00028312029003663

Zirkel, P. A. (1971). Self-concept and the “disadvantage” of ethnic group membership and mixture. *Review of Educational Research*, *41*(3), 211–225. https://doi.org/10.2307/1169528

Appendix A

Variable Descriptions

Appendix A

Variable Descriptions

Variable ID	Level	Measure	TIMSS 2019 Question Number	TIMSS 2019 Variable Name	Variable Description	TIMSS 2019 Item Scale	Recoded Scale
IDSTUD	L1	scale		IDSTUD	student ID		
TOTWGT	L1	Scale		TOTWGT	Student Total weight		
L1SACH1	SL1	Scale		BSSSCI01	Science Plausible values 01		
L1SACH2	SL1			BSSSCI02	Science Plausible values 02		
L1SACH3	SL1			BSSSCI03	Science Plausible values 03		
L1SACH4	SL1			BSSSCI04	Science Plausible values 04		
L1SACH5	SL1			BSSSCI05	Science Plausible values 05		
L1MACH1	ML1	Scale		BSMMAT01	Math Plausible values 01		
L1MACH2	ML1			BSMMAT02	Math Plausible values 02		
L1MACH3	ML1			BSMMAT03	Math Plausible values 03		
L1MACH4	ML1			BSMMAT04	Math Plausible values 04		
L1MACH5	ML1			BSMMAT05	Math Plausible values 05		
L1GND	L1	Nominal		ITSEX	Gender	1: girl 2: boy	1 > 0 "girl"; 2 > 1 "boy"; 9 > sysmis
L1SES	L1	Ordinal	SQG-04	BSBG04	About how many books are there in your home?	1: 0–10 books; 2: 11–25 books; 3: 26–100 books; 4: 101–200 books; 5: More than 200	
	L1		SQG-05c	BSBG05C	Do you have your own room at home?	1: yes 2: no	
	L1		SQG-05d	BSBG05D	Do you have internet connection at home?	1:yes 2: no	

		Derived	BSDG05S			Compute: BSBG05C +BSBG05D: 4 > 0 "neither own room nor internet connection"; 3 > 1 "either own room or internet connection"; 2 > 2 "both own room and internet connection"
L1	Ordinal	SQG-06A	BSBG06A	What is the highest level of education completed by your mother (or female legal guardian)?	1: Some Primary or Lower secondary or did not go to school; 2: Lower secondary; 3: Upper secondary; 4: Post-secondary, non-tertiary; 5: Short-cycle tertiary; 6: Bachelor’s or equivalent; 7: Postgraduate degree; 8: Don’t know	
L1	Ordinal	SQG-06B	BSBG06	What is the highest level of education completed by your father (or male legal guardian)?	1: Some Primary or Lower secondary or did not go to school; 2: Lower secondary; 3: Upper secondary; 4: Post-secondary, non-tertiary; 5: Short-cycle tertiary; 6: Bachelor’s or equivalent; 7: Postgraduate degree; 8: Don’t know	
		Derived	BSDGEUP		1: Some Primary or Lower secondary or did not go to school; 2: Lower secondary; 3:	COMPUTE: max (BSBG06A, BSBG06B) original scale retained: 8 > 0 "don’t know"; 5 > 5 "university or

						Upper secondary; 4: Post-secondary, non-tertiary; 5: Short-cycle tertiary; 6: Bachelor's or equivalent; 7: Postgraduate degree; 8: Don't know	higher"; 6 > 5 "university or higher; 7 > 5 "university or higher"; 9, 99 > sysmis
	L1	Nominal		COMPOSITE			Compute: BSBG04 +BSDGEDUP +BSDG05S: 1 "more disadvantaged' if BCBG03A <=2 and BCBG03B >=3; 2 "neither more affluent or advantaged; 3 = "more affluent" if BCBG03A >=3 and BCBG03B <=2
L1ATS	SL1	Ordinal	SQIS-22a	BSBS22A	I enjoy learning science	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded: 4 > 1 "Disagree a lot"; 3 > 2 "Disagree a little" ; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-22c	BSBS22C	Science is boring	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	
	SL1		SQIS-22e	BSBS22E	I like science	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1	Ordinal		COMPOSITE			COMPUTE BSBS22A + BSBS22C + BSBS22E: 1 "Do not like science" if L1ATS <= 8.3; 2 "somewhat like science"; 3 "very much like science" if

							L1ATM >= 10.6
L1ATM	ML1	Ordinal	SQM-16a	BSBM16A	I enjoy learning math	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-16c	SBM16C	Math is boring	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	
	ML1		SQM-16e	BSBM16E	I like math	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1	Nominal		COMPOSITE			COMPUTE BSBM16A + BSBM16C + BSBM16E: 1 "Do not like math" if L1ATM <= 9.4; 2 "somewhat like math"; 3 "very much like math" if L1ATM >= 11.4.
L1SSC	SL1	Ordinal	SQIS-24a	BSBS24A	I usually do well in science	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-24b	BSBS24B	science is more difficult for me than many of my classmates	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	
	SL1		SQIS-24c	BSBS24C	science is not one of my strengths	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	

	SL1		SQIS-24d	BSBS24D	I learn things quickly in science	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1	Nominal		COMPOSITE			COMPUTE BSBS24A + BSBS24B + BSBS24C+ BSBS24D: 1 "low science self-concept" if L1SSC <= 8.2; 2 "moderate science self-concept"; 3 "high science self-concept" if L1SSC >= 10.2
L1MSC	ML1	Ordinal	SQM-19a	BSBM19A	I usually do well in math	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-19b	BSBM19B	Math is more difficult for me than many of my classmates	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	
	ML1		SQM-19c	BSBM19C	Math is not one of my strengths	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	
	ML1		SQM-19d	BSBM19D	I learn things quickly in math	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1	Nominal		COMPOSITE			COMPUTE BSBM19A + BSBM19B + BSBM19C+ BSBM19D: 1 "low math self-concept"

							if L1MSC <= 9.5; 2 "moderate math self-concept"; 3 "high math self-concept" if L1MSC >= 12.1
L1VOS	SL1	Ordinal	SQIS-25a	BSBS25A	I think learning science will help me in my daily life	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-25b	BSBS25B	I need science to learn other school subjects	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-25c	BSBS25C	I need to do well in science to get to the university of my choice	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-25f	BSBS25F	It is important to learn about science to get ahead in the world	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-25g	BSBS25G	Learning science will give me more job opportunities when I am an adult	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	SL1		SQIS-25i	BSBS25I	It is important to do well in science	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
		Nominal		COMPOSITE			COMPUTE BSBS25A+

							BSBS25B+ BSBS25C+ BSBS25F+ BSBS25G+ BSBS25I: 1 “do not value science” if L1VOS <=8.5; 2 “somewhat value science” 3 “strongly value science” if L1VOS >=10.6
L1VOM	ML1	Ordinal	SQM-20a	BSBM20A	I think learning math will help me in my daily life	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-20b	BSBM20B	I need math to learn other school subjects	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-20c	BSBM20C	I need to do well in math to get to the university of my choice	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-20f	BSBM20F	It is important to learn about math to get ahead in the world	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-20g	BSBM20G	Learning math will give me more job opportunities when I am an adult	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1		SQM-20i	BSBM20I	It is important to do well in math	1: Agree a lot; 2: Agree a little; 3: Disagree a little; 4: Disagree a lot	Reverse Coded - 1: Disagree a lot; 2: Disagree a little; 3: Agree a little; 4: Agree a lot; 9: sysmis
	ML1	Nominal		COMPOSITE			

					COMPUTE BSBM20A+ BSBM20B+ BSBM20C+ BSBM20F+ BSBM20G+ BSBM20I: 1 “do not value math” if L1VOM <=7.8; 2 “somewhat value math” 3 “strongly value math” if L1VOM >=10.3.
IDSCHL L2SACH1	L2	scale	IDSCHL	School ID	
L2SACH2	SL2	Scale	BSSSCI01	L2 Aggregated value of L1SACH1	
L2SACH3	SL2	Scale	BSSSCI02	L2 Aggregated value of L1SACH2	
L2SACH4	SL2	Scale	BSSSCI03	L2 Aggregated value of L1SACH3	
L2SACH5	SL2	Scale	BSSSCI04	L2 Aggregated value of L1SACH4	
L2MACH1	SL2	Scale	BSSSCI05	L2 Aggregated value of L1SACH5	
L2MACH2	ML2	Scale	BSMMAT01	L2 Aggregated value of L1MACH1	
L2MACH3	ML2	Scale	BSMMAT02	L2 Aggregated value of L1MACH2	
L2MACH4	ML2	Scale	BSMMAT03	L2 Aggregated value of L1MACH3	
L2MACH5	ML2	Scale	BSMMAT04	L2 Aggregated value of L1MACH4	

L2SES	ML2	Scale		BSMMAT05	L2 Aggregated value of L1MACH5		
	L2		ScQ - 03a	BCBG03A	Approximately what percentage of students in your school come from economically disadvantaged homes?	1: 0 to 10%; 2: 11 to 25%; 3: 26 to 50%; 4: More than 50%	
	L2		ScQ - 03b	BCBG03B	Approximately what percentage of students in your school come from economically affluent homes?	1: 0 to 10%; 2: 11 to 25%; 3: 26 to 50%; 4: More than 50%	
			Derived	BCDGSBC	COMPUTE: max (BSBG03A, BSBG03B)	1: more disadvantaged ; 2: neither disadvantaged or affluent; 3: more affluent	
L2LOC	L2	Nominal	ScQ-05B	BCBG05B	GEN\IMMEDIATE AREA OF SCH LOCATION	1: Urban 2: Suburban 3: medium city 4: Small town 5: remote rural	recoded dummy coded (3, 4, 5 > 0"rural," 1 & 2 1 "urban")
L2SSC	SL2	Ordinal			L2 Aggregated value of L1SSC		
L2MSC	ML2	Ordinal			L2 Aggregated value of L1MSC		
L2ATS	SL2	Ordinal			L2 Aggregated value of L1ATS		
L2ATM	ML2	ordinal			L2 Aggregated value of L1ATM		
L2VOS	L2	Ordinal			L2 Aggregated value of L1VOS		

L2VOM	L2	Ordinal			L2 Aggregated value of L1VOM		
L2CLM	L2	Ordinal	SCQ-14a	BCBG14A	Teacher's understandings of school's curricular goals	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14b	BCBG14B	Teacher's degree of success in implementing the school's curriculum	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14c	BCBG14C	Teachers' expectation for student achievement	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14d	BCBG14D	Teachers' ability to inspire students	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14e	BCBG14E	Parental involvement in school activities	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14f	BCBG14F	Parental commitment to ensure that students are ready to learn	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14g	BCBG14G	Parental expectations for student achievement	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14h	BCBG14H	Parental support for student achievement	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14i	BCBG14I	Students' desire to do well in school	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2		SCQ-14j	BCBG14J	Students' ability to reach schools'	1: Very high; 2: High; 3: Medium; 4:	reverse coded- 1: Very low; 2: low; 3: Medium; 4:

					academic goals	Low; 5: Very low	high; 5: Very high
	L2		SCQ-14k	BCBG14K	Students' respect for classmates who excel in school	1: Very high; 2: High; 3: Medium; 4: Low; 5: Very low	reverse coded- 1: Very low; 2: low; 3: Medium; 4: high; 5: Very high
	L2	Nominal		COMPOSITE			Compute BCBG14A+ BCBG14B+ BCBG14C+ BCBG14D+ BCBG14E+ BCBG14F+ BCBG14G+ BCBG14H+ BCBG14I+ BCBG14J+ BCBG14K: 1 "medium emphasis on success" if L2CLM <= 9.6; 2 "high emphasis"; 3 "very high emphasis" if L2CLM >=13.1
IDCNTRY	L3			IDCNTRY	Country ID		
HOUWGT	L3			HOUWGT	House Weight		
L3SACH1	SL3	Scale		BSSSCI01	L3 Aggregated value of L1ACH1		
L3SACH2	SL3	Scale		BSSSCI02	L3 Aggregated value of L1ACH2		
L3SACH3	SL3	Scale		BSSSCI03	L3 Aggregated value of L1ACH3		
L3SACH4	SL3	Scale		BSSSCI04	L3 Aggregated value of L1ACH4		
L3SACH5	SL3	scale		BSSSCI05	L3 Aggregated value of L1ACH5		
L3MACH1	ML3	Scale		BSMMAT01	L3 Aggregated value of L1MACH1		
L3MACH2	ML3	Scale		BSMMAT02			

					L3 Aggregated value of L1MACH2		
L3MACH3	ML3	Scale		BSMMAT03	L3 Aggregated value of L1MACH3		
L3MACH4	ML3	Scale		BSMMAT04	L3 Aggregated value of L1MACH4		
L3MACH5	ML3	Scale		BSMMAT05	L3 Aggregated value of L1MACH5		
L3SSC	SL3	Ordinal			L3 Aggregated value of L1SSC		
L3MSC	ML3	Ordinal			L3 Aggregated value of L1MSC		
L3ATS	SL3	Ordinal			L3 Aggregated value of L1ATS		
L3ATM	ML3	Ordinal			L3 Aggregated value of L1ATM		
L3VOS	SL3	Ordinal			L3 Aggregated value of L1VOS		
L3VOM	ML3	Ordinal			L3 Aggregated value of L1VOM		
L3IPC					income per capita	US dollars	
L3TRK	L3	Nominal	CQG-11A	GEN11A	Does an educational authority in your country (e.g., National Ministry of Education) administer examinations	1: no, tracking assessments for primary/secondary/tertiary 2: tracking assessments for tertiary	1: no, tracking assessments for primary/secondary/tertiary 2: tracking assessments for tertiary only 3: tracking

					that have consequences for individual students, such as entry to a higher school system, entry to a university, and/or exiting or graduating from secondary school? If Yes… Please describe the grades at which the exams are given, the subjects that are assessed, and the purpose of each exam.	only 3: tracking assessments for primary/secondary, and/or tertiary	assessments for primary/secondary, and/or tertiary
L3IDV	L3	Scale	CQG-11B	GEN11B	National IDV scale (Hofsted, 2001)	0-100; 0: extreme collectivism; 100: extreme individualism	

Note. All TIMSS 2019 question numbers, variables names, variable descriptions and scales are published in TIMSS 2019 Codebook.

Appendix B

Data Screening Procedures and Syntax

Appendix B

Data Screening Procedures and Syntax

1. Downloaded variables from IDB Analyzer merge mode

2. Files created by merge mode were trimmed to include only the 65 variables Needed for the analysis (saved as T19G8COMPLETE.data.sav)

3. L3TRK variable was created from variable GEN11A (CQG-11B) available in the curriculum questionnaire (see Exhibit 17: National Policies Regarding Examinations with Consequences for Students as reported by National Research Coordinator). Dichotomous yes/no responses were recalibrated on an ordinal range from 1-3 (1 "no tracking"; 2 "tracking practices for tertiary placement only"; 3 "tracking practices for primary and/or secondary as well as tertiary placement"). ID country was recoded as a new variable whereby country ID values were changed to country's tracking measure.

```
GET
FILE=C:\Users\edu44\OneDrive\Desktop\Mirror\pHd
HPENVY\dissertation\Benchmark 3. report\2019 SPSS
data\IDB merge output\T19G8COMPLETE_raw.sav.
DATASET NAME DataSet1 WINDOW=FRONT.
   RECODE IDCNTRY (36=3) (48=3) (152=2) (158=3) (818=3)
(926=3) (344=3) (364=3) (372=3) (376=3)(380=3) (392=3)
(400=2) (410=2) (414=3) (458=3) (554=2) (578=3) (512=3)
(634=3) (682=2)   (702=3) (710=2) (792=3) (784=2)
(840=1) INTO L3TRK_1.
EXECUTE.
```

4. L3IPC was created from info available in Exhibit: Selected Characteristics of TIMSS 2019 Countries in TIMSS 2019 Encyclopedia. ID country was recoded as a new variable whereby country ID values were changed to country's gross national income per capita in US dollars (see TIMMS 2019 Encyclopedia Introduction Exhibit 1: Selected Characteristics of TIMSS 2019 Countries.

```
RECODE IDCNTRY (36=54910) (48=22110) (152=15010)
(158=25501) (818=2690)
(926=42370) (344=50840)
(364=5420) (372=62210) (376=43290) (380=34460)
(392=41690) (400=4300)
(410=33720)(414=34290)(458=11200) (554=42670)
(578=82500) (512=15330) (634=63410) (682=22850)
(702=59590) (710=6040)(792=9610) (784=43470)
(840=65760) INTO L3IPC_1.
EXECUTE.
```

5. L3IDV was created from Hofstede's individualism measure (https://www.hofstede-insights.com/product/compare-countries/). ID country was recoded as a new variable wherein country ID values were changed to country's level of individualism measures as a value form 1-100 with 100 being

extremely individualistic.

```
RECODE IDCNTRY (36=90) (48=25) (152=23) (158=17)
(818=25) (926=89) (344=25) (372=70) (364=41)
(376=54) (380=76) (392=46) (400=30) (414=25) (458=26)
(554=79) (578=69) (512=25) (634=25) (410=18)(682=25)
(702=20) (710=65) (792=37) (784=25) (840=91) INTO
L3IDV.
EXECUTE.
```

