Fall 12-15-2021

Levels of Influence: How Teacher Self-Efficacy Reflects Individual, School, and Country Level Variables

Rachel B. Goldberg
Seton Hall University, rachel.goldberg@student.shu.edu

Levels of Influence:

How Teacher Self-Efficacy Reflects Individual, School, and Country Level Variables

Rachel B. Goldberg

Dissertation Committee:

Wendiann R. Sethi, Ph.D.,Mentor
Daniel Gutmore, Ph.D.
Brian Osborne, Ed.D.

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Education

in the Department of Educational Leadership, Management, and Policy
Seton Hall University
2021

COLLEGE OF EDUCATION & HUMAN SERVICES
DEPARTMENT OF EDUCATION LEADERSHIP MANAGEMENT & POLICY

APPROVAL FOR SUCCESSFUL DEFENSE

Rachel Goldberg has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Fall** Semester.

DISSERTATION COMMITTEE
(please sign and date)

Dr. Wendiann Sethi
Mentor 10/19/2021
Date

Dr. Daniel Gutmore
Committee Member 10/19/2021
Date

Dr. Brian Osborne
Committee Member 10/19/2021
Date

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

ABSTRACT

The primary goal of the study was to investigate relationships between teacher self-efficacy and associated factors not previously studied at an international level. This study used the data gathered through the Programme for International Student Assessment (PISA) assessment conducted in 2018, with an initial sample size of 107,367 participating teachers from 6,128 schools across 19 countries to gain a global perspective regarding the individual and environmental factors that impact teacher self-efficacy. A blocked hierarchical regression model was chosen to support the theoretical structure of the analysis by examining the relationships between three levels of independent variables and teacher self-efficacy. The model predicted over 30 percent of teacher self-efficacy based on the full Teacher Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001) and a subset of questions related to classroom instruction and student engagement, exemplifying the strength of the inclusion of indicators specific to classroom instruction and student engagement when measuring teacher self-efficacy. Confirming prior research, demographic variables were weak predictors of teacher self-efficacy, while professional development participation was a stronger predictor. New to the extant body of research were the positive relationships between school leadership, school, and country-level student achievement, which served as the strongest predictors of teacher self-efficacy. Student achievement by country served as the most significant predictor of teacher self-efficacy, with an inverse relationship at the school and country level between student achievement and teacher self-efficacy. The study findings suggest that the external context is a significant factor in teacher self-efficacy.

Keywords: teacher self-efficacy, PISA, professional development, school leadership, student achievement, federal policy, economic indicators, teacher doubt, Dunning-Krueger effect

ACKNOWLEDGEMENTS

I am deeply grateful to the support of my husband, Scott, and my three amazing children Talia, Doran, and Noah. My graduate work and the dissertation writing have taken place at the same time as the world has been shaken by the COVID pandemic. My family has been patient and kind and have served as my strongest cheerleaders. They always knew when I needed quiet, or when I needed a break. There is no question that this work would not be complete without the incredible strength of my husband. He is a partner in all things and my buoy when the waters are rough.

My mother, Dr. Anna Brandon, was our family's first "doctor", and her courage to challenge the status quo, determination to meet goals, work ethic to achieve, and ongoing love inspired me to complete this program.

I would not have started this program without my mentor and champion Pablo Muñoz. Throughout my career, he has challenged me to stay focused on student learning, to think deeper, and reminded me that the work in the arena is not easy nor for the faint of heart, but in the end, it makes a difference.

My deep gratitude goes to my dissertation advisor Dr. Wendiann Sethi, whose calm guidance and direction supported my research and learning. From the moment she welcomed me into my first statistics class and through the ups and downs of learning more advanced measures, she listened, directed, and supported my work.

My two other committee members both played significant roles in this research. Dr. Daniel Gutmore began my first class in the doctoral program with a study on teacher self-efficacy and that experience paved the path toward this research. Dr. Brian Osborne has served

as a critical thought partner, supported me throughout my career transitions, and continues to challenge me to think deeper about the work of teaching and learning.

Finally, I am grateful for the educators that led me down this path. From my high school English teachers, Mrs. Goolsby and Ms. Moore, through three New Jersey school districts, I am continually inspired by the people that transform the lives of students.

TABLE OF CONTENTS

FIGURES, TABLES, & GRAPHS

CHAPTER I: INTRODUCTION

In 2001 the Organisation for Economic Co-operation and Development (OECD) introduced a new large-scale international assessment, the Programme for International Student Assessment (PISA), with the goal of meeting growing demand for educational global data and providing student achievement data to inform policymakers around the world. It joined the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS), assessments that provide tools to support international achievement comparisons. In addition, OECD created the Teaching and Learning International Survey (TALIS) to gather information about teaching conditions for international comparatives. Each of these international large-scale assessments collects a significant amount of data related to the structural, financial, and political aspects of educational structures and policies (Chmielewski & Dhuey, 2017).

International assessment provides a rich source of data for researchers, government leaders, and policy think-tanks to identify the next "thing" for education. The OECD uses the findings from PISA to provide guidance for developing countries and encourage member-nation dialogue, as exemplified by policy guidance documents that compare PISA findings with country policies (OECD, 2018). Breakspear (2014) argues that policy makers across countries use PISA to inform policymaking, with policy makers identifying high performing countries as critical to the study and development of education policies. High achieving countries gain international attention and inform high level discussions about policy development on a global scale.

In the United States, the first PISA results were released soon after President George W. Bush introduced the framework for the No Child Left Behind Act (No Child Left Behind [NCLB], 2002)—a significant revision of Elementary and Secondary Education Act of 1969

(ESEA). Signed into law in early 2002, NCLB demanded stronger educational standards and expectations for student growth through the use of new reading curricula, standardized tests, and accountability mechanisms. It also marked an unprecedented level of federal investment in school reform efforts with billions invested to increase student academic performance (Bloomfield & Cooper, 2003).

The findings of the first PISA confirmed the historical concerns of many in the U.S.—the nation was not at the top of the world in secondary education achievement. While other countries reacted with "shock," leaders in the U.S. appeared to disregard these findings, a symptom of the hangover from the 1957 Sputnik moment that highlighted weaknesses in the American education system, fears of global competition, and significant federal oversight of policy initiatives (Martens & Niemann, 2010; Singer et al., 2018).

This low international achievement was acknowledged fourteen years later when President Barak Obama introduced Every Student Succeeds Act (Every Student Succeeds Act [ESSA], 2015), the next iteration of the ESEA. He announced that the policy included higher expectations to place the U.S. in a "position to out-teach and out-compete other nations at a time when knowledge is really the single-biggest determinant of economic performance," at the same time lamenting that the nation's educational achievement levels were falling behind those of other countries (Obama, 2015).

At the same time of the release of the first PISA data, U.S. policymakers were paying close attention to measuring teacher quality and efficacy in relation to student achievement. Highly cited research confirmed the quality of a single teacher has a greater impact on student achievement than the school a student attends (Rivken et al., 2005). Results from the United States-specific National Assessment of Educational Progress (NAEP) also indicate significant

variation in student achievement at the school level as well as city and state levels, highlighting the wide range of achievement across the nation (Singer et al., 2018). Researchers used this data to qualify and quantify quality teaching practices and teachers, with a clear vision of replicating the most successful teachers and educational systems. Policymakers looked to the research to guide development of federal and state regulations regarding all facets of the public education sphere. In order to catalyze state action around teacher evaluation and education President Obama introduced the Race to the Top (RTTT) grant, investing additional funding in state systems that promoted new teacher evaluation systems combined with a push toward national standards, known as the "Common Core." Private foundations such as the Bill and Melinda Gates Foundation and the Carnegie Foundation contributed hundreds of millions toward research and the development of teacher training and quality measures (Bleiberg & Harbatkin, 2020; Reckhow & Snyder, 2014; Zeichner & Peña-Sandoval, 2015).

Ultimately, the investments did little to shift the United States' place in the international assessment landscape. In the first PISA results announced in 2001, the U.S. ranked twelfth in reading, fifteenth in math, and twelfth in science among 34 countries/economies (Lemke et al., 2001). Following the release of the 2018 PISA data, the New York Times led with "'It Just Isn't Working': PISA Test Scores Cast Doubt on U.S. Education Efforts" (Goldstein, 2019); the Washington Post titled their coverage, "U.S. students continue to lag behind peers in East Asia and Europe in reading, math and science, exams show" (Balingit & Van Dam, 2019). Darling-Hammond (2014) proposes international assessments such as PISA spurred a faulty feedback loop, ultimately moving the U.S. backward in student achievement.

Continued research on the impact of reforms, including stronger external definitions and evaluations of teacher efficacy, have not shown much progress. Kraft and Gilmore's (2017)

analysis of the shifts in teacher evaluation systems identified significant variations in how such reforms were implemented, finding that teacher evaluation ratings remained inconsistent and the new teacher evaluation systems have not resulted in significant differentiation between teacher quality.

Despite efforts to quantify the impact a teacher has on student learning, the question of what specific skills determine teacher efficacy continue to drive researchers. Today, the industry around defining the qualities of a successful teacher is a multi-million-dollar business (Chambers et al., 2013). While teacher evaluation materials and processes largely work to define external perceptions of educator quality, several of the most popular include questions about teachers' internal definitions and understanding of quality teaching. In a comparative study of five widely used evaluation systems, Gill et al. (2016) found that two included specific language around teacher professionalism, a dimension of teacher practice that includes reflection on efficacy.

One of the most widely used evaluation ratings programs is the Charlotte Danielson Framework for Teaching (Danielson, 2008; 2013). Among the ratings categories is "Reflecting on Teaching," which includes an indicator that describes a "Distinguished Teacher" as one who is able to draw "on an extensive repertoire of skills, [and] offers specific alternative actions, complete with the probable success of different courses of action" (Danielson, 2008, p. 85).

The Marzano Model (Marzano, 2007) has been developed into rubrics such as the one used by the State of Washington for public school teachers. It includes a component that rates the degree to which "the teacher reflects on and evaluates the effectiveness of instructional performance to identify areas of pedagogical strength and weakness" (p. 28).

The UTeach Observational Protocol (UTOP), developed by the University of Texas (2014) includes a rating of teachers' ability to reflect on a lesson and the decision-making

relating to the lesson. The rating requires a post-lesson interview or survey and expects that the teacher's reflection regarding their efficacy relates to observations of the lesson (UTOP, 2014).

Breakspear (2014) recommends OECD utilize PISA to shift its impact on policy from a narrow, assessment-centered lens to one that "would serve to highlight the multidimensionality of school systems and support the implementation of coherent and sustainable reform that can impact positively upon the educational and life outcomes of young people and the societies to which they will contribute" (p. 14). While it may not be clear that is the exact path taken to this point, the addition of teacher questionnaires in 2015 and then more substantively in 2018 demonstrates an increase in OECD's exploration of educational contexts and dimensions of the teacher experience. Internationally and in the United States, it is clear that external evaluators are looking closely at a teacher's ability to meta-analyze, understand, and judge the efficacy of their personal practice. Such a capacity for teachers to understand the skills necessary for effective instruction and reflect on their own practice in relationship to those skills is identified in psychological research as self-efficacy.

Self-Efficacy

The concept of self-efficacy in the context of individual behaviors and performance was first presented by psychologist Albert Bandura in 1977. As defined by Bandura, self-efficacy is a construct originating in psychological social learning theory that frames how a person perceives their personal capacity and mastery of behaviors and how such perceptions impact individual agency, making self-efficacy central to understanding learning and behavior development. Over the last forty years, self-efficacy has been used in many sectors as a tool to guide organizations as they support individual development. In the education sector, self-efficacy has been applied across organizations, from pre-schools to university settings and student-teaching experiences

through retirement. As a critical learning construct, self-efficacy continues to serve as a frequent subject for research (Burić & Kim, 2020; Guo et al., 2012; Klassen et al., 2011; Klassen & Tze, 2014; Morris et al., 2017; Shahid & Thompson, 2001; Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001; Zee & Koomen, 2016).

Teacher Self- Efficacy

Teacher self-efficacy (TSE) was first presented in a research context via a 1976 RAND study of urban student reading proficiency and has since served as a popular construct in educational research (Gibson & Dembo, 1984; Klassen & Tze 2014; Zee & Koomen 2016). In 1998 Tschannen-Moran et al. presented a definition of TSE as the "teacher's belief in her and his ability to organize and execute the courses of action required to successfully accomplish a specific teaching task in a particular context" (p. 233) that is widely used in TSE research studies.

Built on theories of organizational behavior presented by Rotter (1966) and Bandura (1977; 1986; 1993; 1997; 2012), TSE captures the role of a teacher in the classroom and positions them as both the teacher and learner. In this context, teacher agency is developed through ongoing learning of content and pedagogy, where teachers actualize their own learning about teaching practice through their students' learning processes. Over the last several decades of study, TSE has been included on many of the dimensions related to teacher practice, including teacher education, the professional life cycle of teachers, and the relationship between teacher self-efficacy and student self-efficacy (Burić & Kim, 2020; Guo et al., 2012; Klassen et al., 2011; Klassen & Tze 2014; Morris et al., 2017; Shahid & Thompson, 2001; Skaalvik & Skaalvik, 2007, 2010, 2014; Tschannen-Moran & Woolfolk Hoy, 2001; Zee & Koomen 2016).

A significant amount of prior research has theorized that strong levels of teacher self-efficacy leads to higher student academic outcomes (Guo, et al., 2012; Klassen & Tze, 2014; Ross, 1992; Woolfolk Hoy et al., 2009; Zee & Koomen, 2016). However, these findings have not identified clearly linear relationships between TSE and student achievement; rather, they found TSE to be a critical component to effective instructional practice, thereby impacting student achievement (Burić & Kim, 2020; Guo et al., 2012; Klassen & Tze, 2014; Künsting et al., 2016).

Studies of how individual characteristics predict TSE have revealed a wide variety of often conflicting findings. For example, in studying the relationship between gender and TSE, some studies have found little significance (Tschannen-Moran & Johnson, 2011), some have found that females exhibit overall greater TSE (Viesi et al., 2015; Atta et al., 2012; Perera et al., 2019) while other studies have identified differences between male and female TSE specifically relating to classroom management (Klassen & Chiu, 2010; Malinen et al., 2013; Pajares, 1997; Perera et al., 2019; Riggs, 1991; Tschannen-Moran & Johnson, 2011).

Teacher self-efficacy has evolved into an internationally recognized factor in teacher practice and student outcomes. The first twenty years of research was conducted primarily in the U.S. and early efforts to study TSE found cultural discrepancies (Ho & Hau, 2004; Klassen et. al, 2009; Lin et al., 2002). However, in the last twenty years, more research has been conducted internationally with significant research about TSE from countries throughout the world (Scherer et al., 2016). In addition, the use of large-scale international assessments now allows researchers to compare a broader data set, enabling a truly international assessment of TSE (Glassow et al., 2021; Fackler et al., 2021).

As a subject of study, researchers have looked carefully at how TSE is formed and developed across the teacher professional career span. In the consideration of years of experience, the findings are also inconsistent. Several studies identify non-linear relationships between TSE and teaching experience (Guo et al., 2010; Klassen & Chiu, 2010; Morris et al., 2017; Swan et al., 2011; Woolfolk Hoy & Burke Spero, 2005; Zee & Koomen, 2016). Limited studies of teacher education result in similarly mixed findings, primarily due to the variety of scales used and the way teacher education measurement is defined (Darling-Hammond et al., 2002; Fackler et al., 2021; Forsbach-Rothman et al., 2007; Raudenbush et al., 1992). Studies of the interaction between TSE and professional development have yielded more consistent results, finding increases in TSE after participating in specific professional development programs (Althauser, 2015; Dixon et al., 2014; Henson, 2001; Palmer, 2011; Posnanski, 2002; Tschannen-Moran & McMaster, 2009; Yang, 2020).

School context variables are less frequently studied. First, in studies comparing teachers within countries, TSE is positively related to teaching in private schools (Butucha, 2013; Moradkhani & Haghi, 2017; Zamir, et al., 2017). Larger international studies report a weak negative relationship between private school teachers and TSE (Fackler et al., 2021). Another context variable is community size, and the limited research available has indicated a weak to nonexistent relationship between community size designation (i.e., rural, suburban, urban) and TSE (Fackler et al., 2021; Hoy, 2007; Knoblauch & Woolfolk Hoy, 2008; Tschannen-Moran & Woolfolk Hoy, 2007).

The relationship between teacher evaluation and TSE has primarily been studied in the United States and findings indicate that TSE is positively impacted through clear and positive feedback (Mireles-Rios & Becchio, 2018; Schunk & Pajares, 2002; Smith et al., 2020).

Relatively little research exists on school-level use of student assessments and achievement data and its relationship to TSE. However, one study did not find a significant difference in TSE between teachers of high-stakes content areas in a small study of high school teachers (Gonzalez, et al., 2017). At the same time, Skaalnik & Skaalnik (2007) hypothesize that external definitions of education systems may lead to perceptions of teaching efficacy that could limit teacher autonomy and threaten TSE.

Theoretically, country-level variables may also have an impact on TSE. One example is OECD status, which could serve as a mechanism to study whether economic status has a relationship to other variables in the model and TSE. While other studies utilizing large-scale international assessments such as TALIS and TIMSS are cited throughout this research study, none were identified that specifically control for OECD status. Fackler and Malmberg (2016) conclude their study by underscoring the need to include country-level variables to determine whether a relationship between external variables and TSE can be identified.

Another country level variable is student outcomes. Vieluf et al. (2013) utilized country-level PISA scores in a cross-national study of TSE using the 2008 TALIS data and found no relationship between TSE and reading achievement scores. The PISA assessment provides student achievement data in reading, mathematics, and science, but the data is directly associated with specific teachers. Prior research indicates positive relationships are expected between TSE and student achievement but the findings are challenged by the wide variety of measurement tools utilized and the small-scale nature of much of the research (Ashton et al., 1983; Guo et al., 2012; Klassen et al., 2011; Klassen & Tze, 2014; Ross, 1992; Zee & Koomen, 2016).

PISA Teacher Questionnaire

In 2015, 108,292 teachers from 18 countries participated in the PISA Teacher Questionnaire; in 2018, 107,367 teachers from 19 countries participated. The teacher questionnaire expanded from 240 questions about teachers' skills and experiences in 2015 to 311 questions in 2018. The questions included topics such as level of education, professional development, content preparation, technology, and equity. While many of the background questions were the same across the two questionnaires, there were shifts directly related to research around teacher efficacy, content knowledge, and pedagogy. In 2015, PISA included a measurement of teacher self-efficacy in their raw data, but little other information was available about the formula used for the score, nor did published research use or refer to the score. The 2018 PISA included 12 questions specifically relating to teacher self-efficacy. While not directly attributed, the questions most closely match those developed by Tschannen-Moran and Woolfolk Hoy (2001). The responses to these questions provide the opportunity to research relationships between teacher self-efficacy and student achievement with a significant international data set.

Inquiry and Measurement

The 2018 PISA administration included a teacher questionnaire inclusive of survey items using the Bandura-based Teacher Sense of Efficacy Survey (TSES) questions developed by Tschannen-Moran and Woolfolk Hoy (2001). Nineteen countries administered the questionnaire to 62,325 teachers while also measuring student achievement across those countries. The participating countries represented a dichotomy of countries identified as either OECD member countries that met specific standards for economic development, or non-member countries who participated in the PISA assessment but have not met the requirements to join the OECD. This is the first time a large-scale international student achievement assessment has included a measure

for teacher self-efficacy, providing a unique opportunity to explore the relationship between this measure and student achievement on an international scale.