6. Recoded missing values – TRANSFORM > RECODE INTO SAME VARIABLE >
VARIABLES (included all ITEMS except plausible values)

```
RECODE ITSEX BSBG04 BSBG05C BSBG05D BSBG06A BSBG06B
BSBM16A BSBM16C BSBM16E BSBM19A BSBM19B BSBM19C BSBM19D
BSBM20A BSBM20B BSBM20C BSBM20F BSBM20G BSBM20I BSBS22A
BSBS22C BSBS22E BSBS24A BSBS24B BSBS24C BSBS24D BSBS25A
BSBS25B BSBS25C BSBS25F BSBS25G BSBS25I BCBG03A BCBG03B
BCBG05B BCBG14A BCBG14B BCBG14C BCBG14D BCBG14E BCBG14F
BCBG14G BCBG14H BCBG14I BCBG14J BCBG14K(9=SYSMIS)
(99=SYSMIS).
EXECUTE.
RECODE BSBG06A BSBG06B (8=SYSMIS).
EXECUTE.
```

7. Applied ESTIMATION MAXIMIZATION (EM) to raw data to impute missing data as all missing data was initially less than 5% missing (Marsh, 2019, 2020).

```
   DATASET DECLARE  L1SES.
   MVA VARIABLES=BSBG04 BSBG05C BSBG05D BSBG06A BSBG06B
/EM(TOLERANCE=0.001 CONVERGENCE=0.0001 ITERATIONS=25
OUTFILE=L1SES)
```

8. Examined Descriptive Statistics of imputed data

```
    DESCRIPTIVES VARIABLES=BSBS22A BSBS22C BSBS22E
 /STATISTICS=MEAN STDDEV MIN MAX SEMEAN KURTOSIS
 SKEWNESS.
```

9. Computed L1 and L2 Derived Variables (see User Guide Supplement 3 section 2.1 pg. 22-23): **BSDG05S** = BSBG05C + BSBG05D
0 “neither own room nor internet connection IF (BSBG05C = 2 AND; 3 > 1 “either own room or internet connection”; 2 >2 “both own room and internet connection”

```
COMPUTE BSDG05S=BSBG05C+BSBG05D.
IF  (BSBG05C=2 AND BSBG05D=2) BSDG05S=0.
```

```
IF  ((BSBG05C=1 AND BSBG05D=2) OR (BSBG05C=2 AND
BSBG05D=1)) BSDG05S=1.
IF  (BSBG05C=1 AND BSBG05D=1) BSDG05S=2.
EXECUTE.
RECODE BSDG05S (SYSMIS=2).
EXECUTE.
```

BSBGEDUP = max (BSBG06A, BSB06B) original scale retained 1 "Some primary or lower secondary", 2 "lower secondary", 3 "upper secondary", 4 "upper secondary, non-tertiary", 5 "university or higher".

RECODED: 8 > 0 "don't know"; 6 > 5 "university or higher; 7 > 5 "university or higher."

```
RECODE BSBG06A BSBG06B (6=5) (7=5).
EXECUTE.
COMPUTE BSBGEDUP=max (BCBG06A, BCBG06B).
EXECUTE.
```

BCDGSBC (see User Guide Supplement 3 section 2.4 pg. 37) : 1 "more disadvantaged"; 2 "neither disadvantaged nor affluent"; 3 "more affluent"

```
 COMPUTE BCDGSBC = 2.
 IF (BCBG03A <=2 AND BCBG03B >=3) BCDGSBC = 3.
 IF (BCBG03A >=3 AND BCBG03B <=2) BCDGSBC = 1.
 EXECUTE.
```

10. Reverse coded all variables so higher numbers indicated a higher value of the construct 4 > 1. "

```
RECODE BSBM16A BSBM16E BSBM19A BSBM19D BSBM20A BSBM20B
BSBM20C BSBM20F BSBM20G BSBM20I BSBS22A
BSBS22E BSBS24A BSBS24D BSBS25A BSBS25B BSBS25C BSBS25F
BSBS25G BSBS25I (4=1) (3=2) (2=3)
(1=4).
EXECUTE.

RECODE BCBG14A BCBG14B BCBG14C BCBG14D BCBG14E BCBG14F
BCBG14G BCBG14H BCBG14I BCBG14J BCBG14K
(5=1) (4=2) (3=3) (2=4) (1=5).
EXECUTE.
```

11. Renamed ITSEX to L1GND and Dummy coded L1GND (1>0 "girl", 2>1 "boy")

```
RECODE ITSEX (1.9999999999999999 thru Highest=0)
(Lowest thru 2.9999999999999999=1) INTO
L1GND.EXECUTE.
```

12. Renamed BCBG05B to L2LOC dummy code L2LOC (3, 4, 5 > 0 "rural", 1 & 2 >

1

“urban”)

```
RECODE BCBG05B (5 thru Highest=0) (0 thru
1.999999999999999999=1) (2 thru
2.999999999999999999=1)(3 thru 3.999999999999999999=0)
(4 thru 4.999999999999999999=0) INTO L2LOC.
EXECUTE.
```

13. Rename BSMMAT01-05 to L1MACH1-5 and BSSCI01- 05 to L1SACH1-5.

14. Renamed BCDGSBC to L2SES.

15. Examined correlations/covariance of Raw data (including all individual items) with
SPSS 27: ANALYZE – CORRELATE – BIVARIATE.

```
CORRELATIONS
/VARIABLES=BSBG04 BSBG05A BSBG05B BSBG05C BSBG05D
BSBG05E BSBG06A BSBG06B BSBM16A BSBM16C BSBM16E BSBM19A
BSBM19B BSBM19C BSBM19D BSBM20A BSBM20B BSBM20C BSBM20F
BSBM20G BSBM20I BSBS22A BSBS22C BSBS22E BSBS24A BSBS24B
BSBS24C BSBS24D BSBS25A BSBS25B BSBS25C BSBS25F BSBS25G
ITSEX BSMMAT01 BSMMAT02 BSMMAT03 BSMMAT04 BSMMAT05
BSSSCI01 BSSSCI02 BSSSCI03 BSSSCI04 BSSSCI05 BCBG03A
BCBG03B BCBG05B BCBG14A BCBG14B BCBG14C BCBG14D BCBG14E
BCBG14F BCBG14G BCBG14H BCBG14I BCBG14J BCBG14K BSBS25I
/PRINT=TWOTAIL NOSIG FULL
/STATISTICS XPROD
/MISSING=PAIRWISE.
```

16. Conducted Principal Component Analysis (PCA) and Alpha Cronbach’s reliability
analysis in SPSS (TIMSS 2019 Methods and Procedures Technical Report CH.16: Creating Contextual questionnaires scales, pg. 16.168).

```
FACTOR
/VARIABLES BSBM19A BSBM19B BSBM19C BSBM19D
/MISSING LISTWISE
/ANALYSIS BSBM19A BSBM19B BSBM19C BSBM19D
/PRINT UNIVARIATE INITIAL DET KMO EXTRACTION
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/ROTATION NOROTATE
/METHOD=CORRELATION.

 RELIABILITY
/VARIABLES=BSBM19A BSBM19B BSBM19C BSBM19D
```

```
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE CORR COV
/SUMMARY=TOTAL
```

17. Conducted Confirmatory Factor Analysis (CFA) in R to compare with PCA results.

```
  install.packages("lavaan")
  library(lavaan)
  install.packages("haven")
  library(haven)
  install.packages("hmisc")
  library(hmisc)
  install.packages("psych")
  library(psych)
  T19G8IMPUTED <-
read_sav("C:/Users/edu44/OneDrive/Desktop/Mirror/pHd
HPENVY/dissertation/Benchmark 3. report/2019 SPSS
data/T19G8IMPUTED.sav")
  View(T19G8IMPUTED)
  CFA_ModelL1<-'L1SES=~BSBG04+BSDG05S+BSDGEDUP
  L1ATS=~BSBS22A+BSBS22C+BSBS22E
  L1ATM=~BSBM16A+BSBM16C+BSBM16E
  L1SSC=~BSBS24A+BSBS24B+BSBS24C+BSBS24D
  L1MSC=~BSBM19A+BSBM19B+BSBM19C+BSBM19D

L1VOS=~BSBS25A+BSBS25B+BSBS25C+BSBS25F+BSBS25G+BSBS25I

L1VOM=~BSBM20A+BSBM20B+BSBM20C+BSBM20F+BSBM20G+BSBM20I
  L1ATS~~L1ATS
  L1ATM~~L1ATM
  L1SSC~~L1SSC
  L1MSC~~L1MSC
  L1VOS~~L1VOS
  L1VOM~~L1VOM'
  fit<-cfa(CFA_ModelL1, data=T19G8IMPUTED,
estimator="WLSMV")
  summary(fit, fit.measures=TRUE, standardized=TRUE)

  CFA_ModelL2<-'
  L2CLM=~BCBG14A+BCBG14B+BCBG14C+
  BCBG14D+BCBG14E+BCBG14F+BCBG14G+BCBG14H+BCBG14I+
BCBG14J+BCBG14K
  L2SES2=~BCBG03A+BCBG03B
L2CLM~~L2CLM
L2SES2~~L2SES2'
```

```
  fit<-cfa(CFA_ModelL2, data=T19G8IMPUTED,
estimator="WLSMV")
  summary(fit, fit.measures=TRUE, standardized=TRUE)

  CFA_ModelL3<-'
  L1ATS=~BSBS22A+BSBS22C+BSBS22E
  L1ATM=~BSBM16A+BSBM16C+BSBM16E
  L1SSC=~BSBS24A+BSBS24B+BSBS24C+BSBS24D
  L1MSC=~BSBM19A+BSBM19B+BSBM19C+BSBM19D
  L1VOS=~BSBS25A+BSBS25B+BSBS25C+
BSBS25F+BSBS25G+BSBS25I
  L1VOM=~BSBM20A+BSBM20B+BSBM20C+
BSBM20F+BSBM20G+BSBM20I
  L1ATS~~L1ATS
  L1ATM~~L1ATM
  L1SSC~~L1SSC
  L1MSC~~L1MSC
  L1VOS~~L1VOS
  L1VOM~~L1VOM
  L2CLM=~BCBG14A+BCBG14B+BCBG14C+
BCBG14D+BCBG14E+BCBG14F+BCBG14G+BCBG14H+BCBG14I+
BCBG14J+BCBG14K
  L2SES2=~BCBG03A+BCBG03B'
  fit<-cfa(CFA_ModelL3, data=T19G8IMPUTED,
estimator="WLSMV")
  summary(fit, fit.measures=TRUE, standardized=TRUE)
```

18. COMPOSITED L1 and L2 items into single construct scales.

```
COMPUTE L1ATS=BSBS22A + BSBS22C + BSBS22E.
EXECUTE.
COMPUTE L1ATM=BSBM16A + BSBM16C + BSBM16E.
EXECUTE.
COMPUTE L1SSC=BSBS24A+BSBS24B+BSBS24C+BSBS24D.
EXECUTE.
COMPUTE L1MSC=BSBM19A+BSBM19B+BSBM19C+BSBM19D.
EXECUTE.
COMPUTE L1VOS=BSBS25A+ BSBS25B+ BSBS25C+ BSBS25F+
BSBS25G+ BSBS25I.
EXECUTE.
COMPUTE L1VOM=BSBM20A+ BSBM20B+ BSBM20C+ BSBM20F+
BSBM20G+ BSBM20I.
EXECUTE.
COMPUTE L1SES=BSBG04 + BSDGEDUP + BSDG05S.
EXECUTE.
```

```
Compute L2CLM=BCBG14A+ BCBG14B+ BCBG14C+ BCBG14D+
BCBG14E+ BCBG14F+ BCBG14G+ BCBG14H+ BCBG14I+ BCBG14J+
BCBG14K.
EXECUTE.
```

19. AGGREGATED all L1 variables to L2 and L3.

```
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=IDSCHOOL
/L2VOS=MEAN(L1VOS)
/L2SSC=MEAN(L1SSC)
/L2ATS=MEAN(L1ATS)
/L2VOM=MEAN(L1VOM)
/L2MSC=MEAN(L1MSC)
/L2ATM=MEAN(L1ATM)
/L2MACH1=MEAN(L1MACH1)
/L2MACH2=MEAN(L1MACH2)
/L2MACH3=MEAN(L1MACH3)
/L2MACH4=MEAN(L1MACH4)
/L2MACH5=MEAN(L1MACH5)
/L2SACH1=MEAN(L1SACH1)
/L2SACH2=MEAN(L1SACH2)
/L2SACH3=MEAN(L1SACH3)
/L2SACH4=MEAN(L1SACH4)
/L2SACH5=MEAN(L1SACH5).

AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=IDCNTRY
/L3VOS=MEAN(L1VOS)
/L3SSC=MEAN(L1SSC)
/L3ATS=MEAN(L1ATS)
/L3VOM=MEAN(L1VOM)
/L3MSC=MEAN(L1MSC)
/L3ATM=MEAN(L1ATM)
/L3MACH1=MEAN(L1MACH1)
/L3MACH2=MEAN(L1MACH2)
   /L3MACH3=MEAN(L1MACH3)
/L3MACH4=MEAN(L1MACH4)
/L3MACH5=MEAN(L1MACH5)
   /L3SACH1=MEAN(L1SACH1)
/L3SACH2=MEAN(L1SACH2)
   /L3SACH3=MEAN(L1SACH3)
/L3SACH4=MEAN(L1SACH4)
 /L3SACH5=MEAN(L1SACH5).
```

20. Examined DESCRIPTIVE STATISTICS and HISTOGRAMS of all final

variables

that were used for analysis.

```
DESCRIPTIVES VARIABLES=L2SACH1 L2SACH2 L2SACH3 L2SACH4
L2SACH5 L2MACH1 L2MACH2 L2MACH3 L2MACH4
L2MACH5 L2VOS L2SSC L2ATS L2VOM L2MSC L2ATM
/STATISTICS=MEAN STDDEV MIN MAX KURTOSIS SKEWNESS.

EXAMINE VARIABLES=L1MACH1gmc L1MACH2gmc L1MACH3gmc
L1MACH4gmc L1MACH5gmc
/PLOT BOXPLOT HISTOGRAM NPPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES EXTREME
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

21. Identified univariate outliers $z < 3$.

```
SAVE
OUTFILE='C:\Users\edu44\OneDrive\Desktop\Mirror\pHd
HPENVY\dissertation\Benchmark 3. '+'report\2019 SPSS
data\T19G8FINAL.sav'/COMPRESSED.
DESCRIPTIVES VARIABLES=L1SACH1 L1SACH2 L1SACH3 L1SACH4
L1SACH5 L1MACH1 L1MACH2 L1MACH3 L1MACH4 L1MACH5 L1GND
L1SES L1ATS L1ATM L1SSC L1MSC L1VOS L1VOM L2SES L2LOC
L2CLM L2SACH1 L2SACH2 L2SACH3 L2SACH4 L2SACH5 L2MACH1
L2MACH2 L2MACH3 L2MACH4 L2MACH5 L2ATS L2ATM L2SSC L2MSC
L2VOS L2VOM L3SACH1 L3SACH2 L3SACH3 L3SACH4 L3SACH5
L3MACH1 L3MACH2 L3MACH3 L3MACH4 L3MACH5 L3ATS L3ATM
L3SSC L3MSC L3VOS L3VOM L3IPC L3TRK L3IDV
/SAVE
/STATISTICS=MEAN STDDEV MIN MAX SEMEAN KURTOSIS
SKEWNESS.
```

22. Identified multivariate outliers using Mahalanobis distance

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT IDSCHOOL
/METHOD=ENTER L1SACH1 L1SACH2 L1SACH3 L1SACH4 L1SACH5
L1MACH1 L1MACH2 L1MACH3 L1MACH4 L1MACH5 L1GND L1SES
L1ATS L1ATM L1SSC L1MSC L1VOS L1VOM L2SES L2LOC L2CLM
L2SACH1 L2SACH2 L2SACH3 L2SACH4 L2SACH5 L2MACH1 L2MACH2
L2MACH3 L2MACH4 L2MACH5 L2ATS L2ATM L2SSC L2MSC L2VOS
L2VOM L3SACH1 L3SACH2 L3SACH3 L3SACH4 L3SACH5 L3MACH1
L3MACH2 L3MACH3 L3MACH4 L3MACH5 L3ATS L3ATM L3SSC L3MSC
L3VOS L3VOM L3IPC L3TRK L3IDV
/SAVE MAHAL.
```

Note: Some syntax is listed for only 1 of the many variables included in the process.

Appendix C

Confirmatory Factor Analysis Results

Appendix C

Confirmatory Factor Analysis Results

```
T19G8IMPUTED <- read_sav("C:/Users/edu44/OneDrive/Desktop/Mirror/pHd
HPENVY/dissertation2022/Benchmark 3. report/2019 SPSS data/spss 2019 data
prep/T19G8IMPUTED.sav")
> View(T19G8IMPUTED)
```

```
> CFA_ModelL1<-'L1SES=~BSBG04+BSDG05S+BSDGEDUP
+     L1ATS=~BSBS22A+BSBS22C+BSBS22E
+     L1ATM=~BSBM16A+BSBM16C+BSBM16E
+     L1SSC=~BSBS24A+BSBS24B+BSBS24C+BSBS24D
+     L1MSC=~BSBM19A+BSBM19B+BSBM19C+BSBM19D
+     L1VOS=~BSBS25A+BSBS25B+BSBS25C+BSBS25F+BSBS25G+BSBS25I
+     L1VOM=~BSBM20A+BSBM20B+BSBM20C+BSBM20F+BSBM20G+BSBM20I
+     L1ATS~~L1ATS
+     L1ATM~~L1ATM
+     L1SSC~~L1SSC
+     L1MSC~~L1MSC
+     L1VOS~~L1VOS
+     L1VOM~~L1VOM'
> fit<-cfa(CFA_ModelL1, data=T19G8IMPUTED, estimator="WLSMV")
> summary(fit, fit.measures=TRUE, standardized=TRUE)

lavaan 0.6.16 ended normally after 74 iterations

  Estimator                                         DWLS
  Optimization method                             NLMINB
  Number of model parameters                          79

  Number of observations                          169957

Model Test User Model:
                                               Standard      Scaled
  Test Statistic                             186148.642  203568.877
  Degrees of freedom                                356         356
  P-value (Chi-square)                            0.000       0.000
  Scaling correction factor                                   0.915
  Shift parameter                                           156.911
    simple second-order correction

Model Test Baseline Model:

  Test statistic                           4739402.537 1347268.990
  Degrees of freedom                              406         406
  P-value                                       0.000       0.000
  Scaling correction factor                                   3.519

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                   0.961       0.849
  Tucker-Lewis Index (TLI)                      0.955       0.828

  Robust Comparative Fit Index (CFI)                        0.961
  Robust Tucker-Lewis Index (TLI)                           0.955

Root Mean Square Error of Approximation:

  RMSEA                                         0.055       0.058
  90 Percent confidence interval - lower        0.055       0.058
  90 Percent confidence interval - upper        0.056       0.058
  P-value H_0: RMSEA <= 0.050                   0.000       0.000
  P-value H_0: RMSEA >= 0.080                   0.000       0.000

  Robust RMSEA                                              0.055
  90 Percent confidence interval - lower                    0.055
  90 Percent confidence interval - upper                    0.056
  P-value H_0: Robust RMSEA <= 0.050                        0.000
  P-value H_0: Robust RMSEA >= 0.080                        0.000

Standardized Root Mean Square Residual:

  SRMR                                          0.053       0.053

Parameter Estimates:

  Standard errors                             Robust.sem
  Information                                   Expected
  Information saturated (h1) model          Unstructured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L1SES =~
    BSBG04            1.000                               0.664    0.525
    BSDG05S           0.385    0.005   71.636    0.000    0.256    0.418
    BSDGEDUP          0.650    0.009   71.635    0.000    0.432    0.454
  L1ATS =~
    BSBS22A           1.000                               0.766    0.845
    BSBS22C           0.682    0.004  166.104    0.000    0.522    0.518
    BSBS22E           1.101    0.003  375.174    0.000    0.843    0.876
  L1ATM =~
    BSBM16A           1.000                               0.841    0.861
    BSBM16C           0.756    0.004  199.756    0.000    0.636    0.606
    BSBM16E           1.092    0.003  362.633    0.000    0.918    0.876
  L1SSC =~
    BSBS24A           1.000                               0.688    0.779
    BSBS24B           0.506    0.005  107.791    0.000    0.348    0.352
    BSBS24C           0.627    0.005  122.339    0.000    0.431    0.416
    BSBS24D           1.100    0.004  276.939    0.000    0.757    0.818
  L1MSC =~
    BSBM19A           1.000                               0.717    0.773
    BSBM19B           0.572    0.005  113.292    0.000    0.410    0.401
    BSBM19C           0.757    0.005  137.855    0.000    0.542    0.503
    BSBM19D           1.083    0.005  237.572    0.000    0.776    0.812
  L1VOS =~
    BSBS25A           1.000                               0.708    0.797
    BSBS25B           1.011    0.003  348.558    0.000    0.716    0.752
    BSBS25C           1.070    0.003  325.691    0.000    0.757    0.787
    BSBS25F           1.070    0.003  342.987    0.000    0.757    0.811
    BSBS25G           1.051    0.003  325.186    0.000    0.744    0.790
    BSBS25I           0.971    0.003  302.279    0.000    0.687    0.777
  L1VOM =~
    BSBM20A           1.000                               0.668    0.728
    BSBM20B           0.955    0.004  272.071    0.000    0.638    0.692
    BSBM20C           0.969    0.004  233.596    0.000    0.647    0.724
    BSBM20F           1.086    0.004  281.301    0.000    0.726    0.787
    BSBM20G           0.978    0.004  252.392    0.000    0.653    0.741
    BSBM20I           0.938    0.004  244.704    0.000    0.627    0.748

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L1SES ~~
    L1ATS            -0.027    0.002  -14.039    0.000   -0.054   -0.054
    L1ATM            -0.039    0.002  -18.191    0.000   -0.070   -0.070
    L1SSC             0.032    0.002   17.073    0.000    0.069    0.069
    L1MSC             0.082    0.002   40.087    0.000    0.172    0.172
    L1VOS            -0.009    0.002   -5.000    0.000   -0.018   -0.018
    L1VOM            -0.001    0.002   -0.557    0.577   -0.002   -0.002
  L1ATS ~~
    L1ATM             0.212    0.002  102.425    0.000    0.329    0.329
    L1SSC             0.439    0.002  188.142    0.000    0.833    0.833
    L1MSC             0.117    0.002   65.473    0.000    0.212    0.212
    L1VOS             0.375    0.002  171.856    0.000    0.692    0.692
    L1VOM             0.199    0.002  113.262    0.000    0.389    0.389
  L1ATM ~~
    L1SSC             0.157    0.002   83.042    0.000    0.272    0.272
    L1MSC             0.449    0.002  186.023    0.000    0.744    0.744
    L1VOS             0.202    0.002  109.013    0.000    0.340    0.340
    L1VOM             0.321    0.002  156.035    0.000    0.572    0.572
  L1SSC ~~
    L1MSC             0.206    0.002  109.062    0.000    0.417    0.417
    L1VOS             0.309    0.002  161.044    0.000    0.634    0.634
    L1VOM             0.168    0.002  106.868    0.000    0.366    0.366
  L1MSC ~~
    L1VOS             0.143    0.002   89.350    0.000    0.282    0.282
    L1VOM             0.222    0.002  129.926    0.000    0.463    0.463
  L1VOS ~~
    L1VOM             0.285    0.002  149.500    0.000    0.603    0.603

Variances:
```

```
                Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    L1ATS          0.587    0.003  182.737    0.000    1.000    1.000
    L1ATM          0.707    0.003  204.711    0.000    1.000    1.000
    L1SSC          0.474    0.003  152.000    0.000    1.000    1.000
    L1MSC          0.514    0.004  145.978    0.000    1.000    1.000
    L1VOS          0.501    0.003  175.061    0.000    1.000    1.000
    L1VOM          0.446    0.003  150.781    0.000    1.000    1.000
   .BSBG04         1.162    0.007  158.115    0.000    1.162    0.725
   .BSDG05S        0.309    0.001  218.788    0.000    0.309    0.826
   .BSDGEDUP       0.717    0.004  174.215    0.000    0.717    0.794
   .BSBS22A        0.235    0.002  108.712    0.000    0.235    0.286
   .BSBS22C        0.746    0.003  225.361    0.000    0.746    0.732
   .BSBS22E        0.215    0.002   90.903    0.000    0.215    0.232
   .BSBM16A        0.246    0.003   97.638    0.000    0.246    0.258
   .BSBM16C        0.696    0.003  205.612    0.000    0.696    0.632
   .BSBM16E        0.257    0.003   88.944    0.000    0.257    0.233
   .BSBS24A        0.308    0.003  116.560    0.000    0.308    0.394
   .BSBS24B        0.858    0.003  300.506    0.000    0.858    0.876
   .BSBS24C        0.889    0.003  279.494    0.000    0.889    0.827
   .BSBS24D        0.283    0.003  102.339    0.000    0.283    0.330
   .BSBM19A        0.347    0.003  113.882    0.000    0.347    0.403
   .BSBM19B        0.874    0.003  289.664    0.000    0.874    0.839
   .BSBM19C        0.868    0.004  239.669    0.000    0.868    0.747
   .BSBM19D        0.312    0.003   95.636    0.000    0.312    0.341
   .BSBS25A        0.289    0.002  154.758    0.000    0.289    0.366
   .BSBS25B        0.393    0.002  172.818    0.000    0.393    0.434
   .BSBS25C        0.353    0.002  158.310    0.000    0.353    0.381
   .BSBS25F        0.299    0.002  148.466    0.000    0.299    0.343
   .BSBS25G        0.333    0.002  158.466    0.000    0.333    0.376
   .BSBS25I        0.310    0.002  152.579    0.000    0.310    0.396
   .BSBM20A        0.396    0.002  161.339    0.000    0.396    0.470
   .BSBM20B        0.444    0.002  177.577    0.000    0.444    0.522
   .BSBM20C        0.381    0.002  160.260    0.000    0.381    0.477
   .BSBM20F        0.323    0.002  144.045    0.000    0.323    0.380
   .BSBM20G        0.350    0.002  158.395    0.000    0.350    0.451
   .BSBM20I        0.309    0.002  150.199    0.000    0.309    0.440
    L1SES          0.441    0.007   61.870    0.000    1.000    1.000
```

```
CFA_ModelL2 <- '
+ L2CLM=~BCBG14A+BCBG14B+BCBG14C+
+   BCBG14D+BCBG14E+BCBG14F+BCBG14G+BCBG14H+BCBG14I+ BCBG14J+BCBG14K
+ L2SES2=~BCBG03A+BCBG03B
+ L2CLM~~L2CLM
+ L2SES2~~L2SES2
+ '
> fit<-cfa(CFA_ModelL2, data=T19G8IMPUTED, estimator="WLSMV")
> summary(fit, fit.measures=TRUE, standardized=TRUE)
```

```
lavaan 0.6.16 ended normally after 33 iterations

  Estimator                                         DWLS
  Optimization method                             NLMINB
  Number of model parameters                          27

  Number of observations                          169957

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                             61639.154  136042.525
  Degrees of freedom                                64          64
  P-value (Chi-square)                           0.000       0.000
  Scaling correction factor                                  0.453
  Shift parameter                                           22.333
    simple second-order correction

Model Test Baseline Model:

  Test statistic                           2259637.872  682008.338
  Degrees of freedom                                78          78
  P-value                                        0.000       0.000
```

```
  Scaling correction factor                                  3.313

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.973       0.801
  Tucker-Lewis Index (TLI)                       0.967       0.757

  Robust Comparative Fit Index (CFI)                         0.973
  Robust Tucker-Lewis Index (TLI)                            0.967

Root Mean Square Error of Approximation:

  RMSEA                                          0.075       0.112
  90 Percent confidence interval - lower         0.075       0.111
  90 Percent confidence interval - upper         0.076       0.112
  P-value H_0: RMSEA <= 0.050                    0.000       0.000
  P-value H_0: RMSEA >= 0.080                    0.000       1.000

  Robust RMSEA                                               0.075
  90 Percent confidence interval - lower                     0.075
  90 Percent confidence interval - upper                     0.076
  P-value H_0: Robust RMSEA <= 0.050                         0.000
  P-value H_0: Robust RMSEA >= 0.080                         0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.074       0.074

Parameter Estimates:

  Standard errors                           Robust.sem
  Information                                 Expected
  Information saturated (h1) model        Unstructured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L2CLM =~
    BCBG14A           1.000                               0.387    0.577
    BCBG14B           1.163    0.004  310.571    0.000    0.451    0.652
    BCBG14C           1.226    0.005  242.627    0.000    0.475    0.652
    BCBG14D           1.231    0.005  257.357    0.000    0.477    0.644
    BCBG14E           1.893    0.009  212.933    0.000    0.733    0.723
    BCBG14F           2.137    0.009  231.532    0.000    0.828    0.835
    BCBG14G           1.630    0.008  201.596    0.000    0.632    0.713
    BCBG14H           2.043    0.009  227.604    0.000    0.791    0.822
    BCBG14I           1.632    0.007  219.733    0.000    0.632    0.781
    BCBG14J           1.540    0.007  219.929    0.000    0.597    0.777
    BCBG14K           1.413    0.007  199.703    0.000    0.547    0.664
  L2SES2 =~
    BCBG03A           1.000                               0.870    0.763
    BCBG03B          -0.921    0.005 -176.456    0.000   -0.801   -0.717

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L2CLM ~~
    L2SES2            0.186    0.001  146.027    0.000    0.552    0.552

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    L2CLM             0.150    0.001  126.284    0.000    1.000    1.000
    L2SES2            0.757    0.005  148.181    0.000    1.000    1.000
   .BCBG14A           0.300    0.001  225.952    0.000    0.300    0.667
   .BCBG14B           0.275    0.001  248.004    0.000    0.275    0.575
   .BCBG14C           0.306    0.001  231.153    0.000    0.306    0.575
   .BCBG14D           0.320    0.001  241.103    0.000    0.320    0.585
   .BCBG14E           0.490    0.002  217.063    0.000    0.490    0.477
   .BCBG14F           0.297    0.002  186.191    0.000    0.297    0.303
   .BCBG14G           0.385    0.002  205.001    0.000    0.385    0.491
   .BCBG14H           0.301    0.001  202.652    0.000    0.301    0.324
   .BCBG14I           0.255    0.001  207.078    0.000    0.255    0.389
   .BCBG14J           0.233    0.001  206.085    0.000    0.233    0.396
```

```
  .BCBG14K           0.380    0.002  225.381    0.000    0.380    0.559
  .BCBG03A           0.545    0.005  115.909    0.000    0.545    0.419
  .BCBG03B           0.608    0.004  145.660    0.000    0.608    0.486
```

```
CFA_ModelL3<-'
+ L1ATS=~BSBS22A+BSBS22C+BSBS22E
+ L1ATM=~BSBM16A+BSBM16C+BSBM16E
+ L1SSC=~BSBS24A+BSBS24B+BSBS24C+BSBS24D
+ L1MSC=~BSBM19A+BSBM19B+BSBM19C+BSBM19D
+ L1VOS=~BSBS25A+BSBS25B+BSBS25C+ BSBS25F+BSBS25G+BSBS25I
+ L1VOM=~BSBM20A+BSBM20B+BSBM20C+ BSBM20F+BSBM20G+BSBM20I
+ L1ATS~~L1ATS
+ L1ATM~~L1ATM
+ L1SSC~~L1SSC
+ L1MSC~~L1MSC
+ L1VOS~~L1VOS
+ L1VOM~~L1VOM
+ L2CLM=~BCBG14A+BCBG14B+BCBG14C+ BCBG14D+BCBG14E+BCBG14F+BCBG14G+BCBG14H+BCBG14I+ BCBG14J+BCBG14K
+ L2SES2=~BCBG03A+BCBG03B'
> fit<-cfa(CFA_ModelL3, data=T19G8IMPUTED, estimator="WLSMV")
> summary(fit, fit.measures=TRUE, standardized=TRUE)
```

```
lavaan 0.6.16 ended normally after 96 iterations

  Estimator                                         DWLS
  Optimization method                             NLMINB
  Number of model parameters                         106

  Number of observations                          169957

Model Test User Model:
                                                Standard      Scaled
  Test Statistic                              259006.007  235846.686
  Degrees of freedom                                 674         674
  P-value (Chi-square)                             0.000       0.000
  Scaling correction factor                                    1.100
  Shift parameter                                            346.524
    simple second-order correction

Model Test Baseline Model:

  Test statistic                            7017364.492 1898673.901
  Degrees of freedom                                741         741
  P-value                                         0.000       0.000
  Scaling correction factor                                   3.697

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                     0.963       0.876
  Tucker-Lewis Index (TLI)                        0.960       0.864

  Robust Comparative Fit Index (CFI)                          0.963
  Robust Tucker-Lewis Index (TLI)                             0.959

Root Mean Square Error of Approximation:

  RMSEA                                           0.047       0.045
  90 Percent confidence interval - lower          0.047       0.045
  90 Percent confidence interval - upper          0.048       0.045
  P-value H_0: RMSEA <= 0.050                     1.000       1.000
  P-value H_0: RMSEA >= 0.080                     0.000       0.000

  Robust RMSEA                                                0.048
  90 Percent confidence interval - lower                      0.047
  90 Percent confidence interval - upper                      0.048
  P-value H_0: Robust RMSEA <= 0.050                          1.000
  P-value H_0: Robust RMSEA >= 0.080                          0.000

Standardized Root Mean Square Residual:
```

```
  SRMR                                           0.048       0.048

Parameter Estimates:

  Standard errors                             Robust.sem
  Information                                   Expected
  Information saturated (h1) model          Unstructured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L1ATS =~
    BSBS22A           1.000                               0.765    0.845
    BSBS22C           0.683    0.004  166.216    0.000    0.523    0.518
    BSBS22E           1.102    0.003  374.405    0.000    0.843    0.876
  L1ATM =~
    BSBM16A           1.000                               0.841    0.861
    BSBM16C           0.758    0.004  200.607    0.000    0.637    0.607
    BSBM16E           1.090    0.003  364.216    0.000    0.917    0.874
  L1SSC =~
    BSBS24A           1.000                               0.687    0.778
    BSBS24B           0.511    0.005  108.632    0.000    0.351    0.355
    BSBS24C           0.630    0.005  122.726    0.000    0.433    0.418
    BSBS24D           1.099    0.004  276.370    0.000    0.756    0.817
  L1MSC =~
    BSBM19A           1.000                               0.716    0.772
    BSBM19B           0.580    0.005  114.752    0.000    0.416    0.407
    BSBM19C           0.760    0.005  138.440    0.000    0.544    0.505
    BSBM19D           1.078    0.005  237.540    0.000    0.772    0.808
  L1VOS =~
    BSBS25A           1.000                               0.706    0.795
    BSBS25B           1.012    0.003  347.440    0.000    0.715    0.751
    BSBS25C           1.073    0.003  324.909    0.000    0.758    0.787
    BSBS25F           1.074    0.003  342.230    0.000    0.758    0.812
    BSBS25G           1.055    0.003  324.794    0.000    0.745    0.791
    BSBS25I           0.974    0.003  301.695    0.000    0.688    0.778
  L1VOM =~
    BSBM20A           1.000                               0.669    0.729
    BSBM20B           0.953    0.003  272.617    0.000    0.638    0.692
    BSBM20C           0.966    0.004  233.862    0.000    0.647    0.723
    BSBM20F           1.085    0.004  282.070    0.000    0.726    0.788
    BSBM20G           0.975    0.004  252.688    0.000    0.653    0.740
    BSBM20I           0.935    0.004  244.952    0.000    0.625    0.747
  L2CLM =~
    BCBG14A           1.000                               0.389    0.579
    BCBG14B           1.164    0.004  310.713    0.000    0.452    0.654
    BCBG14C           1.218    0.005  242.325    0.000    0.473    0.649
    BCBG14D           1.230    0.005  257.600    0.000    0.478    0.646
    BCBG14E           1.881    0.009  212.839    0.000    0.731    0.721
    BCBG14F           2.130    0.009  231.943    0.000    0.828    0.835
    BCBG14G           1.618    0.008  201.735    0.000    0.629    0.711
    BCBG14H           2.029    0.009  227.857    0.000    0.789    0.819
    BCBG14I           1.628    0.007  220.378    0.000    0.633    0.782
    BCBG14J           1.537    0.007  220.435    0.000    0.597    0.778
    BCBG14K           1.415    0.007  200.865    0.000    0.550    0.667
  L2SES2 =~
    BCBG03A           1.000                               0.882    0.773
    BCBG03B          -0.896    0.005 -180.002    0.000   -0.790   -0.707

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  L1ATS ~~
    L1ATM             0.212    0.002  102.439    0.000    0.330    0.330
    L1SSC             0.439    0.002  187.766    0.000    0.834    0.834
    L1MSC             0.116    0.002   65.348    0.000    0.212    0.212
    L1VOS             0.374    0.002  171.430    0.000    0.692    0.692
    L1VOM             0.199    0.002  113.338    0.000    0.389    0.389
    L2CLM            -0.004    0.001   -5.157    0.000   -0.014   -0.014
    L2SES2            0.044    0.002   20.993    0.000    0.065    0.065
  L1ATM ~~
    L1SSC             0.157    0.002   82.967    0.000    0.272    0.272
```

```
    L1MSC             0.449    0.002  186.073    0.000    0.745    0.745
    L1VOS             0.202    0.002  108.959    0.000    0.340    0.340
    L1VOM             0.322    0.002  156.423    0.000    0.573    0.573
    L2CLM             0.007    0.001    8.399    0.000    0.023    0.023
    L2SES2            0.087    0.002   37.802    0.000    0.117    0.117
  L1SSC ~~
    L1MSC             0.206    0.002  108.992    0.000    0.418    0.418
    L1VOS             0.308    0.002  160.542    0.000    0.634    0.634
    L1VOM             0.168    0.002  106.838    0.000    0.366    0.366
    L2CLM            -0.015    0.001  -19.962    0.000   -0.057   -0.057
    L2SES2            0.004    0.002    1.878    0.060    0.006    0.006
  L1MSC ~~
    L1VOS             0.143    0.002   89.165    0.000    0.282    0.282
    L1VOM             0.222    0.002  129.930    0.000    0.463    0.463
    L2CLM            -0.025    0.001  -31.938    0.000   -0.091   -0.091
    L2SES2           -0.028    0.002  -13.592    0.000   -0.044   -0.044
  L1VOS ~~
    L1VOM             0.285    0.002  149.521    0.000    0.603    0.603
    L2CLM            -0.012    0.001  -17.294    0.000   -0.045   -0.045
    L2SES2            0.030    0.002   16.141    0.000    0.048    0.048
  L1VOM ~~
    L2CLM             0.001    0.001    1.599    0.110    0.004    0.004
    L2SES2            0.048    0.002   26.579    0.000    0.081    0.081
  L2CLM ~~
    L2SES2            0.189    0.001  148.091    0.000    0.551    0.551

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    L1ATS             0.586    0.003  182.421    0.000    1.000    1.000
    L1ATM             0.707    0.003  205.284    0.000    1.000    1.000
    L1SSC             0.473    0.003  151.570    0.000    1.000    1.000
    L1MSC             0.513    0.004  145.821    0.000    1.000    1.000
    L1VOS             0.499    0.003  174.309    0.000    1.000    1.000
    L1VOM             0.448    0.003  151.335    0.000    1.000    1.000
   .BSBS22A           0.235    0.002  109.066    0.000    0.235    0.287
   .BSBS22C           0.745    0.003  225.201    0.000    0.745    0.732
   .BSBS22E           0.215    0.002   90.889    0.000    0.215    0.232
   .BSBM16A           0.246    0.003   97.852    0.000    0.246    0.258
   .BSBM16C           0.694    0.003  205.244    0.000    0.694    0.631
   .BSBM16E           0.259    0.003   90.074    0.000    0.259    0.235
   .BSBS24A           0.309    0.003  116.956    0.000    0.309    0.395
   .BSBS24B           0.856    0.003  299.396    0.000    0.856    0.874
   .BSBS24C           0.888    0.003  278.868    0.000    0.888    0.826
   .BSBS24D           0.285    0.003  103.308    0.000    0.285    0.333
   .BSBM19A           0.347    0.003  114.144    0.000    0.347    0.403
   .BSBM19B           0.869    0.003  286.699    0.000    0.869    0.834
   .BSBM19C           0.866    0.004  239.039    0.000    0.866    0.745
   .BSBM19D           0.318    0.003   97.962    0.000    0.318    0.348
   .BSBS25A           0.291    0.002  155.764    0.000    0.291    0.368
   .BSBS25B           0.395    0.002  173.618    0.000    0.395    0.436
   .BSBS25C           0.353    0.002  158.205    0.000    0.353    0.381
   .BSBS25F           0.298    0.002  148.160    0.000    0.298    0.341
   .BSBS25G           0.331    0.002  157.971    0.000    0.331    0.374
   .BSBS25I           0.308    0.002  152.183    0.000    0.308    0.395
   .BSBM20A           0.394    0.002  160.821    0.000    0.394    0.468
   .BSBM20B           0.443    0.002  177.330    0.000    0.443    0.521
   .BSBM20C           0.382    0.002  160.483    0.000    0.382    0.477
   .BSBM20F           0.322    0.002  143.793    0.000    0.322    0.380
   .BSBM20G           0.351    0.002  158.756    0.000    0.351    0.452
   .BSBM20I           0.310    0.002  150.805    0.000    0.310    0.442
   .BCBG14A           0.299    0.001  225.522    0.000    0.299    0.665
   .BCBG14B           0.274    0.001  247.359    0.000    0.274    0.572
   .BCBG14C           0.307    0.001  231.152    0.000    0.307    0.578
   .BCBG14D           0.319    0.001  240.059    0.000    0.319    0.583
   .BCBG14E           0.493    0.002  217.438    0.000    0.493    0.480
   .BCBG14F           0.297    0.002  185.564    0.000    0.297    0.302
   .BCBG14G           0.388    0.002  205.997    0.000    0.388    0.495
   .BCBG14H           0.305    0.002  202.604    0.000    0.305    0.329
   .BCBG14I           0.254    0.001  205.774    0.000    0.254    0.388
   .BCBG14J           0.232    0.001  205.197    0.000    0.232    0.394
```