Statement of the Problem

As a theoretical construct, teacher self-efficacy is clearly valued in research and in practice related to teacher efficacy, teacher career development, and student outcomes, but limited data sets have minimized opportunities to compare teacher responses to gain perspectives necessary to examine the many variables that relate to teacher self-efficacy. Previous research has established scales to measure TSE in relation to teacher experience and teacher practice from early childhood classrooms through higher education, at international levels, across content areas, and within various social contexts. Across the research, conceptual differences, variations in scales of measurement, and factor calculations have led to inconsistencies and mixed findings, most particularly in the area of teacher characteristics such as gender, teacher experience, and teacher preparation. Some of the largest studies of TSE have been dismissed due to disagreement about the instrument. In 1984, Gibson and Dembo introduced the Teacher Efficacy Scale (TES). For many years, the TES was the primary instrument used to gather data about teacher self-efficacy. As researchers looked to Bandura's theories, they found Gibson and Dembo's instrument flawed and argued against its validity to measure teacher self-efficacy (Brouwers & Tomic, 2001; Tschannen-Moran & Woolfok Hoy, 2001; Denzine et al., 2005). In addition, one of the most common challenges to the research around TSE is limited sample size, with many studies focusing on small-scale studies (Morris et al., 2017; Zee and Koomen, 2016; Klassen et al., 2009).

In addition to studying teacher responses to self-efficacy, there is limited research about how variations in teacher perspectives of their own practices may relate to supporting and

developing teachers. From the standpoint of a practitioner, this study presents an opportunity to look at how teacher perspectives of self-efficacy shift may be influenced by school and government contexts. It also presents an opportunity to question what education systems value in terms of teacher self-efficacy and how they arrive at understandings and beliefs that impact policy development and leadership practices.

Purpose and Research Questions

The PISA Teacher Questionnaire and Principal Questionnaire provide an opportunity to study a significant international sample of teachers and multi-level factors gathered across countries and economies. The purpose of this study is to examine the relationship between teacher self-efficacy, as reflected in PISA teacher questionnaire responses, and multiple levels of factors identified from the 19 countries who administered the teacher questionnaire. Understanding how different aspects of the school, region, and governmental environments influence TSE provides additional research to support school-level practitioners and state and federal policymakers as they develop, implement, and evaluate programs aimed at teacher growth and development, and, ultimately, student learning and achievement. The following research questions guide the analysis of the 2018 PISA data.

Research Questions

This study utilizes data gathered in the 2018 PISA General Teacher Questionnaire and the 2018 School Questionnaire to investigate the following questions:

RQ1: What is the relationship between individual characteristics and teacher self-efficacy (TSE)?

- How do individual characteristics such as age, gender, work experience, and completion of a teacher training program relate to TSE?

- How do individual professional development experiences relate to TSE?

RQ2: How do individual characteristics and school-level factors influence teacher self-efficacy (TSE)?

- How are the teacher-level responses influenced by characteristics of the school such as size and school type (public/private)?

- How are the teacher-level responses influenced by school-level environmental experiences such as the principal's perception of school's instructional capacity, and the quality assurance (accountability) approaches?

- How are the teacher-level responses influenced by school level student achievement?

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)?

- How is TSE influenced by a country's economic development status as defined by OECD status?

- Is there a relationship between the country's level of student achievement and TSE?

Null Hypotheses

H_1: Individual teacher characteristics do not influence TSE.

H_2: Individual experiential factors do not influence TSE.

H_3: School-level factors do not influence TSE.

H_4: School-level environmental factors do not influence TSE.

H_5: Country-level factors do not influence TSE.

Significance of the Study

In the United States, following the introduction of PISA and other international assessments, policy makers incentivized states to introduce policies tied to measuring teacher

quality (Breakspear, 2012; OECD, 2018; Weisberg et al., 2009; USDOE, 2009). In an effort to access billions of dollars offered through the Race to the Top Fund (RTTT), one of the largest federal competitive grants in history, state governments set policy to increase requirements for measuring and reporting teacher quality (Hallgren et al., 2014). In 2015, the Every Child Succeeds Act (ESSA) loosened federal evaluation requirements, allowing states more latitude in their choices for teacher evaluation. While some states shifted the way evaluation data was calculated or the use of standardized assessment results, the directives regarding the use of approved evaluation frameworks and reporting requirements have largely remained (Close et al., 2020; Jones et al., 2017). Studying the relationship between teacher self-efficacy and school-level and country-level variables may shed further light on how evaluation tools and teacher quality frameworks can be more effectively used to support professional learning and development of teacher self-efficacy. For practitioners in the field, this research is aimed at challenging perceptions around the use of teacher self-efficacy tools and metrics, and the importance of having an in-depth understanding of the complexity of adult learning and development.

Limitations

Self-efficacy in the context of education has provided a vast history of research, and over the last forty years there have been ongoing debates about studying the relationship between teacher self-efficacy and student achievement. This study does not attempt to establish any type of causality. Rather, it seeks to identify if a relationship using the 2018 PISA data exists, and if so, to pose questions about how that relationship may impact our understanding of how teachers view their own skills and development as well as the external factors that may impact TSE. The PISA teacher questionnaire and principal questionnaire represent a large international data set

based entirely on self-reported information and are therefore subject to social desirability response bias. There is significant room in the research to study how different characteristics also impact the relationship, including teacher age, technology expertise, and content knowledge. While these provide ample opportunities for further research, they are not used in the analysis presented.

Delimitations

This paper revolves around the responses of 107,367 teachers across 19 countries to a 12-question survey. While other survey questions were asked that have similarities to concepts of self-efficacy, this study focuses only on questions that were related to the Tschannen-Moran and Woolfolk Hoy (2001) TSES. Future research may look at other survey responses to identify whether relationships with TSE exist within and between those responses.

Definitions

Self-efficacy is the extent to which an individual believes they have the capacity and ability to fulfill specific tasks. Originally aimed at addressing phobias, Bandura (1977) proposed that an individual's self-efficacy impacts their ability to identify and address personal behaviors. In this study, it is treated as a multi-dimensional construct applied to teacher practice.

Teacher self-efficacy (TSE) is an extension of the theory of self-efficacy proposed by Bandura (1977). TSE is defined as personal belief in his or her capability to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context (Tschannen-Moran & Woolfolk Hoy, 2001, p. 233). Teacher self-efficacy applies the concept directly to the practice of teaching. Researchers have studied the concept of TSE in relation to student learning, motivation, discipline, achievement, development of student self-efficacy (Caprara et al., 2006; Gibson & Dembo, 1984; Pajares, 1997; Schunck & Pajares, 2002;

Tschannen-Moran, Woolfolk Hoy & Hoy, 1998; Woolfolk Hoy, Hoy, & Davis, 2009). Teacher self-efficacy has also served as a primary variable in the study of teacher practice, job satisfaction, teaching effectiveness, teacher professional development, and collective self-efficacy in educational organizations (Goddard et al., 2000; Klassen & Tze, 2014; Malinen et al., 2011; Tschannen-Moran et al., 1998; Zee & Koomen, 2016).

Bandura (2006) argued that scales for measuring self-efficacy must be specific to the context and appropriate to the constructs being measured. In this study the **Constructs of Teacher Self-Efficacy**, as delineated by Tschannen-Moran and Woolfolk Hoy (2001), are classroom management, instructional strategies, and student engagement. **Classroom management** is a construct defined as the management of student behaviors through specific teaching practices. In the Tschannen-Moran and Woolfolk Hoy (2001) instrument, classroom management is identified as a construct critical to teacher self-efficacy. **Instructional strategies** constitute a construct that includes use of assessments, crafting questions, teaching flexibility, and measuring student comprehension (Tschannen-Moran & Woolfolk Hoy, 2001). **Student engagement** is a construct built around the teacher's ability to develop and support student learning behaviors in the context of the classroom. In the Tschannen-Moran and Woolfolk Hoy (2001) instrument student engagement included building student beliefs in learning, motivating students, and developing critical thinking skills.

Student Achievement in this research is defined by the raw scores generated through the administration of the 2018 PISA. The framework of the PISA measurement is based on a baseline of performance on reading, mathematics, and science in addition to academic cross-content competencies such as critical thinking and problem solving at a common level of secondary education. "To do well in PISA, students have to be able to extrapolate from what

they know, think across the boundaries of subject-matter disciplines, apply their knowledge creatively in novel situations and demonstrate effective learning strategies" (OECD, 2019a, p. 5). The scores are based on the mean at the country level, and the mean scores are translated into a range of levels from one to six, with one representing basic skill mastery and six representing high levels of comprehension and critical analysis (OECD, 2019a).

Teacher evaluation, for the purposes of this research, is defined as any process relating to teacher content knowledge and pedagogical skills that is evaluative in nature. Evaluations in this context impact how teachers understand their own practice, how administrators judge practice, and how professional development is provided to support teacher practice. Evaluation instruments may be commercial products or developed by governmental entities.

Organization of Study

The purpose of this study is to identify the impact of external socio-political contexts, internal school-level structures and philosophies, and teacher-level experiences and beliefs on teacher self-efficacy using the 2018 PISA teacher and school questionnaires. Chapter II presents the literature review for the study, providing a historical context for the development of measurement tools and gaps in the research around the relationship between student achievement and teacher self-efficacy. Chapter III outlines the quantitative methodology used to analyze the PISA data. Chapter IV provides an analysis of the research findings. Chapter V examines the implications of the findings, how they may be used to impact policies related to developing teacher self-efficacy and teacher evaluations, as well as how practitioners may use the findings to support teacher development.

CHAPTER II: LITERATURE REVIEW

Humans are inherently social creatures developed through a complex combination of biological and social processes. Throughout the mid-twentieth century, psychologists worked to identify the learning processes related to human development and reasoning. In 1954 Rotter introduced social learning theory, utilizing the concept of "expectancy" to explain how behaviors are formed and shift based on perceptions of task types and outcomes that positively or negatively reinforce the expectancy and behavior. In 1977 psychologist Alfred Bandura introduced "Social Cognitive Theory," expanding on Rotter's work and introducing his theory of self-efficacy. Bandura defined self-efficacy (1977, 1997) as a theoretical construct referring to "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (1997, p. 3).

Over forty years later, the study of self-efficacy continues to drive researchers across public and private sectors. A search of scholarly articles using the search term "self-efficacy" returns over 100,000 articles published in a single year (2020). Of the first 25 articles returned in the search, 13 directly referenced education, and the remaining referenced a wide variety of topics including financial technology, the COVID-19 pandemic, psychological behavioral therapies, and sales performance.

This study utilizes the Programme for International Student Assessment (PISA), a large-scale international assessment of student skills in mathematics, reading, and science. In addition to the assessment questions, the PISA provides a questionnaire to school leaders to provide information regarding the school environment. Of the 59 countries that participated in the 2018 administration of the PISA, 19 also distributed a teacher questionnaire that included demographic data and teacher survey questions. The questions utilized included a set measuring

teacher self-efficacy (TSE) via the Teacher Sense of Efficacy Survey (TSES), which was developed through the research of Tschannen-Moran and Woolfolk Hoy (2001). Together, the principal and teacher questionnaire provide a rich data set that encourages an international comparison of teacher self-efficacy. Specifically, the large amount of data represents an opportunity for a multi-level study of the relationship between TSE and factors at the individual teacher level, the school level, and the country level.

This research began with a review of the initial behavior psychology work of Rotter and Bandura. The literature chosen initially focused on the early studies around education and teacher self-efficacy, along with the tools created to measure teacher self-efficacy. The use of Google Scholar as a primary search engine linked to the Seton Hall University databases allowed for advanced searches that included *teacher self-efficacy* combined with terms such as *student achievement, professional development, demographics, teacher careers, policy,* and *assessment*. While there has been a significant increase in international study of this topic, this review utilized articles available in English and prioritized articles published in peer-reviewed journals. In cases of specific authors, book chapters were utilized to provide further theoretical understanding.

Self-Efficacy

 In terms of behavioral theory, Bandura (1977; 1986; 1997; 2012) posits self-efficacy as central to human agency, and to the choices and actions individuals make on a daily basis based on perceptions of their personal ability to carry out specific actions. Self-efficacy differs from Rotter's (1954) expectancy theory in that self-efficacy is dependent on a person's ability to effectively appraise the limits of their operative capabilities and is not dependent on performance feedback (Bandura, 1986, p. 363). While Bandura's initial research was applied to addressing

phobias, it has since been employed to identify how behaviors are developed and changed across

sectors (Bandura 1977; 1997; Bandura et al., 1980).

As exemplified in the concept of "triadic reciprocal causation" depicted in Figure 1.1,

Bandura frames human agency around the "interplay of intrapersonal influences, the behavior

individuals engage in, and the environmental forces that impinge upon them" with self-efficacy

being central to the intrapersonal influences (Bandura, 1986; 2012). Bandura explains that socio-

structural and personal determinants are critical to the structure as a whole (Bandura, 1997;

2012). In this model, the triadic reciprocal causation leads to decisions and actions, and self-

efficacy impacts how much effort people will expend and how long they will persist in the face

of obstacles and aversive experiences (Bandura, 1977, p. 193).

Figure 1.1 Triadic Reciprocal Causation (Bandura, 1986)
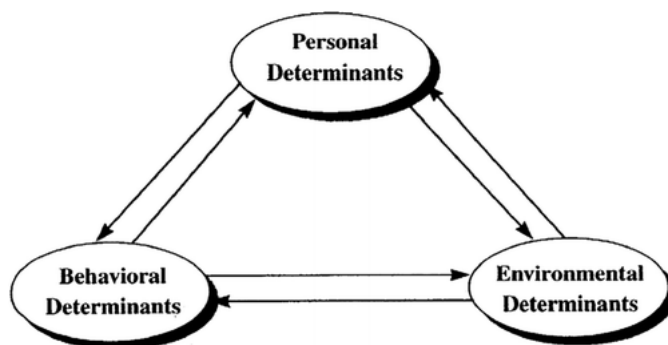
The development of self-efficacy is continuous and multi-faceted. Perceptions of self-

efficacy begin to be constructed during infancy and continue through adulthood and are impacted

by "efficacy promoting influences" (Bandura, 1997, p. 169). Bandura argues such influences are

created through four types of experiences impacting the development of an individual's self-

efficacy: performance accomplishments and mastery, vicarious experiences, verbal/social

persuasion, and emotional arousal or affective sources (1977; 1997). Such experiences occur

continuously, from a child learning how to form words and communicate their needs through

adult development of specific professional skills. Self-efficacy may be developed through what Bandura identifies as "personal enablement" and is achieved by providing the appropriate knowledge, skills, and positive experiences that enhance personal control (1997).

Self-efficacy may be conflated with self-esteem, but there are critical differences between the two. Bandura notes that self-esteem is a different construct built on judgement of self-worth as opposed to judgement of personal capacity. While an individual's self-esteem may impact the perception of personal efficacy, the two operate independently (Bandura, 1997).

Education and Self-Efficacy

Formal school experiences play a significant role in the development of personal efficacy, with efficacy beliefs identified as having a significant role in student cognitive development (Bandura, 1997; Bandura & Schunk, 1981; Pajares & Miller, 1994; Schunk, 1984, 2003; Zimmerman, 2000). Schools serve an essential role in providing the structures and opportunities for students to develop the cognitive and self-regulatory skills necessary for future success (Bandura, 1997). Student perception of self-efficacy has a significant impact on student academic achievement, providing a basis upon which students develop the ability to persist in challenging academic tasks across different grade levels (Bandura & Schunk, 1981; Pajares & Miller, 1994; Schunk, 1984, 2003).

Teacher practice plays a significant role in students' individual development of self-efficacy (Schunk, 2003; Walker, 2003). As noted in Schunk's (1984) research around student mathematical ability, students challenged to build mathematical skill sets were more successful when they were rewarded for successful performance of tasks as opposed to being rewarded for only task completion. Pedagogical practices, including use of student choice, goal setting, modeling, and effective feedback are all noted as impacting student self-efficacy in academic

settings (Locke & Latham, 1990; Pajares & Johnson, 1994; Schunk, 2003; Woolfolk Hoy et al., 2009; Zimmerman, 2000).

In the educator sector, the multi-dimensional nature of the construct provides a broad basis to study all elements of the profession including teacher training, induction, motivation, professional development, and classroom practice (Guo et al., 2012; Klassen et al., 2011; Klassen & Tze, 2014; Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001; Zee & Koomen, 2016). Over forty years after its introduction, researchers across the social and behavioral disciplines continue to study the powerful construct of self-efficacy to support growth and development for individuals and organizations.

Teacher Self-Efficacy

Teacher self-efficacy (TSE) is an extension of the theory of self-efficacy proposed by Bandura (1977). Tschannen-Moran et al. (1998) define TSE as a personal belief in his or her capability to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context (p. 233). Researchers have studied the concept of TSE in relation to student learning, motivation, discipline, achievement, and development of student self-efficacy (Woolfolk Hoy, Hoy, & Davis, 2009; Schunck & Pajares, 2009; Caprara et al., 2006; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Pajares, 1997; Gibson & Dembo, 1984). Teacher self-efficacy has also served as a primary variable in the study of teacher practice, job satisfaction, teaching effectiveness, teacher professional development, and collective self-efficacy in educational organizations (Zee & Koomen, 2016; Klassen & Tze, 2014; Malinen, et al., 2013; Goddard, Hoy & Woolfolk Hoy, 2000; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998).

Multiple meta-analyses have been conducted to study the large amount of research generated around the topic of TSE. Shahid and Thompson (2001) identified 89 studies with 973 research hypotheses in an early effort to synthesize findings on prior TSE studies. Klassen et al. (2011) reviewed 218 studies conducted between 1998 and 2009 with the purpose of identifying key findings, gaps in research, and guidance for future research. Klassen and Tze (2014) reviewed 43 studies specifically related to the relationship between teacher psychological characteristics and TSE. Morris, Usher, and Chen (2017) conducted a meta-analysis of research focused on the sources of TSE between 1977 and 2015, finding 82 empirical studies meeting their specific research domain. Zee and Koomen (2016) identified over 2,000 articles in peer-reviewed journals between 1976 and early 2014, narrowing down their synthesis into 165 quantitative studies focusing on teachers and self-efficacy. This rich body of research provides a strong reference point for specific areas of teacher self-efficacy.

Measuring Teacher Self-Efficacy

Throughout the research there has been an ongoing evolution in the survey instruments used along with a continuous disagreement in the construction of survey tools (Dellinger et al., 2008; Klassen et al., 2011; Labone, 2004; Morris et al., 2017; Shahid & Thompson 2001; Wheatley 2005; Wyatt, 2014). The first survey tool comprised two questions as a part of the 1974 RAND study of Los Angeles public school reading program (Armor et al., 1976). The two questions posed using Rotter's general expectancy theories spurred ongoing debate about theoretical constructs, domain development and definition, differing variables, and almost every other facet of study design.

After the RAND survey findings were shared, a variety of scales were presented that proposed to measure TSE (Tschannen-Moran, et al., 1998). Among those, Rose and Medway's

(1981) Teacher Locus of Control, Guskey's (1981) Responsibility for Student Achievement, Webb Efficacy Scale (Ashton et al., 1983), and Ashton Vignettes (Ashton et al., 1984) were found faulty in appropriately validating the construct (Tschannen-Moran et al., 1998; Henson, 2001).

Gibson and Dembo (1984) began a deeper study of teacher efficacy and presented an instrument explicitly for the study of teacher efficacy. Gibson and Dembo's (1984) Teacher Efficacy Scale (TES) comprised 30 items measuring teacher efficacy, verbal ability, and flexibility across the domains of teacher expectation and classroom outcomes (p. 569). The study developed two modes of efficacy: personal teaching efficacy and general teaching efficacy. Findings from the study were specific to teacher practice, including use of small groups, student engagement, teacher feedback, and high expectations of student learning. Early research of teacher self-efficacy leaned heavily on the use of the TES (Henson, 2001; Ross, 1992; Tschannen-Moran et al., 1998).

As the research increased, the TES was criticized as overly general, violating the domain specificity outlined by Bandura, and not addressing the theoretical basis of self-efficacy (Coladarci & Fink, 1995; Denzine et al., 2005; Guskey & Passaro, 1994; Henson, 2001; Labone, 2004; Soodak & Podell, 1996; Tschannen-Moran et al., 1998). Studies also identified challenges with discriminant validity and found the tool not suitable for obtaining precise and valid information about teacher efficacy beliefs (Brouwers & Tomic, 2001, 2003; Coladarci & Fink, 1995; Henson, 2001).