.BCBG14K	0.378	0.002	223.676	0.000	0.378	0.555
.BCBG03A	0.524	0.005	110.994	0.000	0.524	0.402
.BCBG03B	0.626	0.004	155.180	0.000	0.626	0.501
L2CLM	0.151	0.001	126.735	0.000	1.000	1.000
L2SES2	0.779	0.005	151.740	0.000	1.000	1.000

Appendix D

Final Variable

Descriptive Statistics and Distributions

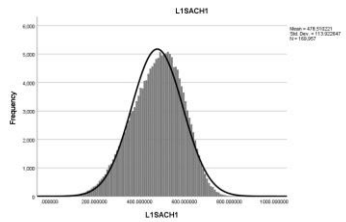
Appendix D
Final Variable
Descriptive Statistics and Distributions

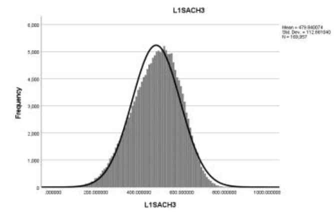
Initial Final Variable Descriptives

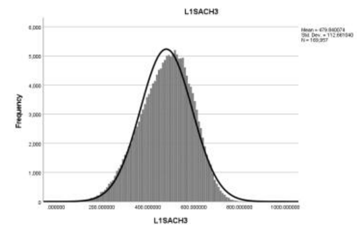
Variable Name	*N*	% Missing	Min	Max	Mean	SEM	Std. Dev.	Skewness	Kurtosis
L1SACH1	169957	0.00%	5	863	478.51	0.280	113.92	-0.29	-0.22
L1SACH2	169957	0.00%	5	873	478.66	0.280	113.43	-0.29	-0.21
L1SACH3	169957	0.00%	12	858	479.84	0.270	112.66	-0.29	-0.21
L1SACH4	169957	0.00%	5	851	478.14	0.280	114.16	-0.28	-0.22
L1SACH5	169957	0.00%	5	866	479.07	0.280	113.86	-0.29	-0.22
L1MACH1	169957	0.00%	60	906	477.48	0.270	109.68	0.13	-0.36
L1MACH2	169957	0.00%	5	903	477.98	0.270	110.40	0.13	-0.35
L1MACH3	169957	0.00%	5	902	477.97	0.270	111.04	0.13	-0.35
L1MACH4	169957	0.00%	11	923	476.93	0.270	111.63	0.13	-0.36
L1MACH5	169957	0.00%	72	911	477.82	0.270	111.17	0.13	-0.36
L1GND	169957	0.00%	0	1	0.50	0.000	0.50	-0.01	-2.00
L1SES	169957	0.00%	2	12	8.22	0.000	2.00	-0.17	-0.29
L1ATS	169957	0.00%	3	12	9.26	0.010	2.40	-0.60	-0.34
L1ATM	169957	0.00%	3	12	8.49	0.010	2.63	-0.37	-0.72
L1SSC	169957	0.00%	4	16	11.54	0.010	2.80	-0.10	-0.46
L1MSC	169957	0.00%	4	16	10.84	0.010	2.97	-0.04	-0.49
L1VOS	169957	0.00%	6	24	19.10	0.010	4.59	-0.80	-0.11
L1VOM	169957	0.00%	6	24	19.56	0.010	4.23	-1.01	0.53
L2SACH1	169957	0.00%	230	620	478.51	0.090	36.57	-2.09	7.49
L2SACH2	169957	0.00%	231	616	478.66	0.090	36.31	-2.08	7.45
L2SACH3	169957	0.00%	235	619	479.84	0.090	35.86	-2.07	7.34
L2SACH4	169957	0.00%	224	620	478.14	0.090	36.42	-2.07	7.50
L2SACH5	169957	0.00%	221	615	479.07	0.090	36.45	-2.12	7.71
L2MACH1	169957	0.00%	317	606	477.48	0.070	30.63	-1.50	4.51
L2MACH2	169957	0.00%	305	609	477.98	0.070	30.79	-1.50	4.50
L2MACH3	169957	0.00%	312	610	477.97	0.070	30.88	-1.46	4.31
L2MACH4	169957	0.00%	310	609	476.93	0.080	30.96	-1.49	4.41
L2MACH5	169957	0.00%	311	604	477.82	0.080	30.93	-1.52	4.53
L2ATS	169957	0.00%	6	12	9.26	0.000	0.37	0.60	6.46
L2ATM	169957	0.00%	6	12	8.49	0.000	0.40	1.88	9.11
L2SSC	169957	0.00%	10	16	11.54	0.000	0.38	1.30	9.31
L2MSC	169957	0.00%	9	16	10.84	0.000	0.35	1.61	15.54
L2VOM	169957	0.00%	15	24	19.56	0.000	0.67	1.38	5.40
L2VOS	169957	0.00%	15	24	19.10	0.000	0.70	1.43	5.44
L2SES	169957	0.00%	1	3	1.89	0.000	0.79	0.20	-1.39
L2LOC	169957	0.00%	0	1	0.77	0.000	0.42	-1.30	-0.30
L2CLM	169957	0.00%	11	48	25.78	0.020	6.85	0.02	-0.18
L3SACH1	169957	0.00%	385	601	478.51	0.137	56.51	-0.01	-0.54
L3SACH2	169957	0.00%	386	601	478.66	0.136	56.17	0.00	-0.53
L3SACH3	169957	0.00%	389	602	479.84	0.135	55.55	0.02	-0.53
L3SACH4	169957	0.00%	386	602	478.14	0.137	56.35	0.01	-0.53
L3SACH5	169957	0.00%	386	603	479.07	0.137	56.46	-0.01	-0.52
L3MACH1	169957	0.00%	401	609	477.48	0.152	62.73	0.56	-0.51
L3MACH2	169957	0.00%	401	612	477.98	0.153	63.15	0.57	-0.51
L3MACH3	169957	0.00%	400	611	477.97	0.154	63.44	0.55	-0.51
L3MACH4	169957	0.00%	399	611	476.93	0.155	63.80	0.55	-0.52
L3MACH5	169957	0.00%	400	611	477.82	0.154	63.66	0.55	-0.52
L3ATS	169957	0.00%	8	10	9.26	0.002	0.65	-0.40	-0.55
L3ATM	169957	0.00%	7	10	8.49	0.002	0.69	0.16	-1.40
L3SSC	169957	0.00%	10	13	11.54	0.002	0.75	-0.94	0.34
L3MSC	169957	0.00%	9	12	10.84	0.002	0.62	-0.80	0.04
L3VOS	169957	0.00%	16	21	19.10	0.003	1.40	-0.60	-1.00
L3VOM	169957	0.00%	16	21	19.56	0.003	1.17	-0.67	0.98
L3IPC	169957	0.00%	2690	82500	31822.98	53.631	22109.93	0.28	-1.02
L3TRK	169957	0.00%	1	3	2.49	0.001	0.59	-0.68	-0.50
L3IDV	169957	0.00%	17	91	45.00	0.061	25.16	0.65	-1.16

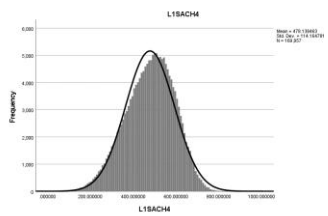
Notes. Results of descriptive statistics for initial final variables shown before univariate outliers were removed (N =169,957).

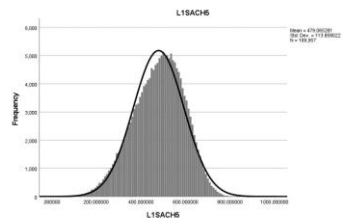
Initial Final Variable Distributions (Student-Level)







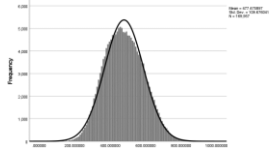


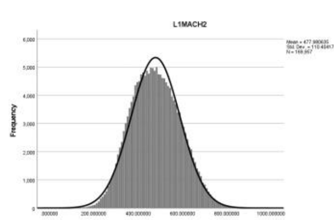


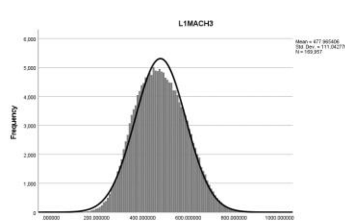
Correlations

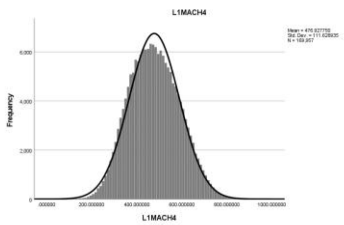
		L1SACH1	L1SACH2	L1SACH3	L1SACH4	L1SACH5
L1SACH1	Pearson Correlation	1	.923**	.923**	.922**	.922**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
	N	169810	169810	169810	169810	169810
L1SACH2	Pearson Correlation	.923**	1	.923**	.922**	.923**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001
	N	169810	169810	169810	169810	169810
L1SACH3	Pearson Correlation	.923**	.923**	1	.922**	.923**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001
	N	169810	169810	169810	169810	169810
L1SACH4	Pearson Correlation	.922**	.922**	.922**	1	.922**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	169810	169810	169810	169810	169810
L1SACH5	Pearson Correlation	.922**	.923**	.923**	.922**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	
	N	169810	169810	169810	169810	169810

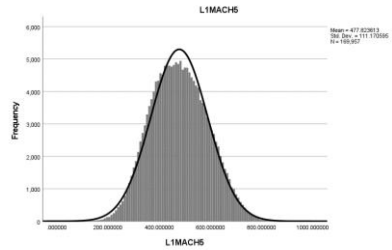
**. Correlation is significant at the 0.01 level (2-tailed).