Brouwers and Tomic (2003) concluded the TES is not suitable for obtaining precise and valid information about teacher efficacy and suggested adaptation of the scale was necessary. Denzine et al. (2005) reported similar issues with the TES and encouraged the abandonment of

previous evidence rather than a re-analysis of the data. Despite those findings, the TES continued to be used in research studies, many of which are cited in this review of literature.

Bandura (1997) argued that scales measuring teacher efficacy are too generalized to appropriately predict student achievement and that scales of teacher self-efficacy should be closely tied to specific content knowledge domains. In a guide to creating self-efficacy scales, Bandura (2006) cautioned researchers to develop scales that accurately reflect the perceived capability of successfully performing difficult tasks, are relevant to the specific domain of functioning, and reflect gradations of the strength of an individual's perception. Bandura (not dated; 2006) presented his iterations of a measurement tool, with versions ranging from 28 to 30 indicators. Ratings included a 9-point scale and a 0 to 100 scale with 0 representing complete inability to complete the actions described and 100 being highly certain of the ability to complete the actions (Bandura, 1997; 2006). Bandura's scales presented constructs of the teaching experience that could be measured separately, responding to the need to provide additional specificity to the measurement tool. While researchers provided commentary on Bandura's scales, they were not widely used or referenced in the research reviewed, nor is there available information about their reliability or validity (Tschannen-Moran & Woolfolk Hoy, 2001).

Tschannen-Moran and Woolfolk Hoy (2001) created the Teacher Sense of Efficacy Survey (TSES), originally titled the "Ohio State Teacher Efficacy Scale," after an extensive review of prior instruments used to collect information regarding teacher self-efficacy, noting issues with balancing specificity and generality along with challenging interpretations of the factor structures and correlations. Utilizing the concepts of domain constructs modeled by Bandura's scales, TSES provides a three-dimensional survey tool to measure teacher self-efficacy in relation to three factors: efficacy for student engagement, efficacy for instructional

strategies, and efficacy for classroom management (Tschannen-Moran & Woolfolk Hoy, 2001).

The tool was constructed after multiple iterations studying teachers in the vicinity of Ohio State

University and was tested on 410 participants, with initial findings indicating the measure to be

reasonably valid and reliable.

The survey questions, both in their long (24 questions) and short (12 questions) versions,

showed satisfactory reliability and construct validity across multiple studies (Fives & Buehl,

2009; Nie et al., 2010; Tschannen-Moran & Woolfolk Hoy, 2001; Zee & Koomen, 2016).

Klassen et al. (2009) provided a cross-national perspective, supporting the prior findings of

reliability and construct validity in five countries. Similar to other international studies, teachers

in East Asian countries (Korea and Singapore) reported lower teacher self-efficacy ratings but

maintained convincing invariance using the scale (Ho & Hau, 2004; Nie et al., 2010).

International Large-Scale Assessments

Teacher self-efficacy has been a construct measured in recent administrations OECD-led

international large-scale assessments; however, there has been a shift from use of the Gibson and

Dembo (1984) four-item TES scale to the use of the twelve-item Tschannen-Moran and

Woolfolk Hoy (2001) TSES. One example is the OECD-sponsored Teaching and Learning

International Survey (TALIS), administered to learn about teaching conditions and experiences

with the goal of providing information for policy development. The 2008 TALIS utilized the

TES items, while the 2013 TALIS shifted to the TSES items. In both cases, researchers

performed construct variation and found TSE to be generalizable across countries (Fackler et al.,

2021; Scherer et al., 2016; Vieluf et al., 2013). In comparing the three constructs of the TSES

questions, OECD (2014a; 2014b) found the three constructs were overall reliable, but also

cautioned that the mean scores could have different meanings between countries.

Glassow et al. (2021) studied TSE in relationship to the Trends in International Mathematics and Science Study (TIMSS) assessment, led by the International Association for the Evaluation of Educational Achievement. It included a nine-item survey of TSE, adapted from the TES tool, with three questions specific to the teaching of mathematics. The study found TSE as a construct could be validly compared internationally (Glassow, et al., 2021).

While PISA 2015 utilized the same questions as the TALIS 2009, there was no evidence of its use in the study of TSE, nor any other references to student achievement being used in relationship to TSE. The rich data gathered through the administration of the PISA 2015, along with its use of the three-construct TSES model, presents a unique opportunity to provide analysis of the teacher factors that have been previously considered, as well as new factors that have not previously been considered at the school and country levels.

Use of the PISA Assessment

The Programme for International Student Achievement (PISA) was first administered by OECD in 2000 and has been administered on a triennial basis since, with the last administration occurring in 2018. Since then, it has served as a global tool for evaluating and comparing educational systems. It has also served as a catalyst for policymaking, impacting nations' development of assessment programs, and norm setting for student achievement (Bieber & Martens, 2011; Breakspear, 2012; OECD, 2020). Education policy leaders such as Darling-Hammond (2014) argue that the United States should look to high achieving nations for policy initiatives including investment in strong teacher education programs.

In 2015, 108,292 teachers from 18 countries participated in the survey; in 2018, 107,367 teachers from 19 countries participated. The teacher questionnaire expanded from 240 questions about teachers' skills and experiences in 2015 to 311 questions in 2018. The questions include

topics such as level of education, professional development, content preparation, technology, and equity. While many of the background questions were the same across the two questionnaires, there were shifts directly related to research around teacher efficacy, content knowledge, and pedagogy. In 2015, PISA included a measurement of teacher self-efficacy in their raw data, but little other information was available about the formula used for the score, nor did published research use or refer to the score. The 2018 PISA included 12 questions specifically relating to teacher self-efficacy. The questions are aligned to the 2015 administration of the TALIS (OECD, 2014a), and while direct attribution was not made in the PISA technical guidance, the questions are identical to questions developed by Tschannen-Moran and Woolfolk Hoy (2001). The responses to these questions provide the opportunity to research relationships between teacher self-efficacy and student achievement with a significant international data set.

The breadth of prior research regarding TSE provides a unique opportunity to approach the PISA data gathered from a different perspective than the smaller scale studies that make up a significant amount of the literature. The remainder of this literature review is organized around the PISA data that will be analyzed in this study. As noted in the introduction, this study will utilize a multi-level study of the relationship between TSE and factors at the individual teacher level, the school level, and the country level.

Individual Teacher Factors

Individual teacher characteristics such as gender, years of teaching experience, and identification of prior professional development experiences serve as personal, or micro-level variables influencing TSE.

Gender

From an international perspective, gender roles are often defined on cultural norms. However, the use of the PISA data allows for an in-depth international consideration of the role gender might play in TSE. An international array of studies finds inconclusive patterns related to gender and TSE. In Iran, India, and Australia female teachers were identified as having higher TSE (Atta et al., 2012; Perera et al., 2019; Veisi et al., 2015). Klassen and Chiu (2010) found male Canadian teachers with a slightly higher TSE related to the classroom management construct; this was similar to the Malinen et al. (2013) study, which found male teachers in Finland demonstrated higher TSE related to classroom management, but that the same patterns were not identified from teachers in China or South Africa. Riggs (1991) found male science teachers with higher TSE than female science teachers, raising questions of the relationship to content experience and gender. In a U.S. study of predominately female fifth grade literacy teachers, Tschannen-Moran and Johnson (2011) found gender contributed a small variance to TSE. In two international studies, findings included positive relationships between female teachers and TSE (Fackler et al., 2021; Perera et al., 2019) found a positive relationship between female teachers and TSE in three individual constructs. A number of studies have also found non-significant variance in gender and TSE (Darling-Hammond et al., 2002; Shoulders & Krei, 2015; Tschannen-Moran & Woolfolk Hoy, 2007).

Age

As an individual characteristic, teacher age may often be aligned directly with teacher experience. However, similar to other individual characteristics, the findings related to age are mixed. Considered as a single variable, Colodarci and Breton (1997) found age was related to

teacher efficacy. Tschannen-Moran and Woolfolk Hoy (2007) found age was not significantly related to TSE for either novice or career teachers.

Teaching Experience

As noted by Klassen and Chiu (2010), studies of teacher experience and TSE depict non-static and non-linear relationships. Morris et al. (2017) found positive relationships primarily in research regarding mentoring and pre-service to early career teachers. Swan et al. (2011) found teacher TSE was highest pre-service, declined by the conclusion of the first year of teaching, increased in the second year, and experienced another slight decline in the third year. Woolfolk Hoy and Burke Spero (2005) reported similar findings in a study of TSE between the student teaching experience and early career teaching. Haverback and Parault (2011) suggest that it is a benefit for pre-service teachers to enter the teaching profession with a lower TSE, as it may indicate more realistic expectations of the occupational challenges. Research of pre-service teachers indicates high levels of TSE correspond with career longevity but found weaker relationships between practicing teachers' decisions to stay in the field (Zee & Koomen, 2016).

Tschannen-Moran and Woolfolk Hoy (2007) used the TSES instrument to measure differences in early and late career teachers and found small differences in the classroom management and instructional strategies constructs, but no difference in TSE with respect to the construct of student engagement. Burić and Kim (2020) found that years of teaching experience exhibited a significant negative correlation with TSE.

One challenge in the research is the equation of years of teaching with direct task-related experience. Labone's (2004) analysis of prior research found that the accuracy of a teacher's judgement of their self-efficacy is related to their experience with the task, but in the case of classroom teachers, the task may shift based on a change of grade level or a specific content area

(Klassen & Chiu, 2010; Klassen & Tze, 2014; Morris et al., 2017; Tschannen-Moran & Johnson, 2011).

Klassen and Chiu (2010) suggest that the most significant weakness in the research around teacher experience and TSE involves the lack of granular data. Most of the studies utilize grouping data, identifying teachers as either novice or experienced, or with very specific ranges that do not allow for a nuanced understanding of when TSE is most likely to evolve and under what conditions such changes in TSE may occur.

Teacher Education

In the thousands of studies relating to teacher self-efficacy, relatively few were identified that specifically compared a measure of TSE with different types of teacher education programs. Raudenbush, Rowan, and Chong (1992) found teachers' level of education had no effect on TSE. Darling-Hammond, Chung, and Frelow (2002) found novice teachers that entered the profession from alternative pathways felt less prepared, as indicated through a correlation between teacher academic preparation and TSE. Forsbach-Rothman, Margolin, and Bloom (2007) found novice teachers who engaged in an undergraduate teacher education that included fieldwork and mastery experiences had higher levels of TSE than teachers who went directly to graduate school or enrolled in an alternative pathway to the teaching certification. Of the three studies, Darling-Hammond et al. (2002) and Forsbach-Rothman et al. (2007) utilized questionnaire items from the RAND study (Gibson & Dembo, 1984). All three of the studies were conducted in the United States. Fackler et al. (2021) utilized the 2008 TALIS international assessment data to evaluate the relationship between teacher characteristics and TSE, using the three-construct model proposed by Tschannen-Moran and Woolfolk Hoy (2001). At an international level, Fackler et

al. (2021) found teachers' level of education was only significant to one of the three TSE constructs measured.

Professional Development

There is a significant body of research centered on the relationship between professional development and teacher self-efficacy. For the purposes of narrowing down the studies, the search terms excluded any research that involved "pre-service" teachers, focusing rather on teachers engaged in professional practice. Findings across the body of research are largely consistent. Teachers that participate in mastery experiences that involve research, coaching, and feedback have positive relationships with TSE (Althauser, 2015; Dixon et al., 2014; Henson, 2001; Palmer, 2011; Posnanski, 2002; Tschannen-Moran & McMaster, 2009). Positive relationships between amount of professional development and TSE were also identified (Dixon et al., 2014; Yang, 2020). Tschannen-Moran and McMaster (2009) also identified small decreases in TSE after initial introduction to new instructional strategies, similar to the shift in TSE from pre-service to practicing teaching. All of the findings indicate that TSE can be fluid in the context of specific adult learning. Unlike other areas, there appeared to be a more significant number of studies related to professional development and TSE conducted in the United States, and none were identified that included a cross-cultural comparison.

Teacher Classroom Practice

Throughout the body of research, TSE is framed as having an indirect relationship to student achievement, with teaching practices and teaching quality as a mediating factor. A significant number of studies have identified a positive relationship between TSE and instructional quality (Burić & Kim, 2020; Gibson & Dembo, 1984; Guo et al., 2012; Guskey, 1988; Klassen & Tze, 2014; Morris et al., 2017; Shahid & Thompson, 2001; Tschannen-Moran

et al., 1998; Woolfolk Hoy et al., 2009). When specifically measuring shifts in TSE related to teacher knowledge and skill, there is evidence that TSE is positively related to a teacher's confidence in their understanding of the content and pedagogical skills (Morris et al., 2017). Caprara et al. (2006) identified a relationship between TSE, teacher job satisfaction, and teachers' perception of their impact on student academic achievement and found a modest relationship between student academic achievement and TSE. Klassen and Tze (2014) found TSE has a stronger relationship to teaching effectiveness than to student achievement.

In a study of German teachers, Künsting et al. (2016) found TSE was predictive of classroom climate and classroom management, using a longitudinal model to identify a positive relationship between long-term stability of TSE and higher instructional quality. Multiple studies have also found relationships between TSE and the use of specific classroom practices such as differentiation (Dixon et al., 2014), science strategies (Palmer, 2011; Posnanski, 2002), reading strategies (Tschannen-Moran & McMaster, 2009), and mathematical thinking strategies (Carney et al., 2016).

School Level Factors

There are a wide variety of environmental, or macro-variables that may impact teacher experiences and TSE, including the type of school and the size of the community. Whether a school is identified as a public or private school has been studied more extensively outside of the United States, and country-specific research studies of Iranian, Pakhastani, and Ethiopian teachers have found private school teachers demonstrate higher levels of TSE (Butucha, 2013; Moradkhani & Haghi, 2017; Zamir et al., 2017). In Singapore, whether a school was academically selective was significant to TSE (Chong et. al, 2010). Larger international studies report a weak negative relationship between private school teachers and TSE (Fackler et al.,

2021). The size of the community, or its designation as urban, suburban, or rural appears to have little relationship to TSE. Research findings have ranged from weak to no evidence of relationships between the setting (urban, suburban, or rural) and TSE (Fackler et al., 2021; Tschannen-Moran & Johnson, 2011; Tschannen-Moran & Woolfolk Hoy, 2007). In a study of pre-service teachers, Knoblauch and Woolfolk Hoy (2008) found no difference between the external setting's relationship to early development of TSE; however, they suggest that external influences, such as public perception of educational quality, may impact TSE in different contexts, thereby impacting TSE in different settings.

A variety of other school-level factors have been considered in relation to TSE. Ross et al. (1996) found that teacher perceptions of their own practice varied based on the teaching assignment they were given and whether they were placed in leadership roles. Norwegian studies found school context variables such as autonomy, time pressure, and teacher relationship to parents have relationships to TSE (Skaalvik & Skaalvik, 2007; 2010; 2014). A variety of studies have identified relationships between the school climate and TSE (Chong et al., 2010; Künsting et al., 2016).

School Leadership

In the first study of TSE, Armor et al. (1976) proposed that school level policies have a mediating impact on teacher morale and commitment. It is expected that the general organization of most school environments posit the principal/school as a critical element in school success, and over the last twenty years research identifying the relationships between school success and school leadership has emerged to support that hypothesis (Bendikson et al., 2012; Hallinger & Heck, 1996; Marks & Printy, 2003; Sehgal et al., 2017). While a significant amount of study on

the principal-TSE dynamic centers on the principal's impact on collective efficacy, this brief review will focus on the impact at the individual teacher level.

Several influential studies have identified relationships between leadership style and student achievement. Marks and Printy (2003) identified differences in leadership styles, inclusive of how a principal values teacher knowledge and skill, resulted in differences in student achievement. Bendikson et al. (2012) found relationships between direct and indirect styles of leadership and high performing schools.

School leaders, and their beliefs about teachers, have a relationship to teachers' feelings of self, thereby impacting TSE as a mediating influence. Nir and Kranot (2006) identified a positive relationship between school leadership style and personal teaching efficacy in Israeli schools. Kurt, Duyar, and Çalik (2012) identified a positive relationship between transformational leadership behaviors and TSE and a negative relationship between transactional leadership and TSE in Turkish primary schools. In a qualitative study of Canadian teachers, Lambersky (2016) found that principal leadership behaviors had a positive relationship to teacher emotions and feelings of success in the workplace. Sehgal et al. (2017) found a positive relationship between principal leadership and TSE in a study of teachers in India, noting that the principal has a relationship with teacher self-perceptions, thereby impacting TSE.

Insight to the relationship between school leaders' perspectives of general skill and knowledge provides an opportunity to study the relationship between school leaders and TSE. The PISA principal questionnaire asks several questions that represent the principal's perspective of the teachers' general efficacy and provides an opportunity to measure those perspectives in the context of TSE.

Assessments, and Evaluation

   Student achievement has been identified as a correlate and antecedent to TSE (Caprara et al., 2006; Chong et al., 2010; Guo et al., 2012; Shahid & Thompson, 2001). How student achievement data is used in the public sphere may have an impact on TSE, but there are few peer-reviewed studies available that examine relationships between the use of assessments in school contexts and TSE. Gonzalez et al. (2017) found no significant difference in TSE for Texas teachers who taught in classes that were measured through high-stakes standardized assessments. Von der Embse et al. (2016) found teachers with higher levels of TSE were less impacted by standardized test-based accountability systems than teachers with lower levels of TSE. There was no literature identified that compared the use of assessments and TSE in an international context.

   Policy measures focused on increasing teacher quality have been focused on improving teacher evaluation processes. In a meta-analysis of 43 studies, Klassen & Tze (2014) found TSE was strongly associated with evaluated teaching performance. As a resource, social persuasion has been primarily researched through the lens of evaluation and feedback, with instructionally credible and specific feedback from observers, including students, to have positive correlations with TSE (Morris et al., 2017). Palmer (2011) found that forms of professional development that impacted TSE included observation, coaching, and feedback. Mireles-Rios and Bechio (2018) found higher levels of TSE when a pre- and post-evaluation conference were used to set goals and included positive feedback. Schunk and Pajares (2002) identified a relationship between TSE and affirming, positive feedback. Smith et al. (2020) identified a positive relationship between teachers who received highly specific feedback and TSE.

Skaalvik and Skaalvik (2010, 2014) hypothesize that external controls are largely related to public perceptions of teaching efficacy and norms for establishing teacher practice standards may undermine autonomy, thereby negatively impacting TSE. In a qualitative study of Louisiana teachers, Ford et al. (2017) found diminished teacher self-efficacy after two years of teacher evaluation and high stakes testing accountability policies were enacted. While use of standardized assessments, external reporting of student progress, and external evaluations are used throughout the world to strengthen teacher quality, their impact on TSE is largely theoretical. The PISA questionnaires provide a significant opportunity to study the use of the assessments, as reported in the principal questionnaire, in relationship to TSE.

Country-Level Factors

While other studies of large-scale international assessments such as (TALIS) and (TIMSS) are cited throughout the study, none of the studies specifically control for OECD status. Fackler and Malmberg (2016) conclude their study specifically identifying the need to include country-level variables to identify whether a relationship between external mezzo-variables and TSE can be identified. Only one study utilized large scale assessment data as an external variable, reporting a significant finding between country-level reading assessment scores and country-level TSE (Fackler et al., 2021).

Student Achievement

As a research construct, TSE was first studied by RAND researchers in a 1976 survey of Los Angeles schools (Armor et al., 1976). The study used Rotter's (1966) work on expectancy to develop two survey questions measuring the extent to which the teacher believed he or she has the capacity to produce an effect on the learning of students and found a positive correlation between the survey responses and student reading gains.

Ashton, Webb, and Doda (1983) used the RAND questions in a "Teacher Efficacy Study," finding a positive relationship between personal teaching efficacy and student performance, but also noting efficacy attitudes are elusive and changing. Ross (1992) conducted a small study of 18 teachers and found a positive correlation between student achievement and personal teaching efficacy.

Klassen et al.'s (2011) meta-analysis of 218 studies conducted between 1998 and 2009 and found that only 2.8% of the research included links between TSE and student academic outcomes, and only two studies specifically examined the relationship between TSE and student academic achievement. Klassen and Tze (2014) found self-efficacy is modestly but significantly associated with the achievement levels of students.