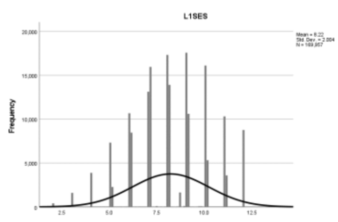


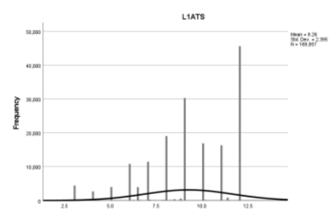


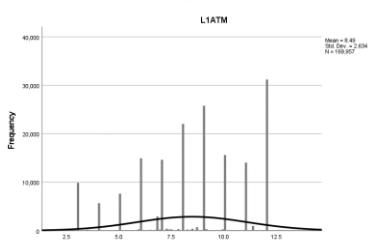
Correlations

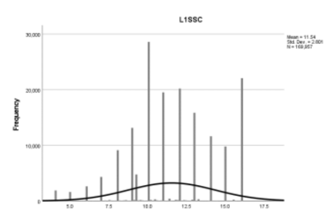
		L1MACH1	L1MACH2	L1MACH3	L1MACH4	L1MACH5
L1MACH1	Pearson Correlation	1	.935**	.935**	.935**	.934**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
	N	169810	169810	169810	169810	169810
L1MACH2	Pearson Correlation	.935**	1	.934**	.935**	.934**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001
	N	169810	169810	169810	169810	169810
L1MACH3	Pearson Correlation	.935**	.934**	1	.934**	.935**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001
	N	169810	169810	169810	169810	169810
L1MACH4	Pearson Correlation	.935**	.935**	.934**	1	.934**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	169810	169810	169810	169810	169810
L1MACH5	Pearson Correlation	.934**	.934**	.935**	.934**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	
	N	169810	169810	169810	169810	169810

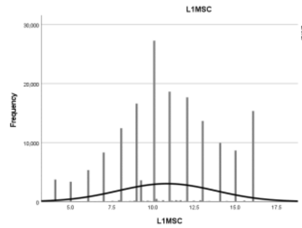
**. Correlation is significant at the 0.01 level (2-tailed).

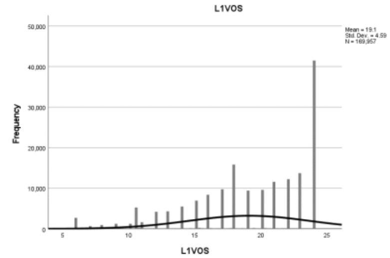


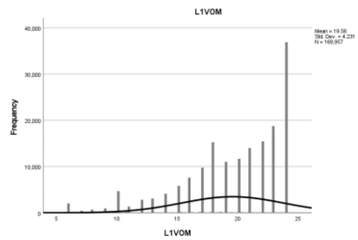




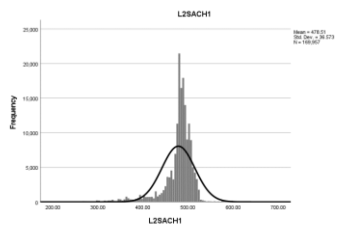


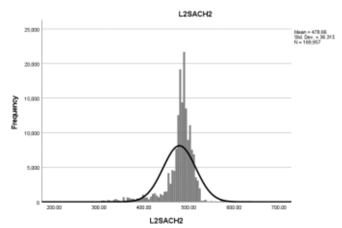


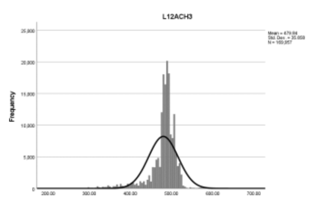


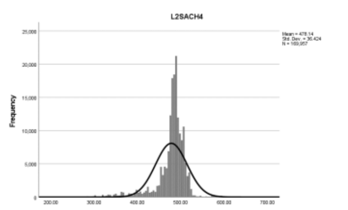


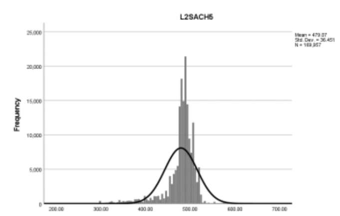
Initial Final Variable Distributions (School-Level)

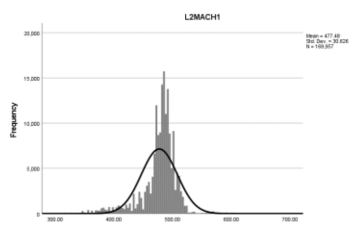


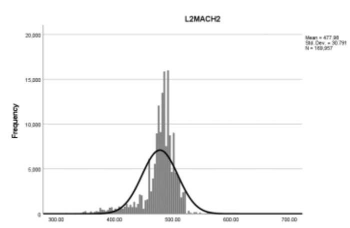


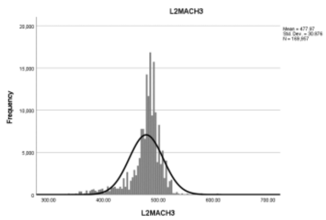


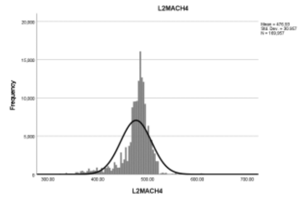


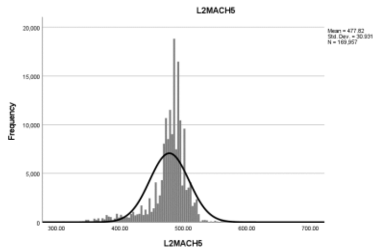


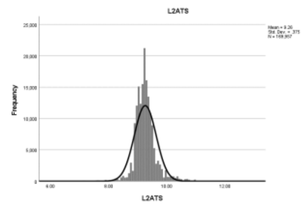


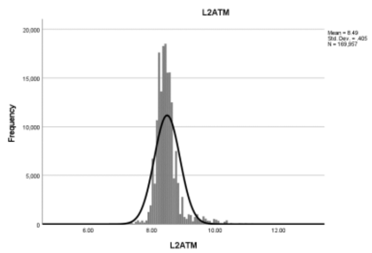


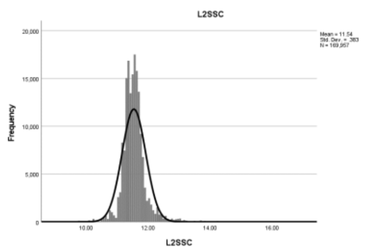


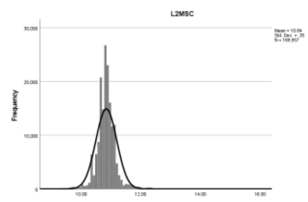


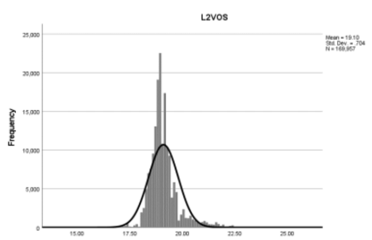


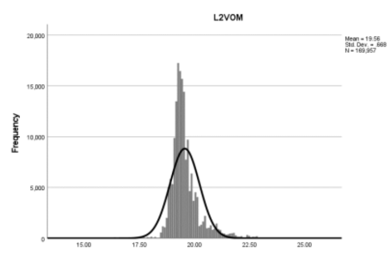




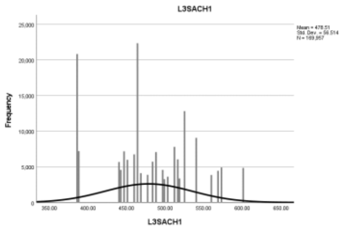


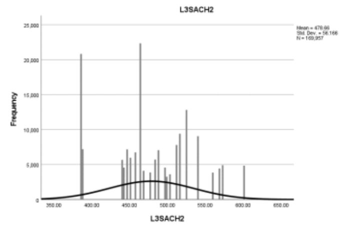


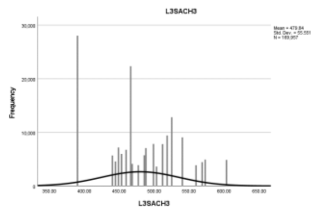


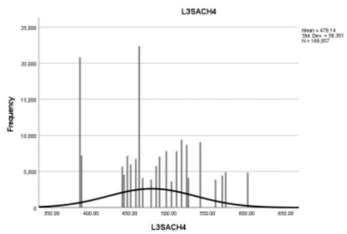


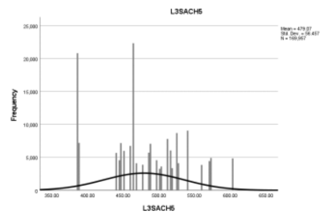
Initial Final Variable Distributions (Country-Level)

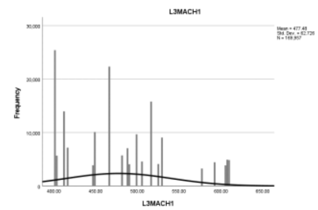

L3SACH1
Frequency
L3SACH1

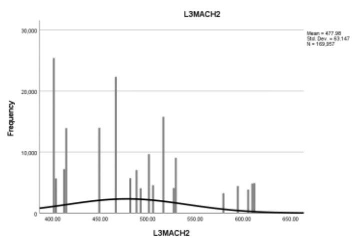

L3SACH2
Frequency
L3SACH2

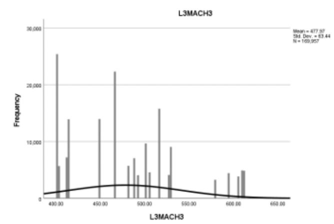

L3SACH3
Frequency
L3SACH3

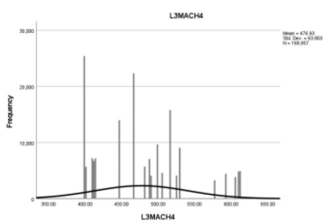

L3SACH4
Frequency
L3SACH4

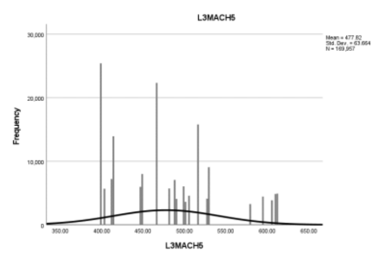

L3SACH5
Frequency
L3SACH5

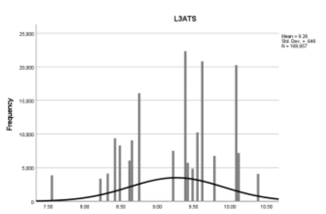

L3MACH1
Frequency
L3MACH1

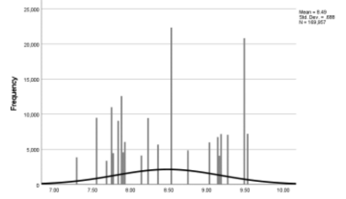

L3MACH2
Frequency
L3MACH2

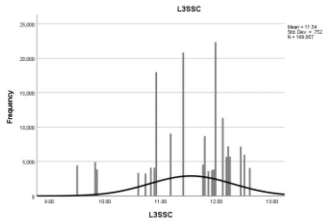

L3MACH3
Frequency
L3MACH3

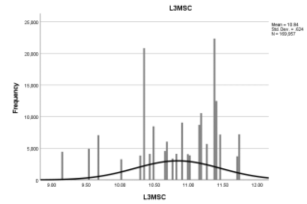

L3MACH4
Frequency
L3MACH4

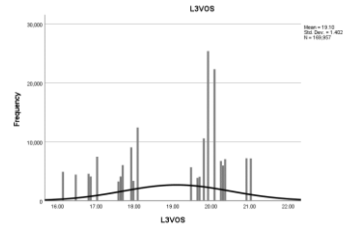

L3MACH5
Frequency
L3MACH5

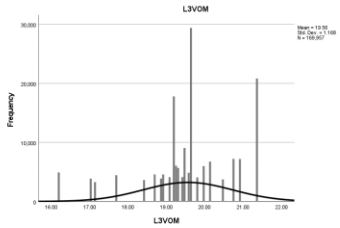

L3ATS
Frequency
L3ATS

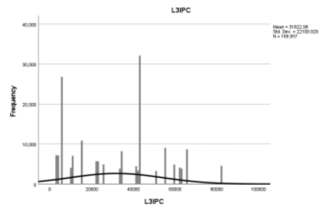

L3ATM
Frequency
L3ATM

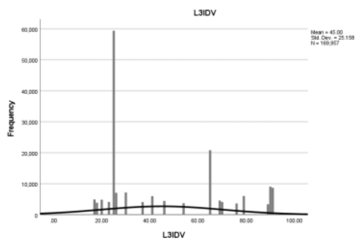

L3SSC
Frequency
L3SSC


L3MSC
Frequency
L3MSC


L3VOS
Frequency
L3VOS


L3VOM
Frequency
L3VOM


L3IPC
Frequency
L3IPC


L3IDV
Frequency
L3IDV

Appendix E

All Results for L2BFLPE Moderation Effects in Math and Science

Appendix E
All Results for L2BFLPE Moderation Effects in Math and Science

L2BFLPE Moderation Effects of Student-Level (L1) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 1 L1 INT			
FIXED EFFECTS	**β**	***SE***	**pvalue**	**-2LL**
L1NULL				849877.02
L1ATM1xL2MACH	0.0003	0.0002	0.080	741547.59
L1GNDxL2MACH	0.0010	0.0009	0.150	44435.05
L1VOM1xL2MACH	0.0005**	0.0002**	0.030**	56419.47
L1SESxL2MACH	0.0008	0.0006	0.214	43181.07
L1MACH1xL2MACH	0.00001	0.00001	0.111	42907.96
L1ATM moderator	0.622*	0.025*	0.000*	
L1GND moderator	-0.402*	0.088*	0.000*	
L1VOM moderator	0.182*	0.011*	0.000*	
L1SES moderator	0.016	0.017	0.342	
L1MACH moderator	0.018*	0.001*	0.000*	
L1MACH (L1ATM)	0.011*	0.0005*	0.000*	
L1MACH (L1GND)	0.018*	0.001*	0.000*	
L1MACH (L1VOM)	0.016*	0.001*	0.000*	
L1MACH (L1SES)	0.018*	0.001*	0.000*	
L2MACH (L1ATM)	-0.003*	0.001*	0.001*	
L2MACH (L1GND)	-0.006*	0.002*	0.011*	
L2MACH (L1VOM)	-0.005*	0.001*	0.000*	
L2MACH (L1SES)	-0.006*	0.002*	0.001*	
L2MACH (L1MACH)	-0.006**	0.002**	0.011**	
RANDOM EFFECTS	**β**	***SD***	**pvalue**	
L1 Res Var NULL	8.405	2.899		
L1 Res Var L1ATM	4.355	2.087		
L1 Res Var L1GND	6.225	3.244		
L1 Res Var L1VOM	5.818	2.412		
L1 Res Var L1SES	6.285	2.507		
L1 Res Var L1MACH	6.331	2.516		
L2 Res Var NULL	0.537*	0.733*	0.000*	
L2 Res Var L1ATM	0.197*	0.444*	0.000*	
L2 Res Var L1GND	0.738*	0.859*	0.000*	
L2 Res Var L1VOM	0.582*	0.763*	0.000*	
L2 Res Var L1SES	0.757*	0.870*	0.000*	
L2 Res Var L1MACH	0.770*	0.878*	0.000*	
L2 Res Var ATM slope	0.011*	0.107*	0.000*	
L2 Res Var L1GND slope	0.320*	0.565*	0.000*	
L2 Res Var L1VOM slope	0.005*	0.068*	0.000*	
L2 Res Var L1SES slope	0.012*	0.112*	0.000*	
L2 Res Var L1MACH slope	0.00002*	0.005*	0.000*	
L2 Res Var L1MACH (L1ATM)	0.00001*	0.003*	0.000*	
L2 Res Var L1MACH (L1GND)	0.00002*	0.004*	0.000*	
L2 Res Var L1MACH (L1VOM)	0.00002*	0.004*	0.000*	
L2 Res Var L1MACH (L1SES)	0.00002*	0.004*	0.000*	
L2 Res Var L1MACH (L1MACH)	0.00002*	0.004*	0.000*	
L3 Res Var Int Null	0.537*	0.733*	0.000*	
L3 Res Var L1ATM	1.025*	1.013*	0.000*	
L3 Res Var L1GND	2.462*	1.569*	0.000*	
L3 Res Var L1VOM	1.805*	1.343*	0.000*	