Zee and Koomen (2016) identified 165 articles as critical to the historical study of TSE, of which 23 studies included student academic achievement in relation to TSE. The studies ranged from sample sizes of 20 to over 2,000 and resulted in a criticism that the wide variety of sample sizes and methods limited the value of the findings, leading them to the same conclusion as Klassen and Tze (2014). Their meta-analysis also supported the development of a heuristic model of TSE that relates TSE to student achievement in a variety of contexts (Zee & Koomen, 2016).

Guo et al. (2012) used structural equation modeling to map the relationship between TSE and student literacy in an elementary setting. In one model, they found a positive relationship between teacher self-efficacy and student outcomes. In a second model, they identified a positive indirect relationship between TSE, teacher support for learning, and student literacy outcomes. In a third model, they found a negative relationship between TSE, teaching experience, and student academic outcomes.

Within international large-scale assessments, there has been limited study of the relationship between TSE and student achievement. Vieluf et al. (2013) used country-level PISA 2009 reading achievement scores as a variable along with 2009 TALIS data and found no significant relationships. Fackler and Malmberg (2016) used a teacher-reported student achievement variable collected on the 2007 TALIS, identifying the degree to which the class as a whole achieved at higher or lower than average levels, and reported a strong relationship between student achievement and TSE.

International Contexts

Bandura (1997) argues that regardless of whether a society operates as a collectivist or individualist society, self-efficacy may be generalized cross-culturally. Bandura proposed that self-efficacy should not be misconstrued as individualistic, given that choices an individual makes may be aligned to any type of cultural norms and every culture is dependent on successful adaptation and change regardless of how that is defined or valued (p. 32).

Early education researchers raised concerns regarding the application of TSE across cultural norms (Ho & Hau, 2004). However, with the increase in large-scale international assessments and international research collaborations, TSE has become identified as a universal construct allowing for generalized findings across countries (Fackler & Malmberg, 2016; Fackler et al., 2021; Klassen et al., 2010; Tschannen-Moran & Woolfolk Hoy, 2001; Vieluf et al., 2013; Yang, 2020).

Chong et al. (2010) studied TSE in Singpore, and determined that TSE had relevance when studied in the context of Asian education systems. Vieluf, Kumter, and van de Vijver (2013) utilized the 2008 TALIS assessment to study the validity of the use of the TSE construct across countries and found that the greatest variation occurs between teachers. Further, while

different norms may impact teacher perceptions, TSE has the same meaning as a construct on an international scale. Glassow, Rolfe, and Hansen (2021) utilized a measurement invariance calculation for studying TSE on the TIMSS assessment and found measurement of TSE generalizable across cultures utilizing a single-construct measure of TSE. Fackler, Malmberg, and Sammons (2021) used structural equation modeling to study the 2013 TALIS assessment from 32 countries and found the greatest variance between teachers, followed by the country, with the least variance occurring between schools

It is also important to note that TSE as an international construct has been validated using items from both the Gibson and Dembo (1984) measure (Fackler & Malmberg, 2016; Vieluf et al., 2013) and the TSES measure (Fackler et al., 2021; Yang, 2020).

Critical Perspectives

Wyatt (2014) argued that a variety of tensions between the scales and theories make it difficult to appropriately measure TSE. The first set of tensions occurs in the relationship between the ends and means of TSE and teacher agency—that is, whether TSE is measured based on the teachers' outcomes or the instructional practices used in the teaching process. Another set of tensions occurs between the level of specificity the teacher is relating their measure of self-efficacy to (Wyatt, 2014); is it as granular as the teaching of a specific concept or as broad as the teaching experience as a whole? Shahid and Thompson (2001) found measures of TSE varying widely, lacking common variables, and complicating meta-analysis.

Wheatley (2005) notes that teacher responses on surveys such as the TSES may not clearly reflect the teachers' understanding of their practice as opposed to their personal belief in a particular type of practice, resulting in ambiguous findings. This dynamic is further complicated by cultural challenges when conducted on a global scale. Wheatley (2005)

concludes the review noting that generalizations of higher levels of TSE as beneficial to students is largely unsubstantiated, and questions whether overvaluing TSE could lead to suppression of potentially beneficial teacher doubts.

Wheatley (2002) argues that TSE models overvalue teacher ratings of self-efficacy, and do not adequately allow for the self-doubt that is critical to the reflection and learning process. Furthermore, he notes systemic school improvement models are dependent on teachers recognizing and working through the self-doubt and cognitive dissonance that are essential to strengthening practice. Wheatley's (2005) critical perspective expands to argue that prior research on TSE is predicated on teacher confidence, and thereby undermines the importance of doubt in the reflection and learning process. To support this claim, the author primarily depends on research about student learning.

Bandura (1997) notes that teachers with lower self-efficacy are "beset by self-doubt" and therefore construct classroom experiences that are likely to be "custodial" and result in student experiences that are lower in cognitive challenges (p. 241). Theoretically, this argument frames many of the research hypotheses in the literature around teacher self-efficacy and student achievement outcomes.

Morris et al. (2017) point out that the construct of TSE has not been measured consistently throughout the literature and argue the research does not adequately address the nuances and ambiguous nature of how a mastery experience is defined, leading to a body of study that includes a wide range of topics inclusive of teacher career cycles and student engagement (Morris et al., 2017).

Wyatt (2014) notes that a significant amount of the prior studies overly generalize the difference between individual perceptions of what a teacher could do on a general basis versus

what teachers believe about their personal instructional practice. Ross et al. (1996) argue that the weaknesses of prior studies include a lack of study between teachers in terms of variance in teacher efficacy. Wyatt (2014) and Wheatley (2002) both point to research as being overly normative, ultimately equating high TSE with successful practice.

Klassen et al. (2011) calls for more research investigating the relationship between TSE and academic achievement with the need for specificity and clarity relating to concept measurement. Fackler et al. (2021) notes that the use of test results, as opposed to self-reported student achievement perspectives, may provide a more robust variable.

Burić and Kim (2020) argue prior research on TSE and student motivation had significant methodological shortcomings and suggest the future use of structural equation measurement (SEM) as a more in-depth study of the data (Zee & Koomen, 2016). A variety of later studies utilized SEM to analyze findings related to the availability of international large-scale assessment data.

Summary

Over twenty-five years and hundreds of thousands of research studies around TSE illustrate how powerful this construct appears across educational systems. The introduction of international large-scale assessments such as the TALIS, TIMSS, and PISA provides the opportunity to investigate new relationships between teacher self-efficacy and factors not previously studied at an international level.

Developing and supporting stronger, more quantitatively effective teachers has become a subject of intense policy study impacting all levels of teacher education and development as demonstrated by the commitment of the OECD member and partner countries who participate in the PISA. A variety of research has established a critical relationship between teachers and

student achievement, noting it is stronger than almost every other factor on student achievement (Chetty et al., 2014; Hattie, 2003; Kane & Staiger, 2008).

Significant research regarding individual teacher factors such as gender, age, education, teaching experience, professional development, and pedagogical philosophies provide interesting perspectives on the person-level characteristics that impact TSE. At the school level, research has included the type of school environment, the school community size, student social economic status, principal leadership style, and academic achievement.

More recent studies have begun examining the relationship between the school leader and TSE, resulting in calls for more research about principals' roles in TSE (Fackler & Malmberg, 2016; Guo et al., 2012). However, there is a scarcity of research examining factors external to the school environment, such as use of assessments, accountability initiatives, or country-level economic indicators as related to TSE.

As we look to researchers to provide insight into how to develop stronger systems of teacher preparation and support teachers' growth in pedagogical skills, the behaviors and beliefs of teachers that make up teacher self-efficacy merit significant consideration. Breakspear (2014) encourages OECD to study the multidimensionality of school systems, and the study of TSE in relationship to other impactful variables measured by PISA provides a unique opportunity to learn how different factors impact teacher self-efficacy. Furthermore, studying the relationship between teacher self-efficacy and school-level and country-level variables may shed further light onto how evaluation tools and teacher quality frameworks can be more effectively used to support professional learning and development of teacher self-efficacy.

CHAPTER III: METHODS

The purpose of this study is to examine the influence of three levels of factors, individual, school, and government, on teacher self-efficacy (TSE). The data used for this analysis has been gathered through the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) assessment conducted in 2018. This chapter begins with an overview of the research questions and theoretical constructs that drive the analysis. This is followed by an overview of the PISA assessment, including the teacher questionnaire and school questionnaire, along with the processes used to prepare the data set for statistical analysis. Next, I present the research plan for the analysis, along with the analytical methods utilized. The chapter concludes with a discussion of limitations of the research.

Problem Statement

Previous research on teacher self-efficacy (TSE) has used established scales such as Gibson and Dembo's (1984) Teacher Efficacy Scale or Tschannen-Moran and Woolfolk Hoy's (2001) Teacher Sense of Efficacy Scale (TSES). While the concept of TSE has been thoroughly researched in relation to teacher career development, teacher professional development, and student self-efficacy, few studies have provided the global scope afforded by the PISA data or the variety of environmental data afforded therein. Often, self-efficacy surveys have been limited in their sample size, and therefore data sets have diminished opportunities to compare teacher responses to gain perspectives necessary to examine the many variables that relate to TSE (Klassen et al., 2009; Morris et al., 2017; Zee & Koomen 2016).

Purpose

The PISA teacher questionnaire provides an opportunity to study a significant international sample of teachers and multi-level factors gathered across countries and economies.

The purpose of this study is to examine the relationship between TSE, as measured by the PISA teacher questionnaire responses, and multiple levels of factors identified from the 19 countries who administered the teacher questionnaire. Understanding how different aspects of the school, region, and governmental environments influence TSE provides additional research to support school-level practitioners and state and federal policymakers as they develop, implement, and evaluate programs aimed at teacher growth and development and ultimately student learning and achievement. Breakspear (2014) challenged researchers to utilize the extensive OECD data to study the multidimensionality of school systems and this study seeks to apply that challenge to the study of many facets of TSE.

Research Design

This study is rooted in a post-positivist approach. As such, the concept of TSE is regarded as a phenomenon subject to the individual experiences, while also recognizing that individuals' experiences lend themselves toward quantification and theoretical examination (Phillips & Barbules, 2000). In the context of a teacher, this includes personal experiences framed by age, gender, and years of experience, but it also includes school and political environments that may influence a teacher's concept of self-efficacy. From this perspective, there is acknowledgement that the responses of teachers in the survey are subject to different contexts, and the study of the relationships among these contexts will provide an opportunity to further explore the relationships between and among teachers, schools, economies, and students (Fox, 2008).

Research Questions

The study will utilize data gathered in the 2018 PISA General Teacher Questionnaire and the 2018 School Questionnaire to investigate the following questions:

RQ1:   What is the relationship between individual characteristics and teacher self-

efficacy (TSE)?

- How do individual characteristics such as age, gender, work experience, and

  completion of a teacher training program relate to TSE?

- How do individual professional development experiences relate to TSE?

RQ2: How do individual characteristics and school-level factors influence teacher self-efficacy

(TSE)?

- How are the teacher-level responses influenced by characteristics of the school such

  as size and school type (public/private)?

- How are the teacher-level responses influenced by school-level environmental

  experiences such as the principal's perception of school's instructional capacity, and

  the quality assurance (accountability) approaches?

- How are the teacher-level responses influenced by school level student achievement?

RQ3: How do individual characteristics, school-level factors, and country-level factors influence

teacher self-efficacy (TSE)?

- How is TSE influenced by a country's economic development status as defined by

  OECD status?

- Is there a relationship between the country's level of student achievement and TSE?

Null Hypotheses

H_1: Individual teacher characteristics do not influence TSE.

H_2: Individual experiential factors do not influence TSE.

H_3: School-level factors do not influence TSE.

H_4: School-level environmental factors do not influence TSE.

H_5: Country-level factors do not influence TSE.

Data and Sample

The OECD is an international organization developed to encourage global economic growth through the study and analysis of policy. As countries became more interested in global economic competition, the OECD utilized its member nations and partner countries/economies to participate in a study of education systems. Since its inception in 1997, OECD has offered the triennial assessment measuring student performance in mathematics, reading, and science (OECD, 2019a). In 2018, OECD administered the PISA to approximately 710,000 15-year-olds from 79 countries/economies across the world. The sample of countries include the 37 countries who are members of the OECD, as well as less economically developed partner countries and economies (OECD, 2019a). The sample is further delineated by regions and sub-stratum that allow for comparison across different levels of the individual countries.

The 2018 PISA examination identified reading as the major domain tested, and this testing cycle also included the first computer-based administration using multistage adaptive testing with a significant number of items tested using a field trial process (OECD, 2018). The items tested for students were developed based on frameworks for cognitive processing and included subject matter experts from around the world to develop items appropriate for measuring student knowledge and understanding (OECD, 2018). In addition, test item reliability was tested both within countries and across countries, and coder reliability studies found that the within-country score agreement met or exceeded the set standards (OECD, 2018). In both the School Questionnaire and Teacher Questionnaire, questions were developed based on guidance provided by Questionnaire Expert Groups and tested on respondents through field trials (OECD, 2018). Teachers were sampled across schools, with a maximum of 10 reading teachers and 10

teachers of other subjects in each school (OECD, 2018). Teachers sampled were either current or former teachers of the approximate grade level. In small schools the sample may have included the total population of teachers, while in larger schools it may have comprised a smaller percentage of the teaching staff (OECD, 2018).

In 2018, PISA provided questionnaires to students, school level leaders/principals, and teachers, resulting in a combined 1,641 items. Across all 79 countries, 21,903 schools participated in the 2018 PISA School Questionnaire, a survey tool with 192 items representing school-level perspectives and policies. The school questionnaire was completed by the school leader or principal for each participating school.

Table 3.1: Overall Initial Sample Distribution (OECD, 2019c)

Type	Totals
Countries	19
OECD	7
Non-OECD	12
Schools	6,128
Public	3,939
Private	1,741
Missing	448
OECD	2,665
Non-OECD	3,463
Teachers	107,367

The 2018 PISA Teacher Questionnaire includes 311 questions about teaching conditions, teacher education and professional development, technology, content and pedagogical knowledge, and teacher self-efficacy (OECD, 2017). Of the 79 total participating countries/economies who administered the PISA in 2018, 19 countries with 107,367 teachers participated in the teacher questionnaires as noted in Table 3.1.

Table 3.2 provides an overview of the total teacher questionnaire sample by country, along with the mean student achievement in each of the content areas tested. This data provides a unique opportunity to study a wide variety of individual, school-level, and governmental factors

and how they relate to teacher self-efficacy. It is critical to note OECD identified inconsistences

in student response time on the reading assessment in Spain, and a listwise deletion removed all

of the variables related to Spain from the regression analysis (OECD, 2019a; 2021a).

Table 3.2: Total 2018 PISA Teacher Questionnaire Sample and Mean Achievement Scores (OECD, 2019c)

Country	OECD	No. Participating Teachers	% of Total	Mean Reading	Mean Math	Mean Science
Albania	No	3375	3%	405	437	417
Baku (Azerbaijan)	No	4077	4%	389	420	398
Brazil	No	8969	8%	413	384	404
Chile	Yes	3755	3%	452	417	444
Chinese Taipei	No	4586	4%	503	531	516
Dominican Republic	No	2700	3%	342	325	336
Germany	Yes	6687	6%	498	500	503
Hong Kong (China)	No	3754	3%	524	551	517
Korea	Yes	4068	4%	514	526	519
Macao (China)	No	2823	3%	525	558	544
Malaysia	No	4737	4%	415	440	438
Morocco	No	3451	3%	359	368	377
Panama	No	3632	3%	377	353	365
Peru	No	5146	5%	401	400	404
Portugal	Yes	5452	5%	492	492	492
Spain	Yes	21621	20%	-	481	483
United Arab Emirates	No	12358	12%	432	435	434
United Kingdom	Yes	2650	2%	504	502	505
United States	Yes	3526	3%	505	478	502

As Table 3.3 outlines, these variables will be compared across the macro (OECD status,

country), meso (school), and micro (teacher) levels. The dataset utilized will be a cross-section

of three PISA data sets, the General Teacher Questionnaire (TQ), School Questionnaire (SQ) and

the Student Questionnaire (STQ). Each school participated in a SQ that was completed by the

school leader (principal, dean, etc). Each school is identified via the international school ID

(CNTSCHID). The sample of teachers chosen to participate in each school are also identified

utilizing the international school ID. The two datasets were combined to allow for each school's

values to be tied to each teacher in that school. While PISA does not allow for a direct

relationship to be identified between the teachers that responded to the questionnaire and the

students that were tested, the students' mean achievement scores are able to be compared at the

school level.

Table 3.3: Variables and Levels

Level	Variable	Location
Micro/Individual Characteristics	Female	TQ
	Age	TQ
	How many years of work experience do you have?	TQ
Micro/Individual Environment	Did you complete a teacher education or training programme?	TQ
	Are you required to participate in professional development activities?	TQ
	Participation in development experience over the last 12 months scale	TQ
	Composite current need for professional development scale	TQ
	How much time do you spending reading for your work out of your classes?	TQ
	TSE All Items (DV)	TQ
	TSE 1: Instruction and Engagement (DV)	TQ
	TSE 2: Student Behavior and Classroom Management (DV)	TQ
Meso/School Characteristics	Public School	SQ
	School Size	SQ
Meso/School Environment	Principal Perception of Capacity-Instructional Staff	SQ
	Principal Perception of Capacity-Staff Behaviors	SQ
	Quality Assurance: External Evaluation	SQ
	Quality Assurance: Professional Control	SQ
	Quality Assurance: Management Approach	SQ
	School achievement: Mean Reading score	STQ
Macro/ Country/Economy	OECD Status	All
	Country achievement: Mean Reading score	PISA

Measures

Tschannen-Moran and Woolfolk Hoy (2001) created the Teacher Sense of Efficacy

Survey (TSES), originally titled the Ohio State Teacher Efficacy Scale (OSTES), after an

extensive review of prior instruments. They found the other measures of TSE struggled with

specificity and generality, noting, "there are conceptual problems in the interpretation of the

factor structure and the poor correlation between the factors where two or more have been

found" (p. 792). The TSES was constructed after multiple iterations using two instruments, one

with 24 questions and the other with 12 questions.

Those 12 items identified in the TSES were included in the PISA "General Teacher"

questionnaire. These were the only items in the questionnaire asking teachers to reflect on

specific details of their own instructional practice. Table 3.4 provides the text of each of the

items, along with their correspondence to the PISA, the TSES, and its subscales. The subscale

findings corresponded to Bandura's (1997) proposal that alignment to constructs specific to the

context provided a stronger validation. The subscales, also referred to as constructs, were

identified as classroom management, student engagement, and instructional strategies

(Tschannen-Moran & Woolfolk Hoy, 2001). While the language of the questions is the same, the

scale employed by PISA differs from the scale used in TSES. TSES is scored on 1-9 scale, with

1 representing "Not at all" to 9 representing "A Great Deal." The PISA questions were answered

on a scale of 1 to 4, with 1 representing "Not at all" to 4 indicating "A lot" (OECD, 2017). The

impact of this shift will be further reviewed in Chapter 4.

Table 3.4: TSE Items (OECD, 2017).

Question Code	Question: In your teaching, to what extent can you:	PISA Code	Subscale/Construct
Q1	Get students to believe they can do well in school work	TC199Q01HA	Instructional Strategies
Q2	Help my students value learning	TC199Q02HA	Instructional Strategies
Q3	Craft good questions for my students	TC199Q03HA	Student Engagement
Q4	Control disruptive behaviour in the classroom	TC199Q04HA	Classroom Management
Q5	Motivate students who show low interest in school work	TC199Q05HA	Instructional Strategies
Q6	Make my expectations about student behaviour clear	TC199Q06HA	Classroom Management
Q7	Help students think critically	TC199Q07HA	Instructional Strategies
Q8	Get students to follow classroom rules	TC199Q08HA	Classroom Management
Q9	Calm a student who is disruptive or noisy	TC199Q09HA	Classroom Management
Q10	Use a variety of assessment strategies	TC199Q10HA	Student Engagement
Q11	Provide an alternative explanation for example when students are confused	TC199Q11HA	Student Engagement
Q12	Implement alternative instructional strategies in my classroom	TC199Q12HA	Student Engagement

Study Design

The purpose of this study is to identify if, and how, different factors aligned to different groupings (micro, meso, macro) predict teacher self-efficacy. For this study, a blocked hierarchical linear regression analysis was chosen to analyze the large data set gathered through the 2018 PISA administration with the goal of identifying relationships between independent variables and the dependent variables, the measures of teacher self-efficacy. This method was chosen to provide the theoretical structure to the analysis and identify the relationship between specific types of variables and TSE (Petrocelli, 2003; Wampold & Freund, 1987). In addition, the method serves as a framework to support sequential modeling and study of the relationship between specific measures of TSE, and their relationship to individual and environmental factors at multiple levels. SPSS Subscription Edition 2018 (Version 26) was used to conduct the blocked hierarchal linear regression analysis. Figure 1 provides an overview of the study framework.

In order to conduct the study, 83 questionnaire items that related to the theoretical variables for the analysis were identified across the micro and meso levels. The two variables used in the macro level were identified through OECD published data (OECD, 2019a). In order to conduct the block regression as described in Figure 3.1, factor analysis was utilized to develop the latent variables that are utilized in the study. The findings related to factor analyses are discussed in Chapter 4.

Limitations

PISA notes that the data is subject to nonsampling and sampling errors (OECD, 2021). The nonsampling errors are related to data collection, nonresponse bias, data processing, and reporting (OECD, 2021). Sampling errors may occur when the sample does not statistically represent the population, thereby creating a degree of uncertainty, or sampling variance (OECD,

2021). PISA provides standard errors for their data through their website, and standard errors for the data used in this study are provided in the descriptive statistics for each variable in Chapter 4.

Another limitation in this data is that all of the data is self-reported. For both teachers and principals completing the survey, there is a risk that the data reflects how they wish to be perceived, or how they feel they should respond as opposed to actual perceptions. In the case of the principals' responses, PISA does not methodically compare what is reported as "mandatory" in a policy, and therefore "mandatory" may be a perception rather than an actual policy. Establishing a relationship between TSE and student achievement is also a limitation of the data. While the sampling procedure requires the teacher responding to the questionnaire to either be a current or former teacher of the tested grade level, the only direct connection between the students and teachers would be available in the smallest schools. Due to this limitation, student achievement data will only be considered at the meso and macro levels.

Results of the study are analyzed in Chapter 4. Implications for practice and recommendations for further study are provided in Chapter 5.

Figure 3.1: Proposed Blocked Hierarchical Linear Regression Model

Micro Block 1: Individual Characteristics	Age Gender Experience Teacher Education

Micro Block 2: Individual Environment	Age Gender Experience Teacher Education PD Required PD Participation Current PD Need Time Reading

Meso Block 3: School Characteristics	Age Gender Experience Teacher Education PD Required PD Participation Current PD Need Time Reading Public School School Size

Meso Block 4: School Environment	Age Gender Experience Teacher Education PD Required PD Participation Current PD Need Time Reading Public School School Size Principal Perception Quality Assurance School Reading Achievement

Macro Block 5: Country Characteristics	Age Gender Experience Teacher Education PD Required PD Participation Current PD Need Time Reading Public School School Size Principal Perception Quality Assurance School Reading Achievement OECD Country Reading Achievement

PISA
TSES Survey
Questions

CHAPTER IV: ANALYSIS

The research and discussion surrounding the relationship between teachers and student learning is broad and deep. Significant bodies of research have established the relationship between teacher efficacy and student achievement, moving researchers to identify what factors support and develop highly effective teachers. Since Bandura (1977) introduced the concept of self-efficacy, it has increasingly been applied to research relating to teachers' development of self-efficacy and its relationship to student learning and achievement.

Previous research on teacher self-efficacy (TSE) has used established scales such as Gibson and Dembo's (1984) Teacher Efficacy Scale or Tschannen-Moran and Woolfolk Hoy's (2001) Teacher Sense of Efficacy Survey (TSES). While the concept of teacher self-efficacy has been thoroughly researched in relation to teacher career development, teacher professional development, and student self-efficacy, few studies have provided the global scope afforded by the PISA data or the variety of individual, experiential, and environmental data afforded through the PISA data. Often, self-efficacy surveys have been limited in their sample size, and therefore data sets are limited in their opportunities to compare teacher responses to gain perspectives necessary to examine the many variables that relate to teacher self-efficacy (Morris et al. 2017; Zee & Koomen 2016; Klassen et al., 2009).

The purpose of this study was to investigate the relationship between individual and environmental factors in teachers' self-efficacy utilizing the data gathered through the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) assessment conducted in 2018. An initial sample size of 107,367 participating teachers in 6,128 schools across 19 countries provides a unique opportunity for a global perspective (OECD, 2019c).

Following the design presented in Chapter 3, a blocked hierarchical linear regression analysis was conducted using data generated from the 2018 PISA administration. This method was chosen to provide the theoretical structure to the analysis and identify the relationship between specific types of variables and TSE (Petrocelli, 2003; Wampold & Freund, 1987).

This chapter is divided into three sections. The first section presents a review of the research questions that form the blocks in the regression. The second section provides a review of the process used to identify the appropriate items and development of subscales for use in the regression, and the latent variables used in the analysis. The third section presents the results of the analysis.

Research Questions

RQ1:   What is the relationship between individual characteristics and teacher self-efficacy (TSE)?

- How do individual characteristics such as age, gender, work experience, and completion of a teacher training program relate to TSE?

- How do individual professional development experiences relate to TSE?

RQ2: How do individual characteristics and school-level factors influence teacher self-efficacy (TSE)?

- How are the teacher-level responses influenced by characteristics of the school such as size and school type (public/private)?

- How are the teacher-level responses influenced by school-level environmental experiences such as the principal's perception of school's instructional capacity, and the quality assurance (accountability) approaches?

- How are the teacher-level responses influenced by school level student achievement?

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)?

- How is TSE influenced by a country's economic development status as defined by OECD status?

- Is there a relationship between the country's level of student achievement and TSE?

Null Hypotheses

H_1: Individual teacher characteristics do not influence TSE.

H_2: Individual experiential factors do not influence TSE.

H_3: School-level factors do not influence TSE.

H_4: School-level environmental factors do not influence TSE.

H_5: Country-level factors do not influence TSE.

Variable Identification and Development

All of the data were downloaded from OECD website (OECD, 2019c). Data was gathered from the computer-based questionnaires identified as Student Questionnaire for PISA 2018 Main Survey Version (STU), School Questionnaire for PISA 2018 Main Survey Version (SQ), and Teacher Questionnaire for PISA 2018 General Teacher (TQ). The SPSS data files for each of the questionnaires were downloaded from the same site. It is important to note that all of the data files required initial recoding to indicate missing values, and specific changes to any coding related to latent variables are described in the sections that follow.

Dependent Variables

Tschannen-Moran and Woolfolk Hoy (2001) created the Teacher Sense of Efficacy Survey (TSES) after an extensive review of prior instruments. They found the other measures of TSE struggled with specificity and generality, noting, "there are conceptual problems in the

interpretation of the factor structure and the poor correlation between the factors where two or more have been found" (p. 792). The TSES was constructed after multiple iterations using two instruments, one with 24 questions and the other with 12 questions. Construct validity was established based on correlations between other established measures of TSE, including the RAND Gibson and Tembo TES survey items (Tschannen Moran & Woolfolk-Hoy, 2001). Table 4.1 provides the text of each of the items, along with their correspondence to the PISA, the TSES and its subscales, and the factor analysis that follows.

Table 4.1: Teacher Sense of Efficacy Survey (TSES) Items as Utilized in PISA 2018 (OECD, 2017).

Question Code	Question: In your teaching, to what extent can you do:	PISA Code	Factor	Subscale/Construct
Q1	Get students to believe they can do well in school work	TC199Q01HA	1	Instructional Strategies
Q2	Help my students value learning	TC199Q02HA	1	Instructional Strategies
Q3	Craft good questions for my students	TC199Q03HA	1	Student Engagement
Q4	Control disruptive behaviour in the classroom	TC199Q04HA	2	Classroom Management
Q5	Motivate students who show low interest in school work	TC199Q05HA	2	Instructional Strategies
Q6	Make my expectations about student behaviour clear	TC199Q06HA	2	Classroom Management
Q7	Help students think critically	TC199Q07HA	1	Instructional Strategies
Q8	Get students to follow classroom rules	TC199Q08HA	2	Classroom Management
Q9	Calm a student who is disruptive or noisy	TC199Q09HA	2	Classroom Management
Q10	Use a variety of assessment strategies	TC199Q10HA	1	Student Engagement
Q11	Provide an alternative explanation for example when students are confused	TC199Q11HA	1	Student Engagement
Q12	Implement alternative instructional strategies in my classroom	TC199Q12HA	1	Student Engagement

Tschannen Moran and Woolfolk-Hoy (2001) utilized principal-axis factor analysis to identify three factors in both sets of questions, establishing the validity of a total scale score as well as the use of subscale scores. The subscale findings corresponded to Bandura's (1997) proposal that alignment to constructs specific to the context provided a stronger validation. The subscales, also referred to as constructs, were identified as classroom management, student engagement, and instructional strategies (Tschannen-Moran & Woolfolk Hoy, 2001).

While the language of the items is the same, the scale employed by PISA differs from the scale used in TSES. TSES was validated using a 1 to 9 scale, with 1 representing "Not at all" to 9 representing "A Great Deal." The PISA items were answered on a scale of 1 to 4, with 1 representing "Not at all" to 4= indicating "A lot" (OECD, 2017).

Initially, the intention was to group responses to the questions based on the constructs as developed by Tschannen-Moran and Woolfolk Hoy (2001). However, the shift in scoring protocols suggested that the use of the constructs needed to be revisited in order to be appropriately utilized as dependent variables. To support the analysis, the 12 items were subjected to a principal axis factor analysis using SPSS Statistics Subscription Cloud Edition (2018).

Using the recoded variables, the factorability of the 12 TSES items was reviewed. First, all 12 of the items correlated at least .38 with at least one other item, suggesting reasonable factorability. Second, the Kaiser-Meyer-Olkin (Kaiser, 1970; 1974) measure of sampling adequacy was .94, above the recommended value of .6, and Bartlett's Test of Sphericity (Bartlett, 1954) was significant (χ^2 (66)=38,1238.83, p<.000). Finally, the communalities were all above .3, providing confirmation that common variance existed with other items. All items in

this analysis had primary loadings over .4, and two items had cross-loadings above .3 with primary loadings above .6.

A principal-axis factor analysis with varimax rotation of the 12-items identified two factors with eigenvalues greater than one, accounting for 61% of the variance in the respondents' scores. A scree test indicated two factors could be extracted.

The factor labels were based on the item content and recognizing that the primary shift from the TSES involves the combination of the Instructional Strategies and Student Engagement subscales, while the Classroom Management subscale added only one new item. As noted in Table 4.2, internal consistency for each of the scales was examined using Cronbach's alpha. The alphas indicated good internal consistency: .89 for Instruction and Engagement (8 items), .84 for Student Behavior and Classroom Management (4 items), and .93 for the Full Survey (12 items). The full factor loading tables are provided in Appendix A.

Table 4.2: Means for Total Score and Subscales

	α
TSE 1: Instruction and Engagement Subscale	.89
TSE 2: Student Behavior and Classroom Management Subscale	.84
TSE 3: Full Survey	.92

Composite scores were created for the two factors as well as for the full survey response and all are used as dependent variables in this analysis. Higher scores on all of the scales indicate teachers hold a higher sense of self-efficacy related to their ability to utilize skillsets related to successful learning environments. Table 4.3 provides the descriptive statistics for each of the three variables.

Table 4.3: Dependent Variables Descriptive Statistics

	N	Min.	Max.	Mean	SD	Variance	Kurtosis	Std. Err.
Instruction and Engagement Subscale	48,019	8	32	26.52	4.457	19.866	-.661	.022
Student Behavior and Classroom Management Subscale	48,560	4	16	13.42	2.319	5.379	-.355	.022
Full Survey	47,654	12	48	39.95	6.339	40.181	-.542	.022

As a result of this finding, the model for the study was revised to reflect the addition of the two subscale measures as demonstrated in Figure 4.1.

This study utilized a blocked hierarchal linear regression analysis to identify the relationships between the independent variables (IVs) and the measure of Teacher Self-Efficacy that form the three dependent variables (DVs). The analysis of the models includes an evaluation of the changes of the unstandardized beta (*B*) as new variables enter the model, supporting analysis of the influence of each variable on the dependent variables and other independent variables. Tables 4.12 through 4.14 provide the regression data for each of the three models. Comparisons of the changes in the standardized betas across the models can be found in Appendix B. SPSS Subscription Edition 2018 (Version 26) was used to conduct the blocked hierarchal linear regression analysis.

Figure 4.1: Revised Blocked Hierarchical Linear Regression Model

Micro Block 1:
Individual
Characteristics

Age
Gender
Experience
Teacher Education

Micro Block 2:
Individual
Environment

Age
Gender
Experience
Teacher Education
PD Required
PD Participation
Current PD Need
Time Reading

Meso Block 3:
School
Characteristics

Age
Gender
Experience
Teacher Education
PD Required
PD Participation
Current PD Need
Time Reading
Public School
School Size

Meso Block 4:
School
Environment

Age
Gender
Experience
Teacher Education
PD Required
PD Participation
Current PD Need
Time Reading
Public School
School Size
Principal Perception$_{1,2}$
Quality Assurance$_{1,2}$
School Reading Achievement

Macro Block 5:
Country
Characteristics

Age
Gender
Experience
Teacher Education
PD Required
PD Participation
Current PD Need
Time Reading
Public School
School Size
Principal Perception$_{1,2}$
Quality Assurance$_{1,2,3}$
School Reading Achievement
OECD
Country Reading Achievement

Student Behavior and Classroom Management

Instruction and Engagement

Full Survey

Micro/Individual Variables

Thirty-seven items from the Teacher Questionnaire were considered for the development of the micro variables. From those items, a total of eight variables are categorized at the teacher-level of analysis. Two of those variables were calculated by combining item responses to create a scale variable.

Gender was recoded into a dummy variable titled *Female*, with females coded as 1 and males as 0. *Age* and *Work Experience* are scale variables based on individual teacher data input. *Teacher Education* was originally developed as three options to indicate whether the teacher attended a program, and if so, whether that program was more or less than one year. For this analysis, it was recoded into a dummy variable indicating a yes or no variable (Yes=1, No=0), with yes indicating the individual attended a teacher education program.

Four variables measure teacher experiential factors related to professional development. The requirement to participate in professional development, coded as *Required PD*, maintains the original values of yes or no (Yes=1, No=0). The second, *Time Reading PD*, asks teachers, "How much time per week do you spend reading for your work (e.g. articles, magazines, books, manuals and websites) out of your classes?" (Teacher Questionnaire, p. 18). The responses are categorized into four groups, less than one hour a week, 1–3 hours a week, 4–6 hours a week, and more than 6 hours a week. The variable provides a scale metric to measure its relationship to the dependent variables. The variable *PD Participation Total* is a sum of the responses to two sets of questions with a total of 11 questions asking teachers whether they participated in a variety of professional development activities or activities related to professional growth. The final professional development variable, *PD Need*, asks teachers to indicate their level of need on a 1 to 4 scale with 1 being the least and 4 being the highest level of need. In order to utilize this

variable, the scale was recoded to indicate 1 representing the greatest need and 4 representing the least amount of need to correspond with individual perceptions of skill and the responses were combined to create a scale variable. Table 4.4 provides the descriptive statistics for the teacher level variables.

Table 4.4: Micro/Individual Teacher Level Independent Variables Descriptive Statistics

	N	Min.	Max.	Mean	SD	Variance	Kurtosis	Std. Error
Female	71,354	0	1	0.63	0.483	0.233	-1.707	0.018
Age	71,539	20	70	42.65	10.215	104.341	-0.793	0.018
Work Experience	70,440	0	50	16.30	9.861	97.241	-0.551	0.018
Required PD	71,402	0	1	0.76	0.425	0.181	-0.471	0.018
Teacher				0.89	0.313	0.098	4.178	0.018
Education	71,493	0	1					
Time Reading PD	71,353	1	4	2.83	0.955	0.911	-1.123	0.018
PD Participation				5.37	2.464	6.072	-0.441	0.023
Total	45,981	0	11					
PD Need	44,357	18	72	42.14	12.970	168.226	-0.641	0.023

Meso/School Level Variables

Thirty-three items from the Principal Questionnaire were considered for the development of the meso variables. From those items, seven variables were identified or developed for study at the school level of analysis. Four of the seven items are latent variables developed through factor analysis. The final variable relating to student achievement was calculated through the use of the Student Questionnaire data. Table 4.5 provides the descriptive statistics for each of the variables and the factor analyses are discussed below.

The first two variables provide information relating to the school characteristics. The *Public School* variable is a dummy variable based on the identification of schools as public or private, with public schools coded as 1, private schools as 0. *School Size* is determined through direct input from the school leader on the survey and ranged from 1 to 10,700. Due to this wide variation and the uneven distribution of school sizes, the responses were divided into seven groups categorized by school size and recoded based on the size classification. Schools were

coded as follows: 1–300 students; Very Small (1), 301–500 students; Small (2), 501–800

students; Mid-Small (3), 801–1100 students; Mid-Size (4), 1,101–1,500 students; Mid- Large

(5), 1,501–2,000 students; Large (6), and above 2,001 students; Very Large (7).

Table 4.5: Meso/School Level Variables Descriptive Statistics

	N	Min.	Max.	Mean	SD	Variance	Kurtosis	Std. Error
Public School	68,581	0	1	0.69	0.460	0.212	-1.283	0.019
School Size	62,113	1	7	3.93	1.818	3.304	-0.933	0.020
Principal Perception of Capacity: Instructional Staff	66,360	4	16	12.46	2.820	7.951	-0.396	0.019
Principal Perception of Capacity: Staff Behavior	67,377	5	20	14.77	3.239	10.494	0.279	0.019
Quality Assurance: Professional Control	67,443	0	8	3.98	1.957	3.831	-0.388	0.019
Quality Assurance: Management Approach	66,938	0	10	7.46	2.056	4.228	-0.609	0.019
School Achievement: Mean Reading Score	68,578	201.76	676.46	447.611	79.077	6253.105	-0.730	0.019

Principal Perception of Capacity

Research on leadership (Bendikson et al., 2012; Hallinger & Heck, 1996; Marks &

Printy, 2003; Sehgal et al., 2017) proposes school leader perspectives about school capacity

impact TSE. In the School Questionnaire, the school leader is asked two questions relating to

their perception of instructional capacity in their schools. First, they are asked "Is your school's

capacity to provide instruction hindered by any of the following issues?" (OECD, 2017, p. 6)

with eight items identifying different aspects of the school environment on the school's

instruction including teaching staff, educational material, and infrastructure. Second, they are

asked, "In your school, to what extent is the learning of students hindered by the following

phenomena?" (OECD, 2017, p. 21) with 11 items identifying student and teacher behaviors that

impact student learning. For the construction of the variables related to the principal or school

leader's perception of capacity, both question sets were subjected to a principal axis factor

analysis using SPSS Subscription Edition 2018 (Version 26).

For the first question, "Is your school's capacity to provide instruction hindered by any of

the following issues?" (OECD, 2017, p. 6) a correlation analysis indicated the presence of all

coefficients at .3 or above. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .79,

exceeding the recommended value of .6 (Kaiser, 1970; 1974) and Bartlett's Test of Sphericity

(Bartlett, 1954) was significant (χ^2 (28)=334,055.37, p<.000) supporting the factorability of the

correlation matrix. All eight extracted communalities were above .3, providing confirmation that

common variance existed with other items. All of the items were included between the two

factors because all displayed a primary factor loading greater than .3 and with four items

demonstrating cross factor loading greater than .3.