L3 Res Var L1SES	1.606*	2.580*	0.000*
L3 Res Var L1MACH	2.516*	1.586*	0.000*
L3 Res Var L1MACH (L1ATM)	0.00001*	0.002*	0.000*
L3 Res Var L1MACH (L1GND)	0.00002*	0.004*	0.000*
L3 Res Var L1MACH (L1VOM)	0.00002*	0.004*	0.000*
L3 Res Var L1MACH (L1SES)	0.00002*	0.005*	0.000*
L3 Res Var L1MACH (L1MACH)	0.00002*	0.005*	0.000*
L3 Res Var L1ATM slope	0.00001	0.0004	0.056
L3 Res Var L1GND slope	0.00000**	0.002**	0.500**
L3 Res Var L1VOM slope	0.00000	0.0003	0.128
L3 Res Var L1SES slope	0.00000*	0.001*	0.000*
L3 Res Var L1MACH slope	0.00000**	0.00001**	0.500**
L3 Res Var L2MACH (L1ATM)	0.00001	0.002	0.115
L3 Res Var L2MACH (L1GND)	0.00002*	0.004*	0.001*
L3 Res Var L2MACH (L1VOM)	0.00000	0.002	0.141
L3 Res Var L2MACH (L1SES)	0.00001**	0.003**	0.014**
L3 Res Var L2MACH (L1MACH)	0.00001**	0.004**	0.004**

Notes. Results displayed for random coefficient models of significantly associated L1 predictors of L1MSC. All L1 predictors are included (see Table 9). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05

L2BFLPE Moderation Effects of Student-Level (L1) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 1 L1 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L1NULL			814270.05
L1ATS1xL2SACH	0.0007**	0.0003**	95649.41
L1VOS1xL2SACH	0.0001*	0.0005*	50739.98
L1SESxL2SACH	0.0004	0.0005	26223.16
L1SACHxL2SACH	0.00001	0.00001	25364.59

L1ATS moderator	0.678*	0.012*
L1VOS moderator	0.218*	0.008*
L1SES moderator	0.086*	0.016*
L1SACH moderator	0.013*	0.0007*
L1SACH slope (L1ATS)	0.008*	0.0006*
L1SACH slope (L1SES)	0.012*	0.0008*
L1SACH slope (L1VOS)	0.0008*	0.012*
L2SACH slope (L1ATS)	-0.004**	0.001**
L2SACH slope (L1SES)	0.0016**	-0.005**
L2SACH slope (L1VOS)	0.001**	-0.003**
L2SACH slope (L1SACH)	-0.004**	0.001**
RANDOM EFFECTS	**β**	***SD***
L1 Intercept L1ATS	3.839	1.959
L1 Intercept L1SES	5.727	2.393
L1 Intercept L1VOS	4.977	2.231
L1 Intercept L1SACH	5.790	2.406
L2 Intercept L1ATS	0.153*	0.391*
L2 Intercept L1VOS	0.003*	0.055*
L2 Intercept L1SES	0.557*	0.746*
L2 Res Var L1SACH	0.557*	0.746*
L2 Res Var L1SACH slope (L1ATS)	0.0000*	0.002*
L2 Res Var L1SACH slope (L1VOS)	0.0000*	0.001*
L2 Res Var L1SACH slope (L1SES)	0.0000*	0.002*
L2 Res Var L1ATS slope	0.014*	0.120*
L2 Res Var L1SES slope	0.012*	0.110*
L2 Res Var L1VOS slope	0.339*	0.582*
L2 Res Var L1SACH slope	0.0000*	0.002*
L3 Intercept L1ATS	0.795*	0.892*
L3 Intercept L1SES	2.536*	1.592*
L3 Intercept L1VOS	1.394*	1.181*
L3 Res Var L1SACH	2.405*	1.551*
L3 Res Var L1ATS slope	0.002*	0.048*
L3 Res Var L1VOS slope	0.0007*	0.026*
L3 Res Var L1SES slope	0.001*	0.037*
L3 Res Var L1SACH slope	0.00001*	0.003*
L3 Res Var L1SACH slope (L1ATS)	0.0000*	0.027*
L3 Res Var L1SACH slope (L1VOS)	0.00001*	0.002*
L3 Res Var L1SACH slope (L1SES)	0.0000*	0.002*
L3 Res Var L1SACH slope (L1SACH)	0.00001*	0.003*
L3 Res Var L2SACH slope (L1ATS)	0.00001*	0.003*
L3 Res Var L2MACH slope (L1VOS)	0.00001**	0.003**
L3 Res Var L2SACH slope (L1SES)	0.004**	0.004**
L3 Res Var L2SACH slope (L1SACH)	0.00001**	0.003**
L3 Res Var L1ATSxL2SACH slope	0.0000**	0.0008**
L3 Res Var L1VOSxL2MACH slope	0.00000	0.0002
L3 Res Var L1SESxL2SACH slope	0.0000**	0.0010**
L3 Res Var L1SACHxL2SACH slope	0.00000	0.00002

L2BFLPE Moderation Effects of School-Level (L2) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 2 L2 INT			
FIXED EFFECTS	**β**	***SE***	**pvalue**	**-2LL**
L1NULL				849877.02
L2ATMxL2MACH	0.002	0.001	0.071	42966.12

L2CLMxL2MACH	0.0002	0.0001	0.135	43232.31
L2SESxL2MACH	-0.003**	0.001**	0.003**	42925.14
L2MSCxL2MACH	0.001	0.000	0.596	43025.60
L2VOMxL2MACH	0.00002	0.001	0.914	42945.28
L2ATM moderator	-0.865	0.514	0.110	
L2CLM moderator	-0.068	0.069	0.332	
L2SES moderator	1.299**	0.428**	0.006**	
L2MSC moderator	0.393	0.592	0.361	
L2VOM moderator	0.079	0.376	0.833	
L1MACH (L2ATM)	0.018*	0.001*	0.000*	
L1MACH (L2CLM)	0.018*	0.001*	0.000*	
L1MACH (L2SES)	0.018*	0.001*	0.000*	
L1MACH (L2MSC)	0.018*	0.001*	0.000*	
L1MACH (L2VOM)	0.018*	0.001*	0.000*	
L2MACH (L2ATM)	-0.025	0.012	0.054	
L2MACH (L2CLM)	-0.010	0.005	0.067	
L2MACH (L2SES)	0.008	0.004	0.059	
L2MACH (L2MSC)	0.011	-0.009	0.413	
L2MACH (L2VOM)	-0.006	0.019	0.739	
RANDOM EFFECTS	**β**	***SD***	**pvalue**	
L1 RES Var NULL	8.405	2.899		
L1 Res Var L2ATM	6.332	2.516		
L1 Res Var L2CLM	6.333	2.516		
L1 Res Var L2SES	6.331	2.516		
L1 Res Var L2MSC	6.332	2.516		
L1 Res Var L2VOM	6.331	2.516		
L2 Res Var NULL	0.537*	0.733*	0.000*	
L2 Res Var L2ATM	0.753*	0.868*	0.000*	
L2 Res Var L2CLM	0.708*	0.841*	0.000*	
L2 Res Var L2SES	0.768*	0.877*	0.000*	
L2 Res Var L2MSC	0.736*	0.858*	0.000*	
L2 Res Var L2VOM	0.757*	0.870*	0.000*	
L2 Res Var L1MACH (L2ATM)	0.000*	0.005*	0.000*	
L2 Res Var L1MACH (L2CLM)	0.000*	0.004*	0.000*	
L2 Res Var L1MACH (L2SES)	0.000*	0.005*	0.000*	
L2 Res Var L1MACH (L2MSC)	0.000*	0.005*	0.000*	
L2 Res Var L1MACH (L2VOM)	0.000*	0.005*	0.000*	
L3 Res Var NULL	0.537	0.733		
L3 Res Var L2ATM	2.578	1.606		
L3 Res Var L2CLM	2.577	1.605		
L3 Res Var L2SES	2.126	1.950		
L3 Res Var L2MSC	2.503	1.582		
L3 Res Var L2VOM	2.553	1.598		
L3 Res Var L1MACH (L2ATM)	0.000*	0.005*	0.000*	
L3 Res Var L1MACH (L2CLM)	0.000*	0.005*	0.000*	
L3 Res Var L1MACH (L2SES)	0.000*	0.005*	0.000*	
L3 Res Var L1MACH (L2MSC)	0.000*	0.005*	0.000*	
L3 Res Var L1MACH (L2VOM)	0.000*	0.005*	0.000*	
L3 Res Var L2ATM slope (L2ATM)	0.137**	0.370**	0.003**	
L3 Res Var L2CLM slope (L2CLM)	0.000*	0.022*	0.000*	

L3 Res Var L2SES slope (L2SES)	0.001**	0.024**	0.500**
L3 Res Var L2MSC slope (L2MSC)	0.192*	0.437*	0.000*
L3 Res Var L2VOM slope (L2VOM)	0.059*	0.242*	0.000*
L3 Res Var L2MACH (L2ATM)	0.000**	0.003**	0.008**
L3 Res Var L2MACH (L2CLM)	0.000**	0.003**	0.012**
L3 Res Var L2MACH (L2SES)	0.000**	0.004**	0.002**
L3 Res Var L2MACH (L2MSC)	0.000*	0.005*	0.000*
L3 Res Var L2MACH (L2VOM)	0.000**	0.004**	0.003**

Notes. Results displayed for random coefficient models of significantly associated L2 predictors of L1MSC only. L2LOC not included (see Table 11). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05

L2BFLPE Moderation Effects of School-Level (L2) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 2 L2 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L1NULL			814270.05
L2ATSxL2SACH	0.00004	0.00002	25430.71
L2CLMxL2SACH	0.000004	0.000009	25516.02
L2SESx L2SACH	-0.0009	0.0006	25352.61
L2SSC x L2SACH	0.00004**	0.00002**	25514.39

L2VOS x L2SACH	0.00002	0.00001	25400.83
L2ATS moderator	0.367*	0.041*	
L2CLM moderator	0.015**	0.007**	
L2SES moderator	0.441	0.286	
L2SSC moderator	0.520*	0.076*	
L2VOS moderator	0.173*	0.031*	
L1SACH slope (L2ATS)	0.013*	0.0008*	
L1SACH slope (L2CLM)	0.013*	0.0007*	
L1SACH slope (L2SES)	0.013*	0.0008*	
L1SACH slope (L2SSC)	0.013*	0.0008*	
L1SACH slope (L2VOS)	0.013*	0.0008*	
L2SACH slope (L2ATS)	-0.004**	0.002**	
L2SACH slope (L2CLM)	-0.003	0.001	
L2SACH slope (L2SES)	0.0006	0.003	
L2SACH slope (L2SSC)	-0.005*	0.001*	
L2SACH slope (L2VOS)	-0.004**	0.001**	
RANDOM EFFECTS	**β**	***SD***	
L1 Intercept L2ATS	5.79	2.41	
L1 Intercept L2CLM	5.79	2.41	
L1 Intercept L2SES	5.79	2.41	
L1 Intercept L2SSC	5.79	2.41	
L1 Intercept L2VOS	5.79	2.41	
L2 Intercept L2ATS	0.540*	0.735*	
L2 Intercept L2CLM	0.536*	0.732*	
L2 Intercept L2SES	0.557*	0.746*	
L2 Intercept L2SSC	0.524*	0.724*	
L2 Intercept L2VOS	0.548*	0.740*	
L2 Res Var L1SACH slope (L2ATS)	0.00000*	0.002*	
L2 Res Var L1SACH slope (L2CLM)	0.00000*	0.002*	
L2 Res Var L1SACH slope (L2SES)	0.00000*	0.002*	
L2 Res Var L1SACH slope (L2SSC)	0.00000*	0.002*	
L2 Res Var L1SACH slope (L2VOS)	0.00000*	0.002*	
L3 Intercept L2ATS	2.41*	1.55*	
L3 Intercept L2CLM	2.42*	1.55*	
L3 Intercept L2SES	2.42*	1.55*	
L3 Intercept L2SSC	2.41*	1.55*	
L3 Intercept L2VOS	2.43*	1.56*	
L3 Res Var L2SACH slope (L2ATS)	0.00001**	0.003**	
L3 Res Var L2SACH slope (L2CLM)	0.00001**	0.003**	
L3 Res Var L2SACH slope (L2SES)	0.00001**	0.004**	
L3 Res Var L2SACH slope (L2SSC)	0.00001**	0.003**	
L3 Res Var L2SACH slope (L2VOS)	0.00001**	0.003**	
L3 Res Var L2ATS slope	0.011	0.104	
L3 Res Var L2CLM slope	0.0003*	0.019*	
L3 Res Var L2SES slope	0.035**	0.186**	
L3 Res Var L2SSC slope	0.035	0.187	
L3 Res Var L2VOS slope	0.004	0.062	
L3 Res Var L1SACH slope (L2ATS)	0.00001*	0.003*	
L3 Res Var L1SACH slope (L2CLM)	0.00001*	0.003*	
L3 Res Var L1SACH slope (L2SES)	0.00001*	0.003*	
L3 Res Var L1SACH slope (L2SSC)	0.00001*	0.003*	
L3 Res Var L1SACH slope (L2VOS)	0.00001*	0.003*	
L3 Res Var L2ATSxL2SACH slope	0.00000	0.00003	
L3 Res Var L2CLMxL2SACH slope	0.0000**	0.00002**	

L3 Res Var L2SESxL2SACH slope	0.0000**	0.0004**
L3 Res Var L2SSCxL2SACH slope	0.00000	0.00002
L3 Res Var L2VOSxL2SACH slope	0.00000	0.00002

Notes. Results displayed for interactions with significantly associated L2 predictors of L1SSC only. L2LOC not included (see Table 14). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

L2BFLPE Moderation Effects of Country-Level (L3) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 3 L3 INT			
FIXED EFFECTS	**β**	***SE***	**pvalue**	**-2LL**
L3ATMxL2MACH	-0.002	0.002	0.292	42925.22
L3MSCxL2MACH	0.002	0.001	0.168	42945.91
L3VOMxL2MACH	0.001	0.001	0.291	42939.67
L3MACHxL2MACH	0.00003**	0.00001**	0.037**	42959.10
L3TRKxL2MACH	DNC	DNC	DNC	DNC
L3ATM moderator	1.348**	0.572**	0.002**	
L3MSC moderator	1.851*	0.159*	0.000*	
L3VOM moderator	1.327*	0.210*	0.000*	

L3MACH moderator	-0.026*	0.003*	0.000*
L2MACH (L3ATM)	-0.005**	0.002**	0.003**
L2MACH (L3MSC)	-0.005**	0.002**	0.007**
L2MACH (L3VOM	-0.005**	0.002**	0.012**
L1MACH (L3MACH)	-0.005**	0.002**	0.017**
L1MACH (L3ATM)	0.018*	0.001*	0.000*
L1MACH (L3MSC)	0.018*	0.001*	0.000*
L1MACH (L3VOM)	0.018*	0.001*	0.000*
L1MACH (L3MACH)	0.018*	0.001*	0.000*
RANDOM EFFECTS	**β**	***SD***	**pvalue**
L1 Res Var L3ATM	6.331	2.516	
L1 Res Var L3MSC	6.331	2.516	
L1 Res Var L3VOM	6.327	2.515	
L1 Res Var L3MACH	6.331	2.516	
L2 Res Var L3ATM	0.768*	0.876*	0.000*
L2 Res Var L3MSC	0.768*	0.876*	0.000*
L2 Res Var L3VOM	0.769*	0.877*	0.000*
L2 Res Var L3MACH	0.769*	0.877*	0.000*
L2 Res Var L1MACH (L3ATM)	0.00002*	0.005*	0.000*
L2 Res Var L1MACH (L3MSC)	0.00002*	0.005*	0.000*
L2 Res Var L1MACH (L3VOM)	0.00002*	0.005*	0.000*
L2 Res Var L1MACH (L3MACH)	0.00002*	0.005*	0.000*
L3 Res Var L3ATM	1.448*	1.203*	0.000*
L3 Res Var L3MSC	0.535*	0.731*	0.000*
L3 Res Var L3VOM	0.606*	0.778*	0.000*
L3 Res Var L3MACH	0.280*	0.529*	0.000*
L2 Res Var L1MACH (L3ATM)	0.00002*	0.005*	0.000*
L2 Res Var L1MACH (L3MSC)	0.00002*	0.005*	0.002*
L2 Res Var L1MACH (L3VOM)	0.00002*	0.004*	0.000*
L2 Res Var L1MACH (L3MACH)	0.00002*	0.005*	0.000*
L2 Res Var L2MACH (L3ATM)	0.00001*	0.003*	0.000*
L2 Res Var L2MACH (L3MSC)	0.00001*	0.004*	0.000*
L2 Res Var L2MACH (L3VOM)	0.00002**	0.004**	0.002**
L2 Res Var L2MACH (L3MACH)	0.00001**	0.004**	0.002**

Notes. Results displayed for random coefficient models of significantly associated L3 predictors of L1MSC only. L3IDV and L3IPC are not included (see Table 13). L3TRK did not converge (DNC). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05

L2BFLPE Moderation Effects of Country-Level (L3) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 3 L3 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L1NULL			814270.05
L3ATSxL2SACH	0.001	0.002	25372.66
L3SSCxL2SACH	-0.001	0.001	25420.73
L3VOSxL2SACH	-0.0001	0.001	25370.03
L3SACHxL2SACH	-0.00003	0.00003	25370.1
L3ATS moderator	1.269*	0.217*	
L3SSC moderator	1.347*	0.071*	
L3VOS moderator	0.774*	0.157*	