The principal-axis factor analysis with varimax rotation of the eight items identified two

factors with eigenvalues greater than one, accounting for 65.73% of the variance in the

respondents' scores. A scree test indicated two factors could be extracted. All of the items in this

analysis had primary loadings over .5, and four items had cross-loading above .3 with primary

loadings above .5. The full factor loading tables are provided in Appendix A.

The factor labels were based on the item content and reflect the two very different groups

of items included. For the purposes of this analysis, only the *Instructional Staff* factor will be

used for the study, as it represents a latent variable measuring the school leader's perspective of

the instructional staff capacity. For the purposes of scale development, responses to the questions

were recoded to correspond to the following values, 1= "A lot", 2= "To some extent," 3= "Very

Little," 4= "Not at all," (PISA, 2017, p. 6). The item responses were combined for each of the

variables with higher scores a more positive principal perception of the school's instructional capacity.

For the second question, "In your school, to what extent is the learning of students hindered by the following phenomena?" (OECD, 2017, p. 21), a correlation analysis indicated the presence of all coefficients at .3 or above. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .91, exceeding the recommended value of .6 (Kaiser, 1970; 1974) and Bartlett's Test of Sphericity (Bartlett, 1954) was significant (χ^2 (55)=484,079.04, p<.000) supporting the factorability of the correlation matrix. All eleven extracted communalities were above .3, providing confirmation that common variance existed with other items. All of the items were included between the two factors because all displayed a primary factor loading greater than .5 and four items demonstrated cross factor loading greater than .3.

The principal-axis factor analysis with varimax rotation of the eight items identified two factors with eigenvalues greater than one, accounting for 62.54% of the variance in the respondents' scores. A scree test indicated two factors could be extracted. The full factor loading tables are provided in Appendix A.

The factor labels were based on the item content and reflect the two very different groups of items included. For the purposes of this analysis, only the *Staff Behaviors* factor will be used for the study, as it represents the school leader's perspective of the instructional staff efficacy. For the purposes of scale development, responses to the questions were recoded to correspond to the following values, 1= "A lot", 2= "To some extent," 3= "Very Little," 4= "Not at all," (OECD, 2017, p. 6). The item responses were combined for each of the variables, with higher scores a more positive principal perception of the school's instructional capacity.

Together, two latent variables were created based on factor analysis. As noted in Table 4.6, internal consistency for each of the scales was examined using Cronbach's alpha. For the *Instructional Staff* subscale, the alpha is .77; for the *Staff Behavior* subscale, the alpha is .83.

Table 4.6: Means for Principal Perception of Capacity Items Variables

	α
Principal Perception of Capacity Instructional Staff Subscale	.77
Principal Perception of Capacity Staff Behavior Subscale	.83

Quality Assurance

The PISA School Questionnaire included 10 questions related to school level quality assurance, all responding to the primary question, "Do the following arrangements aimed at quality assurance and improvements exist in your school and where do they come from?" (OECD, 2017, p. 16). In order to better consider these questions for the purpose of developing school level variables related to accountability measures, the 10 items were subjected to a principal axis factor analysis using SPSS Subscription Edition 2018 (Version 26).

Prior to performing the factor analysis, a review of the suitability found the correlation matrix had a range of coefficients (min.=.087, max.=.510). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .82, exceeding the recommended value of .6 (Kaiser, 1970; 1974) and Bartlett's Test of Sphericity (Bartlett, 1954) was significant (χ^2 (45)=154,891.41 p<.000), supporting the factorability of the correlation matrix. Finally, six of the ten extracted communalities were all above .3, providing confirmation that common variance existed with other items. All of the items were included between the two factors because all displayed a primary factor loading greater than .3 and with four items demonstrating cross factor loading greater than .3.

The principal-axis factor analysis with varimax rotation of the 10 items identified two factors with eigenvalues greater than one, accounting for 45.12% of the variance in the respondents' scores. A scree test indicated two factors could be extracted. Nine of the ten items in this analysis had primary loadings over .4, and three items had cross-loadings above .3 with primary loadings above .5. One of the items, "Quality Assurance at School: External Evaluation" had correlations at or below .19 and did not have loadings above .3; it was accordingly removed from the factors. The full factor loading tables are provided in Appendix A.

Accountability initiatives, including the collection and use of data to improve instruction, have been central to the purpose of OECD work related to education—this includes the use of the PISA. While OECD reports the data related to these items individually, the factor analysis reveals the relationships between items correspond to two types of accountability models proposed by Leithwood and Earl (2000): professional control accountability where school level policies and interactions are central, and management accountability where use of data and planning protocols are primary mechanisms. The factor labels reflect the relationship between the items and the accountability models.

Latent variables were created for the two factors and represent independent variables in this analysis; they are Quality Assurance: Professional Control and Quality Assurance: Management Approach. For the purposes of the scale development, responses to the questions were recoded to correspond to the following values: 0= "No", 1= "Yes, based on school initiative", 2= "Yes, this is mandatory, e.g. based on district or ministry policies" (OECD, 2017, p. 16). The item responses were combined for each of the variables, with higher scores indicating greater accountability mechanisms are present in the school environment. As noted in Table 4.7, internal consistency for each of the scales was examined using Cronbach's alpha. The alphas

indicated good internal consistency: .70 for Professional Control (4 items) and .70 for

Management Approach (5 items). As noted above, one question relating to external evaluation

was excluded in the factor analysis and was not included in this study.

Table 4.7: Means for Quality Assurance (QA) At School Variables

	α
QA all items	.76
QA all items excluding External Evaluation	.77
QA Professional Control	.70
QA Management Approach	.70

Student Achievement: Reading

The 2018 PISA was designed with a focus on reading in digital environment and reading

literacy, defining reading literacy as "understanding, using, evaluating, reflecting on and

engaging with texts in or der to achieve one's goals, to develop one's knowledge and potential,

and to participate in society" (OECD, 2019a, p. 4). In order to measure student achievement,

OECD utilized a combination of online platform items and paper-based items. Computer-based

items were constructed using a multistage adaptive testing design that scaled the difficulty level

of the material based on student responses. Paper-based items were administered across countries

with different levels of difficulties assigned by sample size. Student achievement in the content

area sections of the 2018 PISA administration was then calculated based on statistical modeling

using item response theory and latent regression models to create 10 plausible values for each

student drawn from a posterior distribution (OECD, 2021a). The plausible values are presented

specifically to make group-level inferences, rather than individual-level inferences (p. 23).

In order to build a school-level mean of student achievement on the PISA, as identified in

Table 8, a mean score was calculated by collecting the mean of the ten plausible values for each

student identification code (CNTSTUID). From those means, a mean score based on the total

student population for each school identification code (CNTSCHID) was created, resulting in the school level mean in reading used in this analysis. These means are consistent with the range of country mean scores reported by OECD (OECD, 2019a).

It is critical to note that in the construction of this variable, OECD identified inconsistences in student response time on the reading assessment in Spain, and for the regression analysis, all of the variables related to Spain were removed from the calculations. As a result, the N decreases significantly, with shifts evident in the statistics appearing in Table 4.8.

Table 4.8: Meso/School Level Variables Descriptive Statistics

	N	Min.	Max.	Mean	SD	Kurtosis	Std. Error
School Achievement: Mean Reading Score	68,578	201.76	676.46	447.61	79.10	-.730	.019

Macro/Country Level Factors

The outcomes from the PISA demonstrate the close relationship between economics and student achievement, noting, "the quality of the education a student acquires can still best be predicted by the student's or his or her school's socio-economic background" (OECD, 2019a, p. 5). The goal of this study was to identify whether the same applies to a country's socio-economic level and teacher self-efficacy. Two variables utilized to examine this relationship included OECD status, specifically looking at whether controlling for OECD has an impact on TSE, and the mean student achievement in reading on the 2018 PISA. In order to establish the reliability of the country mean score, as discussed above, it is provided in Table 4.9 along with the mean score of the country as established by OECD (2019a). As noted above, Spain is not included in the student achievement analysis. However, its data is included in the variables below in the event that it is used as part of future research on other academic variables (Appendix C).

Membership in OECD is representative of countries with stronger economic and policy development, and for the purposes of this study serves as a proxy for stronger economies. In the data, membership in the OECD is coded as 1, while partner countries/economies are coded as 0. The descriptive statistics related to the study are provided in Table 10.

Table 4.9: Sample Macro/Country Level Variables (OECD, 2019a; 2019c).

Country	OECD	N	% of Total	Country Mean Reading	OECD Mean Reading
Albania	No	2,947	3%	402.85	405
Baku (Azerbaijan)	No	2,154	2%	389.48	389
Brazil	No	6,674	8%	420.32	413
Chile	Yes	3,167	4%	473.19	452
Chinese Taipei	No	4,046	5%	496.86	503
Dominican Republic	No	2,310	3%	344.98	342
Germany	Yes	3,933	5%	501.13	498
Hong Kong (China)	No	3,196	4%	525.61	524
Korea	Yes	3,941	5%	517.94	514
Macao (China)	No	2,655	3%	518.98	525
Malaysia	No	4,645	5%	415.50	415
Morocco	No	3,056	4%	358.06	359
Panama	No	2,701	3%	376.11	377
Peru	No	4,824	6%	403.83	401
Portugal	Yes	4,077	5%	499.23	492
United Arab Emirates	No	10,017	11%	432.63	432
United Kingdom	Yes	1,417	2%	501.30	504
United States	Yes	2,818	3%	505.14	505

Table 4.10: Sample Macro/Country Level Variables Descriptive Statistics

	N	Min.	Max.	Mean	SD	Kurtosis	Std. Error
OECD	91,190	0	1	.44	.50	-1.937	.016
School Achievement: Mean Reading Score	72,297	342	525	445.96	53.45	-1.22	.018

Missing Data

The analysis was completed using a pairwise deletion. As noted in the data presented, this resulted in a reduction of the sample, but the sample sizes remained significant to the study.

Summary of Data

A total of 83 questionnaire items that related to the theoretical variables for the analysis were initially identified to study the micro and meso levels. The macro level data was provided through OECD (OECD, 2019a). Factor analysis, described in the prior sections of this chapter, was used to develop latent IVs and the DVs. Ultimately, the regression analysis utilized a total of 17 independent variables (IV) and 3 dependent variables (DV). Table 4.11 provides the titles for each of the variables used in the analysis and the source of the variable. The remainder of this section is organized by the level of items and provides descriptive statistics relevant to the analysis.

Table 4.11: Variables and Levels (OECD, 2017)

Level	Variable	Location
Micro/Individual	Female	TQ
	Age	TQ
	How many years of work experience do you have?	TQ
	Did you complete a teacher education or training programme?	TQ
	Are you required to participate in professional development	TQ
	activities?	TQ
	Participation in development experience over the last 12	TQ
	months scale*	TQ
	Composite current need for professional development scale*	TQ
	How much time do you spending reading for your work out of	TQ
	classes?	TQ
	TSE All Items (DV)	
	TSE 1: Instruction and Engagement (DV)*	
	TSE 2: Student Behavior and Classroom Management (DV)*	
Meso/School		
	Public School	SQ
	School Size	SQ
	Principal Perception of Capacity-Instructional Staff*	SQ
	Principal Perception of Capacity-Staff Behaviors*	SQ
	Quality Assurance: Professional Control*	SQ
	Quality Assurance: Management Approach*	SQ
	School Achievement: Mean Reading Score*	STQ
Macro/ Country/Economy		
	OECD Status	All

Country Achievement: Mean Reading Score	PISA

*Variables developed based on multiple items in the associated questionnaire

Regression Analysis

Prior to conducting the regression, the relevant assumptions of the model were tested using simple OLS regression for each IV and DV. After the exclusion of the Spain sample and pairwise deletion, the sample size exceeded 40,000, making it adequate for this analysis (Tabachnick & Fidell, 2001). Linearity all of IVs were tested by running standardized residual scatterplots with the DV. In all of the cases, linear relationships were identified, thereby meeting the assumption of linearity. In addition, the scatterplots did not indicate evidence of heteroscedasticity. Due to the large sample size, tests for normality were not necessary to meet the assumption. The Durbin-Watson values for each of the three DVs range between 1.76 and 1.85, meeting the assumption for independent errors. An examination of correlation revealed two moderately correlated variables (OECD and Country-Level Student Achievement in Reading). However, the VIF and Tolerance values were within the accepted ranges, thereby meeting the assumption of multicollinearity. Finally, all of the independent variables in the models are supported by theoretical considerations as presented in Chapter Two (Hair et al., 1998).

A five-step hierarchical multiple regression was conducted for TSE Instruction and Engagement, TSE Classroom Behavior and Management, and the Full Survey. The individual level characteristics were entered in the first model, followed by the individual experiential variables, school level characteristics, school environment variables, and country level characteristics, for a total of five models. The variables were entered in this order to indicate the shift from the micro level through the macro level, so as to provide a perspective about the levels and types of variables that may be related to TSE.

Results

Tables 4.12 through 4.14 provide the correlations, regression coefficients, and t-statistics for each variable as well as the multiple correlation coefficient for each of the models analyzed and discussed below. Each regression contributed to the analysis of the three research questions framing this study.

The evaluation of the study utilizes the predictability gradient hypothesis as proposed by Stankov (2013) in which correlations between .20 and .35 will be treated as 'moderate' and correlations greater than .35 are considered 'high' (Lee & Stankov, 2013). The predictability gradient hypothesis recognizes that using noncognitive variables impact the correlations, but that such correlations are significant to research.

TSE Instruction and Engagement Subscale

RQ 1: How do teacher-level factors influence teacher self-efficacy? In the first model in the hierarchical regression, teacher level characteristics contributed significantly to the regression model, $F_{(4, 38,137)} = 20.79$, $p<.001$) and accounted for .2% of the variation in TSE Instruction and Engagement. In this model, gender, years of work experience, and completion of a teacher education program were significant, $p<.001$. The characteristic of age was not significant, $p<.231$. Female teachers ($B=.284$, s.e.$=.047$, $p<.001$), years of work experience ($B=.018$, s.e.$=.005$, $p<.001$) were positive and significant predictors of TSE Instruction and Engagement. Completion of a teacher education program ($B=-.276$, s.e.$=.073$, $p<.001$) was a negative and significant predictor of TSE Instruction and Engagement.

Introducing the teacher level experiential variables in the second model explained an additional 10.3% of variation in TSE Instruction and Engagement, resulting in 10.5% of the variance, and this change in R^2 was significant, $F_{(8, 38,133)}= 559.52$, $p<.001$). In this model,

gender remained significant (p<.001), age remained insignificant, and years of work experience was significant (p=.010). The new variables that were introduced, required professional development, total participation in professional development, scale of need for professional development, and total time reading for work were all significant (p<.001). A requirement to participate in professional development (*B*=.460, s.e.=.053, p<.001), total participation in professional development activities (*B*=.271, s.e.=.009, p<.001), and time spent reading for work (*B*=.915, s.e.=.023, p<.001) were positive and significant predictors of the TSE Instruction and Engagement Scale. Need for professional development (*B*=-.049, s.e.=.002, p<.001) was a negative and significant predictor of TSE Instruction and Engagement.

RQ2: How do teacher-level and school-level factors influence teacher self-efficacy?

A third model added two variables, public school and school size, as school-level characteristics that explained an additional .9% of the variation in TSE Instruction and Engagement, resulting in 11.4% of the variance. The change in R² was significant, F (10, 38,131)=490.27, p<.001. Both variables added in this model, public school (*B*=.421, s.e.=.048, p<.001) and school size (*B*=.-.194, s.e.=.012, p<.001), were significant.

In the fourth model, the introduction of school-level environmental characteristics to the regression model explained an additional 10.4% of the variation in TSE Instruction and Engagement, resulting in 21.5% of the variance, and this change in R² square was also significant, F (15, 38,126)=695.45, p<.001. Both principal perception of capacity variables, instruction hindered (*B*=.096, s.e.=.008, p<.001) and student learning hindered (*B*=.032, s.e.=.007, p<.001), were significant. Both of the quality assurance variables, professional control (*B*=.238, s.e.=.012, p<.001) and management approach (*B*=.060, s.e.=.012, p<.001), were significant positive and significant predictors of TSE. Finally, the school achievement mean

reading score (B=-.017, s.e.=.000, p<.001) was a negative and significant predictor of TSE Instruction and Engagement. Between the four models, age remained a negative predictor, but shifted from not significant (p=.231) to significant (p=.021). School size remained a negative predictor but was not significant in this model (p=.055).

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)? A fifth model measuring the macro, or country-level characteristics, added two variables, OECD status and country student achievement mean scores in reading. The macro level variables explained an additional 11.3% of the variation in TSE Instruction and Engagement. The change in R^2 was significant, F (17, 38,124)=1,093.29, p<.001. OECD status (B=.166, s.e.=.055, p=.003) was a positive and significant predictor. Country student achievement mean scores in reading (B= -.040, s.e.=.001, p<.001) was a negative and significant predictor.

In the final model, gender, years of work experience, required participation in professional development, time reading for work, participation in professional development, principal perception of capacity-student learning hindered, quality assurance professional control, and quality assurance management approach were all positive and significant predictors. Age, participation in a teacher education program, scale of need for professional development, public school, and school size were all negative and significant predictors. Principal perception of capacity-instruction hindered and country student achievement mean scores in reading were not significant in this model. Comparisons of the changes in the standardized betas across the models can be found in Appendix B.

The fifth model resulted in 32.7% of the variance in TSE Instruction and Engagement. Across the models, the addition of the individual- and school-level environmental factors and the

country-level characteristics make this model the strongest predictor of TSE Instruction and Engagement.

TSE Student Behavior and Classroom Management Subscale

RQ 1: How do teacher-level factors influence teacher self-efficacy? In the first model in the hierarchical regression, teacher level characteristics contributed significantly to the regression model, $F_{(4, 38,137)} = 39.86$, p<.001) and accounted for .4% of the variation in TSE Student Behavior and Classroom Management. In this model, gender ($B=.128$, s.e.=.025, p<.001) and years of work experience ($B=.022$, s.e.=.002, p<.001) were positive and significant predictors of TSE Student Behavior and Classroom Management. Age ($B=-.011$, s.e.=.002, p<.001) was a negative and significant predictor. Finally, completion of a teacher education program ($B=-.016$, s.e.=.038, p=.672) was not significant.

Introducing the teacher-level experiential variables in the second model explained an additional 6.8% of variation in Student Behavior and Classroom Management, resulting in 7.2% of the variance, and this change in R^2 was significant, $F_{(8, 38,133)}= 370.97$, p<.001. In this model, gender ($B=.095$, s.e.=.024, p<.001) and years of experience ($B=.019$, s.e.=.002, p<.001) remained positive and significant predictors. Completion of a teacher program ($B=-.121$, s.e.=.037, p=.001) shifted to serve as a negative and significant predictor. Age ($B=-.012$, s.e.=.002, p<.001) remained a negative and significant predictor. One of the new variables, scale of need for professional development ($B=-.030$, s.e.=.001, p<.001) was a negative and significant predictor. Three of the new variables that were introduced in this model, required professional development ($B=.291$, s.e.=.028, p<.001), total participation in professional development ($B=.093$, s.e.=.005, p<.001), and total time reading for work ($B=.309$, s.e.=.012, p<.001) were positive and significant predictors of TSE Student Behavior and Classroom Management.

RQ2: How do teacher-level and school-level factors influence teacher self-efficacy?

Adding School Level Characteristics to the third block in the regression model explained an additional .2% of the variation in TSE Student Behavior and Classroom Management, resulting in 7.5% of the variance, and this change in R^2 was significant, $F (10, 38,131)=311.05$, $p<.001$. Both variables added in this model, public school ($B=.089$, s.e.$=.025$, $p<.001$) and school size ($B=-.067$, s.e.$=.006$, $p<.001$), were significant. In this model, all of the teacher-level variables remained significant ($p<.001$). Age and completion of a teacher program remained negative and significant predictors.