L3SACH moderator	-0.022*	0.005*
L2SACH slope (L3ATS)	-0.003**	0.002**
L2SACH slope (L3SSC)	-0.003**	0.001**
L2SACH slope (L3VOS	-0.004**	0.002**
L2SACH slope (L3SACH)	-0.003**	0.001**
L1SACH slope (L3ATS)	0.013*	0.0008*
L1SACH slope (L3SSC)	0.013*	0.0008*
L1SACH slope (L3VOS)	0.013*	0.0008*
L1SACH slope (L3SACH)	0.013*	0.0008*
RANDOM EFFECTS	**β**	***SD***
L1 Res Var Int L2ATS	5.79	2.41
L1 Res Var Int L3SSC	5.79	2.41
L1 Res Var Int L3VOS	5.79	2.41
L1 Res Var Int L3SACH	5.79	2.41
L2 Res Var Int L3ATS	0.557*	0.747*
L2 Res Var Int L3SSC	0.557*	0.746*
L2 Res Var Int L3VOS	0.558*	0.747*
L2 Res Var Int L3SACH	0.557*	0.747*
L2 Res Var Int L1SACH slope (L3ATS)	0.00000*	0.002*
L2 Res Var Int L1SACH slope (L3SSC)	0.00000*	0.002*
L2 Res Var Int L1SACH slope (L3VOS)	0.00000*	0.002*
L2 Res Var Int L1SACH slope (L3SACH)	0.00000*	0.002*
L3 Res Var Int L3ATS	0.958*	0.979*
L3 Res Var Int L3SSC	0.221*	0.470*
L3 Res Var Int L3VOS	0.781*	0.864*
L3 Res Var Int L3SACH	0.885*	0.941*
L3 Res Var Int L2SACH slope (L3ATS)	0.00001**	0.003**
L3 Res Var Int L2SACH slope (L3SSC)	0.00001**	0.004**
L3 Res Var Int L2SACH slope (L3VOS)	0.00001**	0.003**
L3 Res Var Int L2SACH slope (L3SACH)	0.00001*	0.003*
L3 Res Var Int L1SACH slope (L3ATS)	0.00001*	0.003*
L3 Res Var Int L1SACH slope (L3SSC)	0.00001*	0.003*
L3 Res Var Int L1SACH slope (L3VOS)	0.00001*	0.003*
L3 Res Var Int L1SACH slope (L3SACH)	0.00001*	0.003*

Notes. Results displayed for interactions with significantly associated L3 predictors of L1SSC. L3IDV and L3TRK are not included (see Table 16). L3IPC did not converge (DNC). L3 Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

Appendix F

All Results for L3BFLPE Moderation Effects in Math and Science

Appendix F
All Results for L3BFLPE Moderation Effects in Math and Science

L3BFLPE Moderation Effects of Student-Level (L1) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 1b L1 INT		
FIXED EFFECTS	β	*SE*	-2LL
L1NULL			849877.02
L1ATM1xL3MACH	-0.005*	0.0004*	108289.71
L1GNDxL3MACH	0.0004	0.0003	44427.17
L1VOM1xL3MACH	-0.0005*	0.0001*	56371.53
L1SESxL3MACH	-0.00007	0.0002	43194.25
L1MACH1xL3MACH	0.004**	0.00001**	42913.43
L1ATM moderator	0.620*	0.022*	
L1GND moderator	-0.344*	0.038*	

L1VOM moderator	0.188*	0.009*
L1SES moderator	0.02	0.015
L1MACH moderator	0.018	0.001
L1MACH (L1ATM)	0.011*	0.0005*
L1MACH (L1GND)	0.011*	0.008*
L1MACH (L1VOM)	0.016*	0.001*
L1MACH (L1SES)	0.018*	0.001*
L3MACH (L1ATM)	-0.016*	0.003*
L3MACH (L1GND)	-0.026*	0.003*
L3MACH (L1VOM)	-0.022*	0.003*
L3MACH (L1SES)	-0.027*	0.003*
L3MACH (L1MACH)	-0.030*	0.003*
RANDOM EFFECTS	**β**	***SD***
L1 Res Var Int L1ATM	4.355	2.087
L1 Res Var Int L1GND	6.223	2.495
L1 Res Var Int L1VOM	5.817	2.412
L1 Res Var Int L1SES	6.285	2.507
L1 Res Var Int L1MACH	6.33	2.516
L2 Res Var Int L1ATM	0.201*	0.448*
L2 Res Var Int L1GND	0.755*	0.869*
L2 Res Var Int L1VOM	0.594*	0.771*
L2 Res Var Int L1SES	0.771*	0.878*
L2 Res Var Int L1MACH	0.783*	0.885*
L2 Res Var ATM slope	0.012*	0.108*
L2 Res Var L1GND slope	0.329*	0.574*
L2 Res Var L1VOM slope	0.005*	0.068*
L2 Res Var L1SES slope	0.013*	0.113*
L2 Res Var L1MACH slope	0.00002*	0.005*
L2 Res Var L1MACH (L1ATM)	0.00001	0.003
L2 Res Var L1MACH (L1GND)	0.00002	0.005
L2 Res Var L1MACH (L1VOM)	0.00002	0.004
L2 Res Var L1MACH (L1SES)	0.00002	0.005
L3 Res Var Int L1ATM	0.386*	0.621*
L3 Res Var Int L1GND	0.278*	0.527*
L3 Res Var Int L1VOM	0.286*	0.535*
L3 Res Var Int L1SES	0.277*	0.526*
L3 Res Var Int L1MACH	0.286*	0.535*
L3 Res Var L1ATM slope	0.007*	0.080*
L3 Res Var L1GND slope	0.017*	0.132*
L3 Res Var L1VOM slope	0.001*	0.038*
L3 Res Var L1SES slope	0.003*	0.052*
L3 Res Var L1MACH (L1ATM)	0.00001*	0.002*
L3 Res Var L1MACH (L1GND)	0.00002*	0.005*
L3 Res Var L1MACH (L1VOM)	0.00002*	0.004*
L3 Res Var L1MACH (L1SES)	0.00002*	0.005*
L3 Res Var L1MACH (L1MACH)	0.00002*	0.004*

Notes. Results displayed for interaction effects of significantly associated L1 predictors of L1MSC. All L1 predictors are included (see Table 11). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

L3BFLPE Moderation Effects of Student-Level (L1) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 1 L1 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L1NULL			814270.05
L1ATSxL3SACH	0.0005*	0.0001*	95531.90
L1VOS1xL3SACH	-0.0002	0.0001	50641.88
L1SESxL3SACH	0.0004**	0.0002**	26150.94
L1SACH1xL3SACH	0.00003*	0.000008*	25323.81
L1ATS moderator	0.676*	0.011*	
L1VOS moderator	0.135*	0.092*	
L1SES moderator	0.085*	0.016*	
L1SACH moderator	0.012*	0.0007*	
L1SACH slope (L1ATS)	0.008*	0.0006*	
L1SACH1 slope (L1VOS)	0.010*	0.0007*	
L1SACH1 slope (L1SES)	0.012*	0.0008*	
L3SACH slope (L1ATS)	-0.012**	0.004**	
L3SACH slope (L1VOS)	-0.017**	0.005**	
L3SACH slope (L1SES)	-0.023*	0.005*	
L3SACH slope (L1SACH)	-0.025*	0.006*	

RANDOM EFFECTS	β	*SD*
L1 Res Var Int L1ATS	3.837	1.959
L1 Res Var Int L1VOS	4.977	2.231
L1 Res Var Int L1SES	5.728	2.393
L1 Res Var Int L1SACH	5.791	2.406
L2 Res Var Int L1ATS	0.156*	0.395*
L2 Res Var Int L1VOS	0.344*	0.587*
L2 Res Var Int L1SES	0.566*	0.753*
L2 Res Var Int L1SACH	0.565*	0.752*
L2 Res Var Int L1SACH slope (L1ATS)	0.00000*	0.002*
L2 Res Var Int L1SACH slope (L1VOS)	0.00000*	0.001*
L2 Res Var Int L1SACH slope (L1SES)	0.00000*	0.002*
L2 Res Var Int L1ATS slope	0.015*	0.121*
L2 Res Var Int L1VOS slope	0.003*	0.056*
L2 Res Var Int L1SES slope	0.012*	0.111*
L2 Res Var Int L1SACH slope	0.00000*	0.002*
L3 Res Var Int L1ATS	0.441*	0.664*
L3 Res Var Int L1VOS	0.679*	0.824*
L3 Res Var Int L1SES	0.879*	0.938*
L3 Res Var Int L1SACH	0.850*	0.922*
L3 Res Var Int L1ATS slope	0.002*	0.040*
L3 Res Var Int L1VOS slope	0.0006*	0.024*
L3 Res Var Int L1SES slope	0.002*	0.042*
L3 Res Var Int L1SACH slope	0.850	0.922
L3 Res Var Int L1SACH slope (L1ATS)	0.00000	0.002
L3 Res Var Int L1SACH slope (L1VOS)	0.00001	0.002
L3 Res Var Int L1SACH slope (L1SES)	0.00001	0.003

Notes. Results displayed for interaction effects of significantly associated L1 predictors of L1SSC only. L1GND is excluded (see Table 11). Sample size = 169,810 students; 5,410 schools; 26 countries. *p < .001, **p <.05.

L3BFLPE Moderation Effects of School-Level (L2) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 2b L2 INT		
FIXED EFFECTS	β	*SE*	-2LL
L2NULL			849877.02
L2ATMxL3MACH	-0.005*	0.001*	177458.96
L2CLMxL3MACH	0.0002	0.000	43327.42
L2SESxL3MACH	-0.0002	0.0007	43385.44
L2MSCxL3MACH	-0.005**	0.002**	42980.22
L2VOMxL3MACH	-0.002**	0.001**	42940.24
L2MACHxL3MACH	-0.00003	0.000	42959.11
L2ATM moderator	0.263*	0.070*	
L2CLM moderator	0.0327*	0.006*	
L2SES moderator	-0.296*	0.040*	
L2MSC moderator	0.422**	0.128**	
L2VOM moderator	0.091	0.093	
L2MACH moderator	-0.005**	0.002**	

L1MACH (L2ATM)	0.018*	0.001*
L1MACH (L2CLM)	0.018*	0.005*
L1MACH (L2SES)	0.018*	0.001*
L1MACH (L2MSC)	0.018*	0.001*
L1MACH (L2VOM)	0.018*	0.001*
L1MACH (L2MACH)	0.018*	0.001*
L3MACH (L2ATM)	-0.026*	0.003*
L3MACH (L2CLM)	-0.030*	0.003*
L3MACH (L2SES)	-0.026*	0.003*
L3MACH (L2MSC)	-0.026*	0.003*
L3MACH (L2VOM)	-0.026*	0.003*
L3MACH (L2MACH)	-0.026*	0.003*
RANDOM EFFECTS	**β**	***SD***
L1 Intercept L2ATM	6.331	2.516
L1 Intercept L2CLM	6.332	2.516
L1 Intercept L2SES	6.33	2.516
L1 Intercept L2MSC	6.330	2.516
L1 Intercept L2VOM	6.329	2.516
L1 Intercept L2MACH	6.331	2.516
L2 Intercept L2ATM	0.765*	0.875*
L2 Intercept L2CLM	0.719*	0.848*
L2 Intercept L2SES	0.685*	0.827*
L2 Intercept L2MSC	0.765*	0.875*
L2 Intercept L2VOM	0.773*	0.879*
L2 Intercept L2MACH	0.769*	0.877*
L2 Res Var L1MACH (L2ATM)	0.00002*	0.005*
L2 Res Var L1MACH (L2CLM)	0.00002*	0.004*
L2 Res Var L1MACH (L2SES)	0.00002*	0.005*
L2 Res Var L1MACH (L2MSC)	0.00002*	0.005*
L2 Res Var L1MACH (L2VOM)	0.00002*	0.005*
L2 Res Var L1MACH (L2MACH)	0.00002*	0.005*
L3 Intercept L2ATM	0.300*	0.546*
L3 Intercept L2CLM	0.332*	0.576*
L3 Intercept L2SES	0.245*	0.494*
L3 Intercept L2MSC	0.284*	0.822*
L3 Intercept L2VOM	0.286*	0.535*
L3 Intercept L2MACH	0.280*	0.530*
L3 Res Var L2ATM slope	0.006	0.072
L3 Res Var L2CLM slope	0.0003*	0.018*
L3 Res Var L2SES slope	0.018*	0.135*
L3 Res Var L2MSC slope	0.061**	0.246**
L3 Res Var L2VOM slope	0.029**	0.171**

L3 Res Var L2MACH slope	0.00001*	0.004*
L3 Res Var L1MACH (L2ATM)	0.0000*	0.005*
L3 Res Var L1MACH (L2CLM)	0.00002*	0.005*
L3 Res Var L1MACH (L2SES)	0.018*	0.135*
L3 Res Var L1MACH (L2MSC)	0.00002*	0.005*
L3 Res Var L1MACH (L2VOM)	0.00002*	0.005*
L3 Res Var L1MACH (L2MACH)	0.00002*	0.005*

Notes. Results displayed for Interaction effects of significantly associated L2 predictors of L1MSC for L3BFLPE. L2LOC not included (see Table 13). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

L3BFLPE Moderation Effects of School-Level (L2) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 2 L2 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
NULL			814270.05
L2ATSxL3SACH	-0.001	0.0006	25388.06
L2CLMxL3SACH	0.0003*	0.00009*	25478.43
L2SESx L3SACH	-0.002*	0.0008*	25521.62
L2SSC x L3SACH	-0.002	0.001	178773.67
L2VOS x L3SACH	-0.001**	0.0004**	25364.23
L2SACH x L3SACH	-0.00003	0.00003	25370.10
L2ATS moderator	0.048*	0.011*	
L2CLM moderator	0.013**	0.005**	
L2SES moderator	-0.193*	0.034*	
L2SSC moderator	0.450*	0.053*	
L2VOS moderator	0.145*	0.020*	
L2SACH moderator	-0.003**	0.001**	
L1SACH slope (L2ATS)	0.013*	0.0008*	
L1SACH slope (L2CLM)	0.013*	0.0007*	
L1SACH slope (L2SES)	0.013*	0.0007*	
L1SACH slope (L2SSC)	0.095*	0.012*	

L1SACH slope (L2VOS)	0.013*	0.0008*
L1SACH slope (L2SACH)	0.013*	0.0008*
L3SACH slope (L2ATS)	-0.022*	0.005*
L3SACH slope (L2CLM)	-0.022*	0.005*
L3SACH slope (L2SES)	-0.022*	0.005*
L3SACH slope (L2SSC)	-0.022*	0.005*
L3SACH slope (L2VOS)	-0.022*	0.005*
L3SACH slope (L2SACH)	-0.022*	0.005*
RANDOM EFFECTS	**β**	***SD***
L1 Intercept L2ATS	5.79	2.41
L1 Intercept L2CLM	5.79	2.41
L1 Intercept L2SES	5.79	2.41
L1 Intercept L2SSC	5.77	2.40
L1 Intercept L2VOS	5.79	2.41
L1 Intercept L2SACH	5.79	2.41
L2 Intercept L2ATS	0.549*	0.741*
L2 Intercept L2CLM	0.547*	0.739*
L2 Intercept L2SES	0.544*	0.738*
L2 Intercept L2SSC	0.541*	0.736*
L2 Intercept L2VOS	0.556*	0.745*
L2 Intercept L2SACH	0.557*	0.746*
L2 Res Var L1SACH slope (L2ATS)	0.00000*	0.002*
L2 Res Var L1SACH slope (L2CLM)	0.00000*	0.002*
L2 Res Var L1SACH slope (L2SES)	0.00000*	0.002*
L2 Res Var L1SACH slope (L2SSC)	0.00000*	0.002*
L2 Res Var L1SACH slope (L2VOS)	0.00000*	0.002*
L2 Res Var L1SACH slope (L2SACH)	0.00000*	0.002*
L3 Intercept L2ATS	0.864*	0.929*
L3 Intercept L2CLM	0.848*	0.921*
L3 Intercept L2SES	0.817*	0.904*
L3 Intercept L2SSC	0.793*	0.887*
L3 Intercept L2VOS	0.870*	0.933*
L3 Intercept L2SACH	0.885*	0.941*
L3 Res Var L2ATS slope	0.005	0.069
L3 Res Var L2CLM slope	0.0002*	0.015*
L3 Res Var L2SES slope	0.015*	0.120*
L3 Res Var L2SSC slope	0.008	0.091
L3 Res Var L2VOS slope	0.0003	0.016
L3 Res Var L2SACH slope	0.00001*	0.003*
L3 Res Var L1SACH (L2ATS)	0.00001*	0.003*
L3 Res Var L1SACH (L2CLM)	0.00001*	0.003*
L3 Res Var L1SACH (L2SES)	0.00001*	0.003*
L3 Res Var L1SACH (L2SSC)	0.00001*	0.003*
L3 Res Var L1SACH (L2VOS)	0.00001*	0.003*
L3 Res Var L1SACH (L2SACH)	0.00001*	0.003*

Notes. Results displayed for interaction effects of significantly associated L2 predictors of L1SSC only. L2LOC is excluded (see Table 14). Sample size = 169,810 students; 5,410 schools; 26 countries *$p < .001$, **$p < .05$.