In the fourth model, introducing school-level environmental characteristics to the regression model explained an additional 5.7% of the variation in TSE Student Behavior and Classroom Management, resulting in 13.1% of the variance, and this change in R^2 square was also significant, $F (15, 38126)=384.39$, $p<.001$. Both of the principal perception of capacity variables, instruction hindered ($B=.039$, s.e.$=.004$, $p<.001$) and student learning hindered ($B=.024$, s.e.$=.004$, $p<.001$), were significant. Both of the quality assurance variables, professional control ($B=.094$, s.e.$=.007$, $p<.001$) and management approach ($B=.030$, s.e.$=.006$, $p<.001$), were significant positive and significant predictors of TSE. Finally, the school achievement mean reading score ($B=-.006$, s.e.$=.000$, $p<.001$) was a negative and significant predictor of TSE Student Behavior and Classroom Management. Between the four models, age remained a significant and negative predictor. Completion of a teacher program ($p=.101$) and school size ($p=.747$) shifted in this model and were not significant predictors.

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)? A fifth model measuring the macro, or country-level characteristics, added two variables, OECD status and country student achievement mean

scores in reading, and explained an additional 4.8% of the variation in TSE Student Behavior and Classroom Management. The change in R² was significant, F (17, 38,124)=490.95, p<.001. Both OECD status (*B*=-.237, s.e.=.032, p<.001) and country student achievement mean scores in reading (*B*=-.012, s.e.=.000, p<.001) were negative and significant predictors of TSE Student Behavior and Classroom Management.

In the final model, gender, years of work experience, required participation in professional development, time reading for work, participation in professional development, principal perception of capacity student learning, quality assurance professional control, and quality assurance management approach were all positive and significant predictors. Age, scale of need for professional development, public school, and school size were all negative and significant predictors. Completion of a teacher program, principal perception of capacity-instructional staff, and country student achievement mean scores in reading were not significant in this model. Comparisons of the changes in the standardized betas across the models can be found in Appendix B.

Together the five models accounted for 17.9% of the variance in TSE Student Behavior and Classroom Management. Across the models, the addition of the individual- and school-level environmental factors and the country-level characteristics make this model the strongest predictor of TSE Student Behavior and Classroom Management.

Full Survey

RQ 1: How do teacher-level factors influence teacher self-efficacy? In the first model in the hierarchical regression, teacher level characteristics contributed significantly to the regression model, F (4, 38,137) = 28.87, p<.001) and accounted for .3% of the variation in Combined TSE. In this model, gender (*B*=.418, s.e.=.067, p<.001) and years of work experience

(B=.040, s.e.=.007, p<.001) were positive and significant predictors of Combined TSE. Age (B=.-015, s.e.=.006, p=.02) and completion of a teacher education program (B=-.301, s.e.=.104, p=.004) were negative and significant predictors.

Introducing the teacher-level experiential variables in the second model explained an additional 10.2% of variation in Combined TSE, resulting in 10.5% of the variance, and this change in R^2 was significant, F (8, 38,133)= 557.49, p<.001. In this model, gender (B=.345, s.e.=.064, p<.001) and years of experience (B=.029, s.e.=.006, p<.001) remained positive and significant predictors. Completion of a teacher program (B=-.685, s.e.=.099, p<.001) and age (B=-.015, s.e.=.006, p=.013) remained negative and significant predictors. One of the new variables, scale of need for professional development (B=-.079, s.e.=.002, p<.001), was a negative and significant predictor. Three of the new variables that were introduced in this model, required professional development (B=.757, s.e.=.075, p<.001), total participation in professional development (B=.364, s.e.=.013, p<.001), and total time reading for work (B=1.223, s.e.=.033, p<.001), were positive and significant predictors of Combined TSE.

RQ2: How do teacher-level and school-level factors influence teacher self-efficacy?
Adding school-level characteristics to the third block in the regression model explained an additional .8% of the variation in Combined TSE, resulting in 11.2% of the variance, and this change in R^2 was significant, F (10, 38,131)=482.42, p<.001. Both variables added in this model, public school (B=.512, s.e.=.068, p<.001) and school size (B=-.262, s.e.=.017, p<.001), were significant. In this model, all of the teacher level variables remained significant (p<.001). Age and completion of a teacher program remained negative and significant predictors.

In the fourth model, introducing school-level environmental characteristics to the regression model explained an additional 9.7% of the variation in Combined TSE, resulting in

20.9% of the variance, and this change in R^2 was also significant, $F (15, 38,126)=671.02$, $p<.001$. Both of the principal perception of capacity variables, instruction hindered ($B=.136$, s.e.=.012, $p<.001$) and student learning hindered ($B=.056$, s.e.=.010, $p<.001$), were significant. Both of the quality assurance variables, professional control ($B=.332$, s.e.=.018, $p<.001$) and management approach ($B=.091$, s.e.=.016, $p<.001$), were significant positive and significant predictors of TSE. Finally, the school achievement mean reading score ($B=-.023$, s.e.=.000, $p<.001$) was a negative and significant predictor of Combined TSE.

In the fourth model, age, completion of a teacher education program, and scale of professional development need remained significant and negative predictors. Gender, work experience, required professional development, time spent reading for work, and professional development remained positive and significant predictors of Combined TSE. School size remained negative but was not significant in this model ($p=.137$).

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)? A fifth model measuring the macro, or country-level characteristics, added two variables, OECD status and country student achievement mean scores in reading, and explained an additional 10% of the variation in Combined TSE. The change in R^2 was significant, $F (17, 38,124)=1,002.11$, $p<.001$. In this model, OECD status ($B=-.069$, s.e.=.080, $p=.390$) was not significant. Country student achievement mean scores in reading ($B=-.052$, s.e.=.001, $p<.001$) was a negative and significant predictor of Combined TSE.

In the final model, gender, years of work experience, required participation in professional development, time reading for work, participation in professional development, principal perception of capacity student learning, quality assurance professional control, and quality assurance management approach, were all positive and significant predictors. Age,

completion of a teacher program, scale of need for professional development, public school, and school size, were all negative and significant predictors. Principal perception of capacity–instructional staff, and school mean student achievement mean scores in reading were not significant in this model. Comparisons of the changes in the standardized betas across the models can be found in Appendix B.

Together the five models accounted for 30.9% of the variance in Combined TSE. Across the models, the addition of the individual- and school-level environmental factors and the country-level characteristics make this model the strongest predictor of Combined TSE.

Summary

This chapter outlined how the PISA data was organized and utilized to study the multidimensional factors related to teacher self-efficacy, utilizing data collected through a large-scale international assessment to identify the predictive relationships between different types of variables and TSE.

The first research question asked whether there is a relationship between individual characteristics and TSE. Across the three measures of TSE, there are statistically significant, but very small relationships between the individual characteristics of age, gender, years of work experience, and participation in a teacher education program. When experiential factors such as required professional development, time spent reading for work, participation in professional development, and scaled identification of need for professional development were added, the model was strengthened across all three measures. These findings allow for us to reject the first two null hypotheses and find that individual teacher characteristics and experiential factors are antecedents for measures of TSE.

The second research question added school-level characteristics and environmental factors, with similar patterns from the first two models. The school level characteristics such as identification as a public school and school size were consistently negative across the three measures and were significant in the final model, although school size had a very small relationship to TSE. The school environmental characteristics such as principal perception of capacity, the quality assurance models, and the school mean achievement in reading had statistically significant differences in all three measures. Similar to the previous question, the null hypotheses may be rejected, and the findings support the notion that school-level characteristics and environmental factors are antecedents to TSE across all three measures.

The third research question included country-level characteristics in the analysis, studying whether the country's OECD status and the country level mean student achievement in reading was significant. In all three measures, the country level characteristics had statistically significant impacts on the TSE. The final null hypothesis can be rejected, with the analysis supporting the finding that country-level factors also serve as antecedents to TSE.

The final chapter will explore the findings and patterns identified in this analysis, their relationship to current literature, how the findings may be used in academia and educational practice, and how further research can continue to strengthen the study of teacher self-efficacy.

Table 4.12: Teacher Self-Efficacy Instruction and Engagement Subscale

Variable	r	Final B	Final β	t	sr²	R	R²	ΔR²
Model 1: Micro/Individual Characteristics						.047	.002	.002***
Female	.030***	.351	.038***	8.939	.001			
Age	.021***	-.015	-.035***	-4.150	.000			
How many years of work experience do you have?	.029***	.035	.077***	9.171	.002			
Did you complete a teacher education or training programme?	-.017***	-.215	-.015***	-3.565	.000			
Model 2: Micro/Individual Environment						.324	.105	.103***
Are you required to participate in professional development activities?	.086***	.397	.038***	8.654	.002			
How much time do you spending reading for your work out of your classes?	.234***	.672	.144***	33.050	.016			
Participation in development experience over the last 12 months scale	.200***	.298	.165***	36.706	.021			
Composite current need for professional development scale	-.164***	-.041	-.120***	-28.053	.019			
Model 3: Meso/School Level Characteristics						.338	.114	.009***
Public School	.042***	-.579	-.060***	-12.338	.002			
School Size	-.076***	-.058	-.023***	-5.281	.000			
Model 4: Meso/School Level Environment						.463	.215	.101 ***
Principal Perception of Capacity-Instructional Staff	.068***	-.003	-.002	-.369	.000			
Principal Perception of Capacity-Staff Behaviors	.060***	.043	.031***	6.703	.001			
Quality Assurance: Professional Control	.217***	.162	.071***	14.129	.003			
Quality Assurance: Management Approach	.149***	.054	.025***	5.101	.001			
School Achievement: Mean Reading Score	-.300***	.001	.019*	3.007	.000			
Model 5: Macro/Country Level Environment						.572	.327	.113 ***
OECD Status	-.317***	.166	.017**	2.996	.000			
Country Student Achievement: Mean Reading Score	-.474***	-.040	-.477***	-73.114	.080			

Note:*p<.05, **p<.01, ***p<.001

Table 4.13: TSE Student Behavior and Classroom Management Subscale

Variable	r	Final B	Final β	t	sr²	R	R²	ΔR²
Model 1: Micro/Individual Characteristics						.065	.004	.004***
Female	.027***	.136	.028***	6.023	.001			
Age	.032***	-.012	-.053***	-5.624	.001			
How many years of work experience do you have?	.053***	.025	.108***	11.625	.003			
Did you complete a teacher education or training programme?	.002	-.021	-.003	-.599	.000			
Model 2: Micro/Individual Environment						.269	.072	.068***
Are you required to participate in professional development activities?	.081***	.237	.044***	9.001	.002			
How much time do you spending reading for your work out of your classes?	.158***	.210	.087***	17.998	.007			
Participation in development experience over the last 12 months scale	.142***	.097	.104***	20.852	.009			
Composite current need for professional development scale	-.183***	-.027	-.150***	-31.909	.022			
Model 3: Meso/School Level Characteristics						.275	.075	.003***
Public School	.015***	-.203	-.040***	-7.518	.001			
School Size	-.043***	-.019	-.015*	-2.970	.000			
Model 4: Meso/School Level Environment						.362	.131	.056 ***
Principal Perception of Capacity-Instructional Staff	.067***	.005	.006	1.101	.000			
Principal Perception of Capacity-Staff Behaviors	.066***	.027	.038***	7.307	.001			
Quality Assurance: Professional Control	.168***	.064	.054***	9.650	.002			
Quality Assurance: Management Approach	.119***	.026	.023***	4.168	.000			
School Achievement: Mean Reading Score	-.205***	-7.14E-6	.000	0.034	.000			
Model 5: Macro/Country Level Environment						.424	.179	.048 ***
OECD Status	-.254***	-.237	-.046**	-7.472	.001			
Country Student Achievement: Mean Reading Score	-.319***	-.012	-.279***	-38.689	.032			

Note: *p<.05, **p<.01, ***p<.001

Table 4.14: Teacher Self-Efficacy Full Survey

Variable	r	Final B	Final β	t	sr^2	R	R^2	ΔR^2
Model 1: Micro/Individual Characteristics						.055	.003	.003***
Female	.032***	.494	.038***	8.604	.001			
Age	.028***	-.026	-.041***	-4.766	.000			
How many years of work experience do you have?	.040***	.060	.093***	10.758	.002			
Did you complete a teacher education or training programme?	-.012**	-.243	-.012**	-2.926	.000			
Model 2: Micro/Individual Environment						.324	.105	.102***
Are you required to participate in professional development activities?	.090***	.640	.043***	9.673	.002			
How much time do you spending reading for your work out of your classes?	.222***	.880	.133***	29.962	.016			
Participation in development experience over the last 12 months scale	.193***	.396	.154***	33.929	.021			
Composite current need for professional development scale	-.182***	-.068	-.139***	-32.171	.019			
Model 3: Meso/School Level Characteristics						.335	.112	.008***
Public School	.034***	-.786	-.057***	-11.756	.002			
School Size	-.069***	-.076	-.022*	-3.602	.000			
Model 4: Meso/School Level Environment						.457	.209	.097***
Principal Perception of Capacity-Instructional Staff	.072***	.002	.001	.298	.000			
Principal Perception of Capacity-Staff Behaviors	.066***	.070	.036***	7.649	.001			
Quality Assurance: Professional Control	.214***	.225	.070***	13.606	.003			
Quality Assurance: Management Approach	.149***	.081	.026***	5.270	.000			
School Achievement: Mean Reading Score	-.287***	.001	.013	1.525	.000			
Model 5: Macro/Country Level Environment						.556	.309	.100***
OECD Status	-.316***	-.069	-.005	-.867	.000			
Country Student Achievement: Mean Reading Score	-.451***	-.052	-.439***	-66.121	.080			

Note: *p<.05, **p<.01, ***p<.00

CHAPTER V: FINDINGS AND DISCUSSION

In 1977, Gibson and Dumbo applied the theory of self-efficacy in a RAND study focused on urban student literacy. Over four decades later, the study of teacher self-efficacy (TSE) has generated a significant body of research relating to a variety of topics including teaching quality, professional learning and development, teacher career span, and student learning. Throughout this time, researchers have continually worked to connect teacher self-efficacy to student achievement, often with inconsistent or weak findings (Burić & Kim, 2020; Guo et al., 2012; Klassen & Tze, 2014; Künsting et al., 2016). Prior research has been criticized for the use of small sample sizes, inconsistent measurement tools (Klassen, 2009; Morris et al., 2017; Zee & Koomen, 2016). The introduction of computer-based, large-scale international assessments such as PISA, PIRLS, and TALIS now make it possible for organizations such as OECD to support the multidimensional study of student education systems and deeper analysis of international comparisons of teacher efficacy using a wide range of variables.

The purpose of this study was to investigate the relationship between individual and environmental factors and teacher self-efficacy (TSE) utilizing the data gathered through the Programme for International Student Assessment (PISA) assessment conducted in 2018. An initial sample size of 107,367 participating teachers in 6,128 schools across 19 countries provides a unique opportunity for a global perspective. Through the use of data gathered from three measures used during the 2018 PISA administration this study evaluates antecedents to TSE not previously studied at an international level.

Research Questions

RQ1:   What is the relationship between individual characteristics and teacher self-efficacy (TSE)?

- How do individual characteristics such as age, gender, work experience, and completion of a teacher training program relate to TSE?

- How do individual professional development experiences relate to TSE?

RQ2: How do individual characteristics and school-level factors influence teacher self-efficacy (TSE)?

- How are the teacher-level responses influenced by characteristics of the school such as size and school type (public/private)?

- How are the teacher-level responses influenced by school-level environmental experiences such as the principal's perception of school's instructional capacity, and the quality assurance (accountability) approaches?

- How are the teacher-level responses influenced by school level student achievement?

RQ3: How do individual characteristics, school-level factors, and country-level factors influence teacher self-efficacy (TSE)?

- How is TSE influenced by a country's economic development status as defined by OECD status?

- Is there a relationship between the country's level of student achievement and TSE?

Summary of Results

Method

A blocked hierarchical regression analysis was conducted using SPSS Subscription Edition 2018 (Version 26). The model developed utilizes 17 independent variables across five blocks, building on a theoretical shift from micro-, meso-, and macro-level variables. This model was run on three dependent variables, all framed based on the Teacher Sense of Efficacy Survey (TSES) developed by Tschannen-Moran and Woolfolk Hoy (2001). The dependent variables

utilize the same questions, but are configured as the Instruction and Engagement Subscale, Student Behavior and Classroom Management Subscale, and the Full Survey.

Summary of Findings

The first model tested the relationship between individual characteristics and TSE. Across the three measures of TSE, there are statistically significant, but very small relationships identified between the individual characteristics of age, gender, years of work experience, and participation in a teacher education program, accounting for less than .5% (p<.001) of the variation in each of the scales.

When experiential factors such as whether professional development participation was required, time spent reading for work, participation in professional development, and scaled identification of need for professional development were added in the Micro/Individual Model, the model was strengthened across all three measures, accounting for 7.2% of variation on the Student Behavior and Classroom Management Subscale, 10.3% of the variation on the Instruction and Engagement Subscale, and 10.5% of the variation on the Full Survey (p<.001).

The fourth variable in the model, teachers' composite need for professional development, had a significant negative relationship to TSE across all three measures. These findings support the hypotheses that individual teacher characteristics and experiential factors have a relationship to TSE, although this relationship is very weak.

The second research question added two models, the school-level characteristics and school-level environment. The Meso/School Level Characteristics Model included two variables, public school and school size. Across the three measures, this model had a very small impact, accounting for less than 1% of the variation in TSE across the three models (p<.001). When only the characteristics were considered, there was a positive relationship between public schools and

TSE; however, when the additional models were added the relationship became negative, confirming prior research findings (Butucha, 2013; Moradkhani & Haghi, 2017; Zamir et al., 2017). School size has a weak negative relationship similar to other research findings (Fackler et al., 2021; Tschannen-Moran & Johnson, 2011; Tschannen-Moran & Woolfolk Hoy, 2007).

The Meso/School Level Environment Model adds school environmental characteristics such as the principal perception of capacity, the quality assurance models, and the school mean achievement in reading. The addition of these variables returned significant differences in all three measures, predicting 13.1% of TSE on the Student Behavior and Classroom Management Subscale, 21.5% of TSE on the Instruction and Engagement Subscale, and 20.9% on the Full Survey ($p<.001$).

Among the variables, the principal's identification of staffing needs was small and ultimately not significant in the final model. However, the principal's perception of staff capacity and the type of management approach used by principals relating to teaching and learning behaviors were both positive and significant in all three of the measures.

In this model, the most significant finding was a negative relationship between TSE and school-level student achievement on the reading assessment. This inverse relationship marks a deviation from prior research that identified positive relationships between TSE and student achievement (Armor et al., 1976; Ashton et al., 1983; Fackler & Malmberg, 2016; Guo et. al, 2012; Klassen & Tze, 2014). It is important to note that the publicly available PISA questionnaire did not allow for a direct relationship between a teacher's response and individual students to be established, but it does allow for school-level data to be aggregated as was done for the purpose of this study.

The final research question utilizes the Macro/Country Level Environment Model and includes country-level characteristics in the analysis, studying whether the country's OECD status and the country level mean student achievement in reading was significant. In all three measures, the country level characteristics had statistically significant impacts on the TSE. The addition of this model accounts for 30.9% of TSE on the Full Survey, 32.7% of TSE on the Instruction and Engagement Subscale, and 17.9% of TSE on the Student Behavior and Classroom Management Subscale.

In these findings, OECD status behaves differently in each of the three measures. When measured with the Instruction and Engagement subscale, it has a small but positive relationship (ß=.017, p<.01). Within the Student Behavior and Classroom Management Subscale and the Full Survey it is not significant. Similar to the fourth model, the country-level student achievement returns a significant negative relationship to TSE on all three models with a significant negative relationship on the Instruction and Engagement subscale (ß=-.477, p<.001) and the Full Survey (ß=-.439, p<.001). In this model, school-level student achievement becomes insignificant in all three models, indicating the power of the country-level student achievement variable in relation to TSE.