L3BFLPE Moderation Effects of Country-Level (L3) Predictors of Students' Math Self-Concept (L1MSC)

	MODEL 3b L3 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L3NULL			849877.02
L3ATMxL3MACH	0.005	0.003	42911.32
L3MSCxL3MACH	0.003*	0.0007*	42984.19
L3VOMxL3MACH	0.003**	0.002**	42915.31
L3TRKxL3MACH	0.0003	0.001	42913.95
L3ATM moderator	-2.810	1.537	
L3MSC moderator	-0.552	0.425	
L3VOM moderator	-1.501	0.759	
L3TRK moderator	-0.484**	0.217**	
L3MACH (L3ATM)	-0.042**	0.027**	
L3MACH (L3MSC)	-0.048*	0.008*	
L3MACH (L3VOM	-0.085**	0.030**	
L3MACH (L3TRK)	-0.025*	0.003*	
L1MACH (L3ATM)	0.018*	0.001*	
L1MACH (L3MSC)	0.018*	0.001*	
L1MACH (L3VOM)	0.018*	0.001*	
L1MACH (L3TRK	0.018*	0.001*	
RANDOM EFFECTS	β	SD	
L1 Intercept L3ATM	6.330	2.516	

L1 Intercept L3MSC	6.330	2.516
L1 Intercept L3VOM	6.330	2.516
L1 Intercept L3TRK	6.330	2.516
L2 Intercept L3ATM	0.784*	0.886*
L2 Intercept L3MSC	0.782*	0.884*
L2 Intercept L3VOM	0.784*	0.885*
L2 Intercept L3TRK	0.784*	0.885*
L2 Res Var L1MACH (L3ATM)	0.00002*	0.005*
L2 Res Var L1MACH (L3MSC)	0.00002*	0.005*
L2 Res Var L1MACH (L3VOM)	0.00002*	0.005*
L2 Res Var L1MACH (L3TRK)	0.00002*	0.005*
L3 Intercept L3ATM	0.259*	0.509*
L3 Intercept L3MSC	0.010*	0.102*
L3 Intercept L3VOM	0.209*	0.457*
L3 Intercept L3TRK	0.220*	0.469*
L3 Res Var L1MACH (L3ATM)	0.00002*	0.0045*
L3 Res Var L1MACH (L3MSC)	0.00002*	0.0045*
L3 Res Var L1MACH (L3VOM)	0.00002*	0.0045*
L3 Res Var L1MACH (L3TRK)	0.00002*	0.0045*

Notes. Results displayed for interaction effects of significantly associated L3 predictors of L1MSC only. L3IDV and L3IPC are not included (see Table 15). Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

L3BFLPE Moderation Effects of Country-Level (L3) Predictors of Students' Science Self-Concept (L1MSC)

	MODEL 3 L3 INT		
FIXED EFFECTS	**β**	***SE***	**-2LL**
L1NULL			814270.05
L3ATSxL3SACH	0.015**	0.006**	25333.39
L3SSCxL3SACH	0.0007	0.001	25422.02
L3VOSxL3SACH	0.008*	0.002*	25336.26
L3ATS L3SACH slope	-0.154**	0.057**	
L3SSC L3SACH slope	0.133	0.108	
L3VOS L3SACH slope	-0.165*	0.032*	
L3ATS L1SACH slope	0.0127*	0.0008*	
L3SSC L1SACH slope	0.013*	0.0008*	
L3VOS L1SACH slope	0.013*	0.0008*	
L3ATS moderator	-6.714**	3.031**	
L3SSC moderator	0.743	0.494	
L3VOS moderator	-3.475*	0.864*	
RANDOM EFFECTS	**β**	***SD***	
L1 Res Var Int L3ATS	5.790	2.406	
L1 Res Var Int L3SSC	5.791	2.406	
L1 Res Var Int L3VOS	5.791	2.406	
L2 Res Var Int L3ATS	0.565*	0.752*	
L2 Res Var Int L3SSC	0.564*	0.751*	

L2 Res Var Int L3VOS	0.565*	0.752*
L2 Res Var Int L1SACH (L3ATS)	0.00000*	0.002*
L2 Res Var Int L1SACH (L3SSC)	0.00000*	0.002*
L2 Res Var Int L1SAC (L3VOS)	0.000000*	0.002*
L3 Res Var Int L3ATS	0.351*	0.592*
L3 Res Var Int L3SSC	0.020*	0.140*
L3 Res Var Int L3VOS	0.306*	0.553*
L3 Res Var Int L1SACH (L3ATS)	0.00001*	0.003*
L3 Res Var Int L1SACH (L3SSC)	0.00001*	0.003*
L3 Res Var Int L1SACH (L3VOS)	0.00001*	0.003*

Notes. Results displayed for interaction effects of significantly associated L3 predictors of L1SSC only. L3IDV and L3TRK are excluded (see Table 16). L3 IDV was excluded as it did not converge in the model. Sample size = 169,810 students; 5,410 schools; 26 countries *p < .001, **p <.05.

Appendix G

Results Visualization Graphs

Appendix G

Results Visualization Graphs

Figure G11

Effects of Student Math Achievement on Student Math Self-Concept

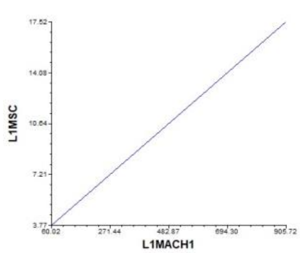


Figure G12

Effects of school-averaged math achievement on the relationship between students' math self-concept and students' math achievement (L2BFLPE).

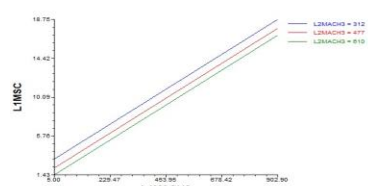


Figure G13

Effects of School-Averaged Math Achievement on Student Math Self-Concept (L2BFLPE)

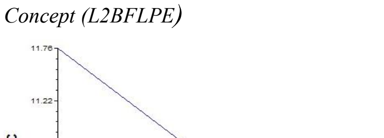


Figure G14

Effects of L2BFLPE for Varying Levels of Student Math Achievement

Figure G15

Effects of Student Science Achievement on Science Student Self-Concept

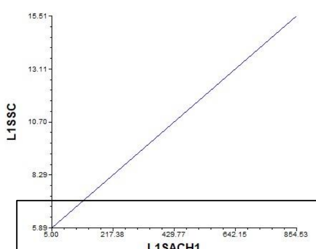


Figure G16

Effects of school-averaged science achievement on the relationship between students' science self-concept and students' science achievemnt (L2BFLPE).

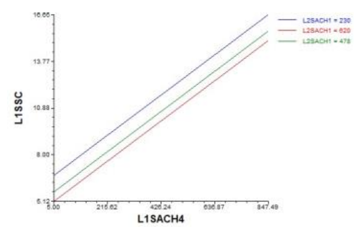


Figure G17

Effects of School-Averaged Science Achievement on Student Science Self-Concept (L2BFLPE)

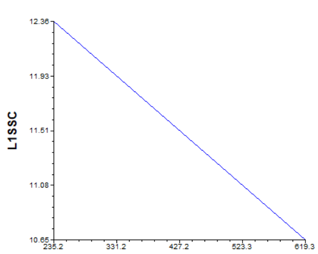


Figure G18

Effects of L2BFLPE for Varying Self-Concept (per student ability) (L2BFLPE)

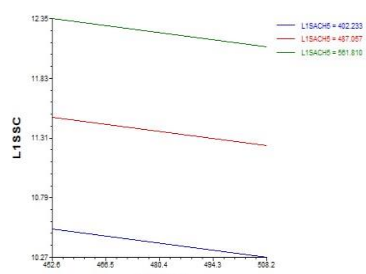


Figure G19

Effects of Country-Averaged Math Achievement on Student Math Self-Concept (L3BFLPE)

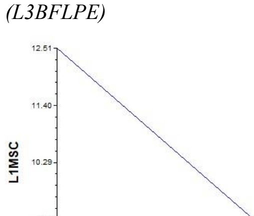


Figure G20

Effects of L3BFLPE for Varying Levels of Student Math Achievement

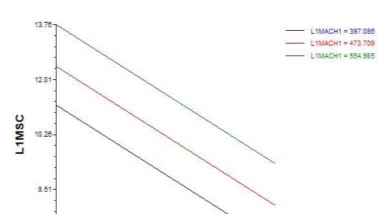


Figure G21

Effects of country-averaged math achievement on the relationship between students' math self-concept and students' math achievement (L3BFLPE).

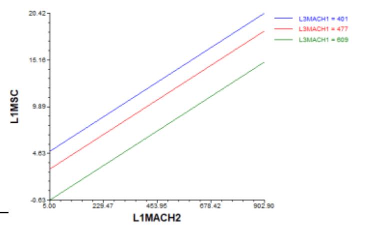


Figure G22

Effects of Country Science Achievement on Student Science Self-Concept (L3BFLPE)

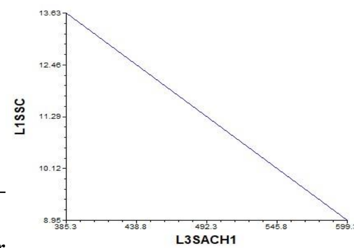
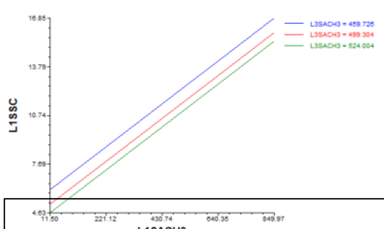


Figure G23

Effects of country-averaged science achievement on the relationship between students' science self-concept and students' science achievement (L3BFLPE).



Figur

Effects of L3BFLPE for Varying Levels of Student Science Achievement

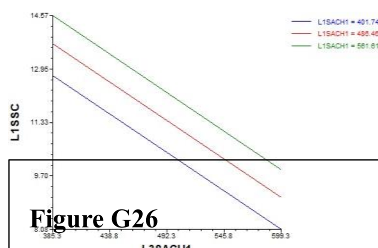


Figure G25

Effects of Various Levels of L1VOM on L2BFLPE in Math

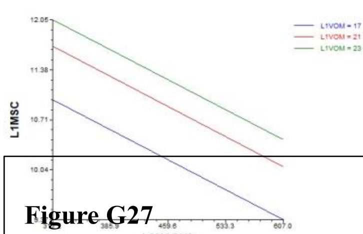


Figure G26

Effects of Various Levels of L2SES on L2BFLPE in Math

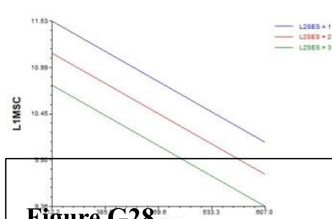


Figure G27

Effects of Various Levels of L3MACH on L2BFLPE in Math

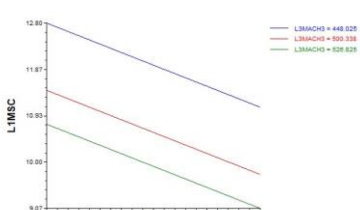


Figure G28

Effects of Various Levels of L1ATS on L2BFLPE in Science

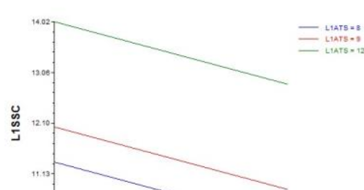


Figure G29

Effects of Various Levels of L1VOS on L2BFLPE in Science

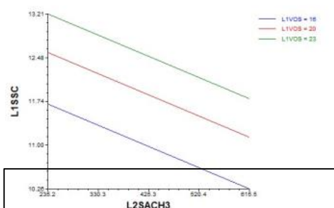


Figure G30

Effects of Various Levels of L2SSC on L2BFLPE in Science

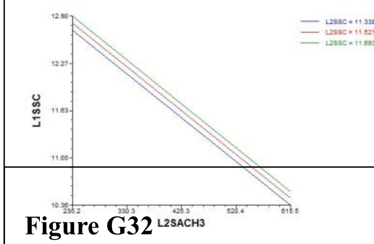


Figure G31

Effects of Various Levels of L1ATM on L3BFLPE in Math

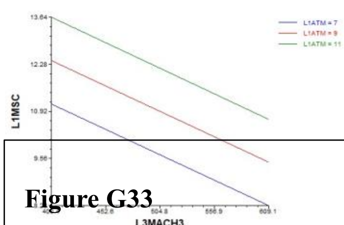


Figure G32

Effects of Various Levels of L1VOM on L3BFLPE in Math

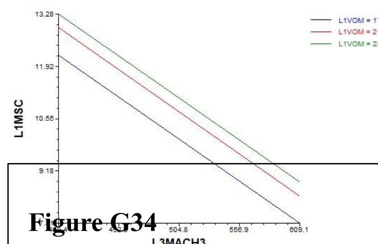


Figure G33

Effects of Various Levels of L1MACH on L3BFLPE in Math

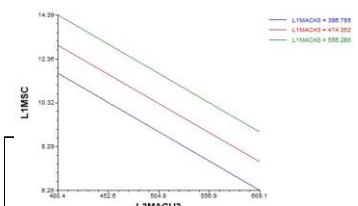
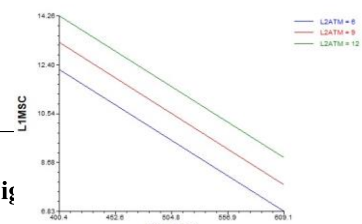
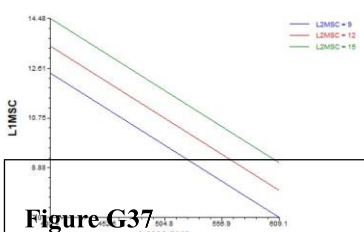


Figure G34

Effects of Various Levels of L2ATM on L3BFLPE in Math



Effects of Various Levels of L2MSC on L3BFLPE in Math



Fi

Effects of Various Levels of L2VOM on L3BFLPE in Math

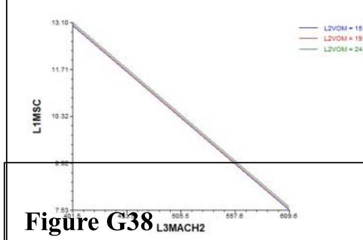


Figure G37

Effects of Various Levels of L3MSC on L3BFLPE in Math

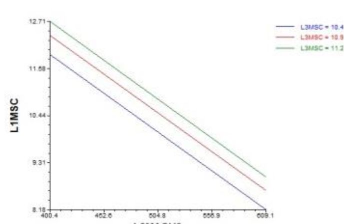


Figure G38

Effects of Various Levels of L3VOM on L3BFLPE in Math

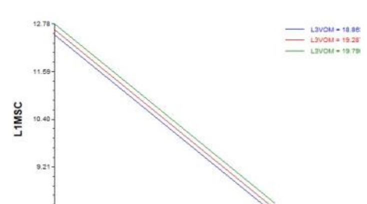


Figure G39

Effects of Various Levels of L1ATS on L3BFLPE in Science

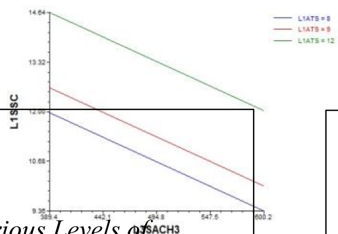


Figure G40

Effects of Various Levels of L1SES on L3BFLPE in Science

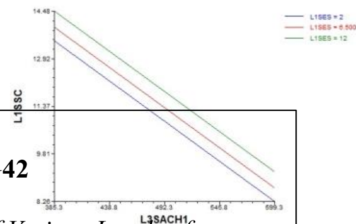


Figure G41

Effects of Various Levels of L1SACH on L3BFLPE in Science

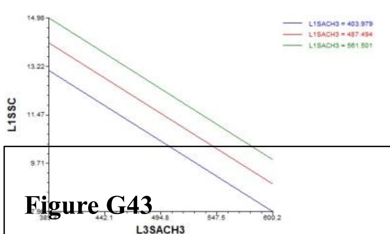


Figure G42

Effects of Various Levels of L2CLM on L3BFLPE in Science

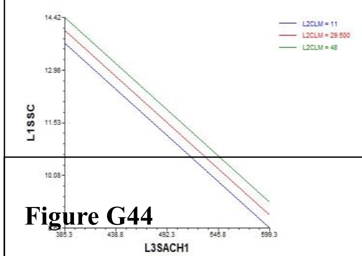


Figure G43

Effects of Various Levels of L2SES on L3BFLPE in Science

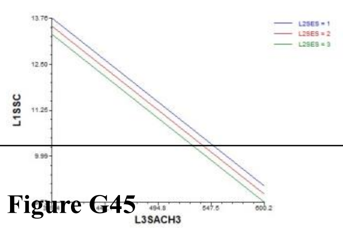


Figure G44

Effects of Various Levels of L2VOS on L3BFLPE in Science

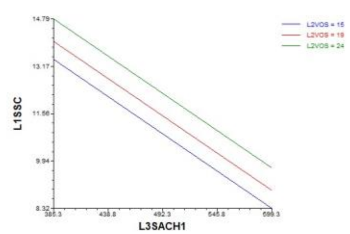


Figure G45

Effects of Various Levels of L3ATS on L3BFLPE in Science