Discussion

Measures of TSE

The three scales developed from the TSES tool provided different levels of predictive ability, with the Classroom Instruction and Engagement Scale providing the most significant. The weak relationship between the variables and the Student Behavior and Classroom Management Subscale supports extensive research regarding the difference between managing the processes and procedures of a classroom and the work of teaching and learning (Hattie,

2008). It demonstrates the variability between the type of questions and teacher responses, supporting Bandura's theory that measuring self-efficacy is deeply dependent on the specificity of the task (Bandura, 1997). For researchers and policymakers studying TSE, this finding supports ensuring policies and research related to TSE may be more effective when directly related to the work around instructional practice, as opposed to classroom management. Similarly, it may indicate the value of investing in teacher education centered on pedagogy and practice over classroom management.

Individual vs. Environmental Antecedents

Across the five models, the characteristics included in Model 1 and Model 3 were weaker predictors than questions that related to the environment in Model 2 and Model 4. Demographics were weak predictors, consistent with prior research. While several prior studies produced mixed findings regarding gender and TSE, this study identified small but positive relationships between female teachers and TSE in each of the three measures, confirming other findings (Fackler et al., 2021; Perera et al., 2019; Tschannen-Moran & Johnson, 2011). In addition, while prior studies did not identify age as a significant variable to TSE (Colodarci & Breton, 1997; Tschannen-Moran & Woolfolk Hoy, 2007), age was negatively related to TSE in this study across all three measures. This finding supports Tschannen-Moran and Woolfolk Hoy's (2007) suggestion that there is not theoretical support for such individual characteristics to be related to TSE, but their relationship suggests that they may be related to vicarious experiences that impacted the teachers' decisions to work in the educational sector (p. 952).

Completion of a teacher education program also had a weak negative relationship to TSE across all three measures and was only significant in the Instruction and Engagement Subscale and the Full Survey. This finding was similar to the research of Swan et al. (2011) that identified

declines in TSE after participation in a teacher education program. This may also be indicative of a relationship between a teacher education program and teacher knowledge and understanding of the challenges of instructional practice.

Professional development participation variables were stronger predictors, although whether a teacher was required to participate was the weakest among the variables in the model while the time spent reading outside of work was among the strongest positive predictors. This finding confirms prior research related to participation in professional development and TSE (Althauser, 2015; Dixon et al., 2014; Henson, 2001; Palmer, 2011; Posnanski, 2002; Tschannen-Moran & McMaster, 2009; Yang, 2020). It also suggests that external mandates are less impactful than self-directed learning opportunities.

The findings suggest that the school environment, inclusive of how the principal perceives staff instructional capacity and the type of leadership style utilized by the principal, is related to TSE, with school quality assurance protocols that support professional control as one of the stronger predictors at the meso-level. This finding supports previous research related to the principal's role in TSE (Bendikson et al., 2012; Nir & Kranot, 2006; Kurt et al., 2012; Sehgal et al., 2017). The Quality Assurance: Professional Control variable was the strongest positive predictor of the model in the full regression, indicative of higher TSE among teachers when school leaders prioritize school level policies and interactions (Leithwood & Earl, 2000).
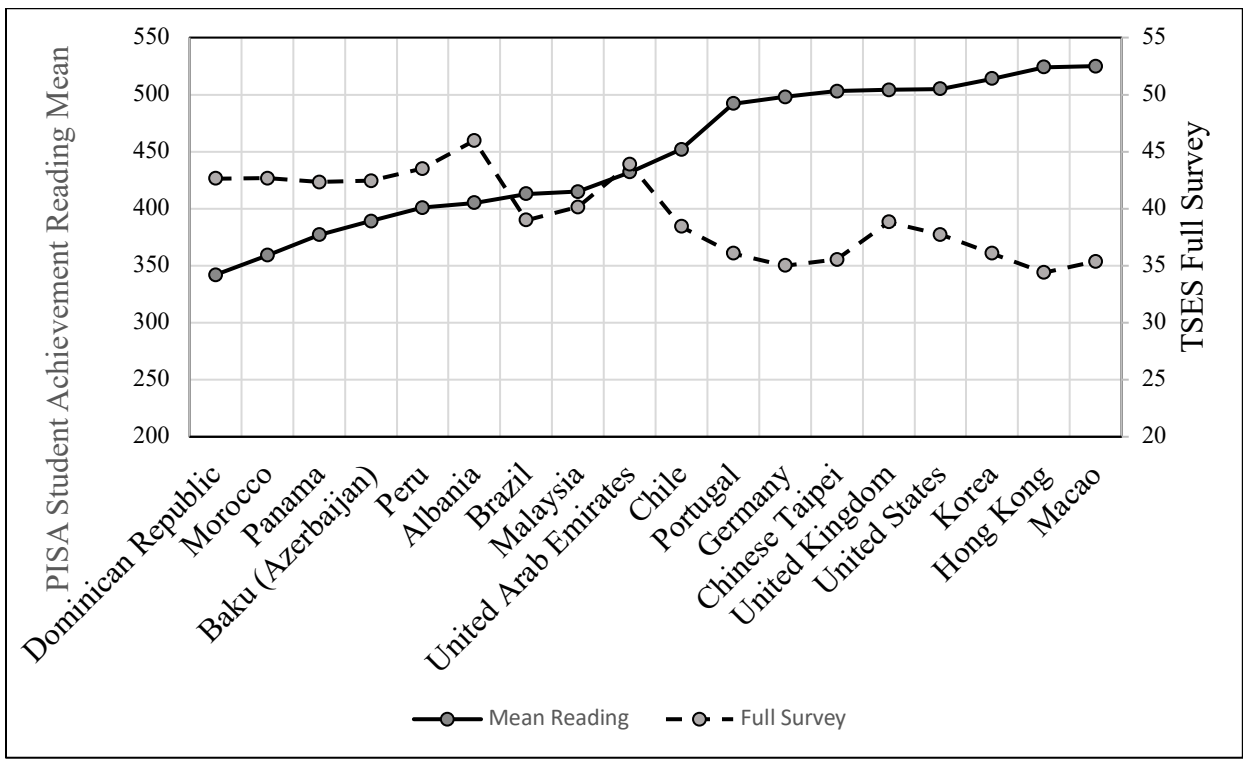
Student Achievement

The most unexpected finding is the relationship that appears between student achievement and TSE. Throughout the body of research, TSE is framed as having a positive relationship to student achievement, often categorized as an indirect relationship based on factors such as teaching practices and teaching quality. Caprara et al., (2006) identified a relationship

between TSE, teacher job satisfaction, and teachers' perception of their impact on student academic achievement and found a modest relationship between student academic achievement and TSE. Klassen and Tze (2014) found TSE has a stronger relationship to teaching effectiveness than to student achievement.

In this regression analysis, country-level student achievement, serves as the strongest predictor of TSE, reducing or eliminating the relationship to school-level student achievement across all three measurement scales. This relationship indicates that the country student achievement variable serves as an important predictor for TSE in an international comparison and may serve as a latent variable related to the educational and economic policies and programs informing educational systems. To provide a visual perspective on this relationship, Graph 5.1 demonstrates the research finding. As depicted in the graph, the countries with the highest level of student achievement are among those with the lowest means of TSE.

It is important to recognize that some researchers have challenged the ability to compare TSE responses across significantly different cultures, as how a teacher may respond might be deeply tied to cultural beliefs (Ho and Hau, 2004; Klassen et. al, 2009; Lin et al., 2002). However, as Graph 5.2 illustrates, the full survey means do not reveal a pattern in the TSE responses, with European countries such as Germany and Portugal demonstrating a close TSE mean to the Asian countries Hong Kong and Macao.

Graph 5.1: TSE & PISA Student Achievement Reading Mean by Achievement (lowest to highest)

Graph 5.2 TSE & PISA Student Achievement Reading Mean by TSE Mean (lowest to highest)
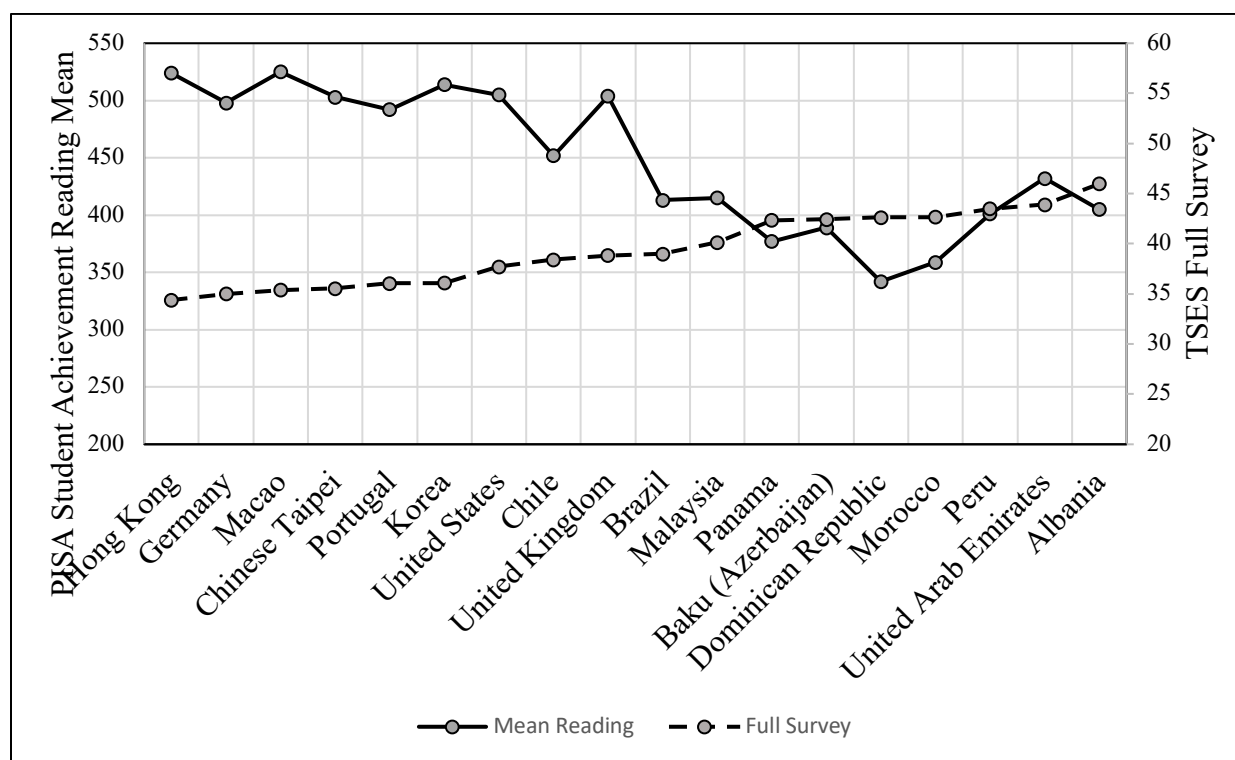
While these findings may be perceived as undermining the value of the TSES or negating the predicted positive relationship between teacher self-efficacy and student achievement outcomes, the patterns that emerge lead to questions about how the differences in economic and social development of international educational systems relate directly to teacher perception of their own skills and capacity.

One hypothesis is that the differences in the relationship between teacher self-efficacy and student outcomes between countries is indicative of information asymmetry (Akerlof, 1970) a condition in which those with less information and knowledge overrate their skills due to a lack of expertise. In terms of self-efficacy, Bandura (1977) predicted "when performance requirements are ill-defined, people who underestimate the situational demands will display positive discrepancies between self-efficacy and performance attainments" (p. 203). The result is

the Dunning-Kruger effect, the belief that "the skills that engender competence in a particular domain are often the very same skills necessary to evaluate competence in that domain" resulting in inflated perceptions of self-efficacy (Kruger & Dunning, 1999). As knowledge increases, so does skill, leading to a deflation in perception of efficacy until such time as the increase in knowledge and skill lead to a metacognitive ability to recognize the growth.

In terms of the TSE and student outcomes, the inverse relationship found between TSE and student academic outcomes in two of the three measures may be indicative of stronger teacher understanding or a more stringent standard of practice related to the student engagement or instructional strategies constructs.

One example occurs in the TSES question, "To what extent can you do: Craft good questions for my students." In the case of the highest achieving countries, the inverse relationship may be due to a stronger teacher understanding or standard of practice related to crafting cognitively challenging questions for students. A teacher that can differentiate between simple knowledge retrieval and cognitively complex questions may also rate themselves more critically than a teacher that has not received training related to inquiry and engagement in the classroom.

This interpretation is supported across other levels of the model. For example, the inverse relationship between age and TSE, while weak, supports a hypothesis that older teachers are likely to have more experience in the classroom, resulting in a deeper understanding of the complexities of classroom instruction and student achievement. Similarly, teachers that demonstrated higher levels of need for professional development had lower TSE, indicating knowledge of the need for learning may be aligned to student achievement.

This perspective supports Wheatley's (2002; 2005) theory that teacher doubts are critical to educational reform and progressive educational systems because such doubts drive productive professional learning and growth. While Wheatley (2002) argues that the Tschannen-Moran and Woolfolk Hoy TSES instrument does not directly support the presence of teacher doubt, this study's model also supports Wheatley's hypothesis, despite utilizing a measure that does not directly address teacher doubt. In addition, Wheatley (2005) notes that the global measures of TSE may minimize the relationship between doubt and TSE because such measures miss the task and content specificity necessary for a deeper understanding. The PISA data has limitations that do not allow for a deeper analysis of context or task specificity, but it may support the addition of the appropriate survey questions to provide a better understanding of the context.

Implications

The teacher-student interaction that happens in classrooms throughout the world every day is a deeply complicated and multidimensional interaction between micro-level factors such as personal experiences, meso-level factors such as school environment, and macro-level factors such as governmental policy. This study utilized data collected through an international large-scale assessment to identify how these levels of antecedents might serve as predictors of TSE and found the significant role that environmental factors at the school and country levels may play in TSE. The findings have implications for policymakers and practitioners.

Educational Policy

The OECD presents PISA as a policymaking tool, noting in its policy analysis that its goal is to help policymakers make informed decisions (OECD, 2018; 2020; 2021b). Skaalvik and Skaalvik (2010) hypothesize that external controls are largely related to public perceptions of teaching efficacy and norms for establishing teacher practice standards may undermine

autonomy, thereby negatively impacting TSE. As Graph 5.2 illustrates, countries that are more economically developed with stronger systems of education may provide such norms, resulting in the trends identified in this research. Additionally, the significant relationship that country-level student achievement has with TSE indicates that governmental policy may play an important role on teacher perceptions of their work at the school level.

Germany provides an interesting case study on the relationship between PISA, policymaking, and student achievement. Since the first administration of the PISA in 2000, Germany has shown some of the most significant growth in student achievement. The impact of the initial data and the resulting shifts in federal- and state-level education policy have been widely documented (Niemann et al., 2017; Neumann et al., 2010; Ringarp, 2016). Such policy shifts included a balance of increased accountability, teacher training, and school-level autonomy (Niemann et al., 2017). Between 2000 and 2015, Germany moved from being ranked in the 20th position to the 10th position, and student mean scores in reading grew 24 points (Niemann et al., 2017). In the 2018 administration, Germany slipped back down in the rankings, but student mean scores were 14 points higher than in the original administration (OECD, 2019a).

Teacher evaluation protocols are one example of country and local policies that have an impact on TSE. The relationship between external evaluation criteria and teacher concepts of self-efficacy exemplifies the types of expectations that Bandura (1977) identified as formative for self-efficacy: enactive or performance accomplishments, vicarious experiences, exhortative or verbal persuasion, and emotional or physiological states. If there are strong frameworks for evaluating teacher practice at the country or school level, those frameworks may be a critical part of how teachers develop a sense of efficacy.

Such a perspective supports Klassen and Tze's (2014) finding where TSE was strongly associated with evaluated teaching performance. While previous research has connected positive feedback and coaching related to the evaluation process as having a positive relationship with TSE, no studies were identified that directly compared types of evaluative environments and base levels of TSE prior to such feedback (Palmer, 2011; Mireles-Rios & Bechio, 2018; Morris et al., 2017; Schunk & Pajares, 2009; Smith et al., 2020).

In addition, many of the most commonly used evaluation frameworks is the United States include a component of self-reflection on efficacy. In one example, the Danielson Framework includes a rating specifically on a teacher's ability to reflect on their practice and identify the probable success of different instructional strategies (Danielson, 2013). The findings of this study lead to questions about the value of rating a teacher on their ability to reflect, and whether school leaders have the appropriate training to value a teacher's self-doubt during such a reflective exercise.

Future research around the teaching environment's impact on TSE would benefit from a deeper look at how evaluation protocols and expectations relate to TSE throughout the career span of a teacher, and how the school environment's use of such protocols impacts TSE.

Practitioners

Professional Development. Studying the relationship between teacher self-efficacy and school-or country-level variables may shed further light on how evaluation tools and teacher quality frameworks can be more effectively used to support inquiry-based professional development. This is supported by the work on learning motivation described as the growth mindset (Dweck, 1999; Dweck & Leggett, 1988). Yaeger and Dweck (2020) note that ongoing controversy around how teachers can support the development of a student growth mindset may

be rooted in the lack of research around addressing teacher mindsets about their own practice. If teachers are expected, and even encouraged to have only positive views of their own practice, the ability to address challenges and failures through critical inquiry and reflection may be diminished.

This study found positive relationships between participation in professional development, supporting extensive prior research. However, one variable, the composite scale of professional development need, had an inverse relationship to TSE. Such a finding indicates teachers who were more likely to indicate interest or need in a variety of professional learning topics were more likely to report lower TSE.

One explanation for this finding may be teachers with lower TSE are better able to identify areas of need related to instructional practice, making them more likely to demonstrate the growth mindset related to their practice. If teachers' interest in professional learning is demonstrative of a growth mindset, it may weaken Bandura's (1997) assertion that self-efficacy beliefs are resistant to change. It is logical then to propose that student achievement may be higher for teachers that recognize their own need to learn. For practitioners, this demonstrates the need to develop educational leaders that understand and value such a growth mindset, and honor self-doubt for the purpose of setting goals and supporting adult learning.

Impact of the Teaching Environment. While there were clear differences in how individual variables performed, there were also similarities among the groups of variables. In all three scales, the characteristics showed significantly less impact on TSE than the environmental variables. This provides an indication of the power of the environment in which a teacher teaches, and how that environment contributes to teacher perspectives relating to their own instructional practice. The environment serves to norm teacher practice, providing the baseline

by which a teacher evaluates their own knowledge and skill levels in relationship to the norms set at the school level, the country level, or some combination thereof.

Prioritizing Instructional Practice

Comparing the measures provided a critical look at the TSES tool and the role that different constructs play in the measurement of TSE. While teacher knowledge and skill around classroom management and student discipline are important to the way the classroom functions, this research suggests that teacher self-efficacy may be better measured through a lens directed at student engagement and instructional practice. While the full survey included all three of the original constructs developed by Tschannen-Moran and Woolfolk Hoy (2001), researchers interested in TSE and its relation to student achievement should be aware of the relationships between teachers and these constructs and the possibility they may reflect different concepts of teacher knowledge and skill.

Limitations

This study has several limitations that should be considered as the results are reviewed. PISA notes that the data is subject to nonsampling and sampling errors (OECD, 2021a). The nonsampling errors are related to data collection, nonresponse bias, data processing, and reporting (OECD, 2021a). The sampling errors may occur when the sample does not statistically represent the population, thereby creating a degree of uncertainty, or sampling variance (OECD, 2021a). PISA provides the standard errors for their data through their website and standard errors for the data used in this study are provided in the descriptive statistics for each variable in Chapter 4.

The Teacher Questionnaire, expanded in the 2018 PISA administration, was not mandatory and countries that administered the Teacher Questionnaire did so based on their own

choice, with significant differences in the administration. In the original sample, Spain is

significantly overrepresented in the sample, representing 21% of the original total sample. The

framework for this study was developed inclusive of Spain's available data; however, due to

testing inconsistencies, Spain's results from the reading section of the PISA were not included in

the final PISA data. For the regression models, Spain was eliminated from the data set. It is

critical to note that Spain's data for the mathematics and science sections remained valid, and

future research including those testing domains should include Spain (See Appendix C).

Likewise, the United Arab Emirates and Brazil are moderately overrepresented in the sample.

Another limitation is the data is self-reported, leading to perception-bias. For both the

teachers and the principals completing the survey, there is a risk that the data reflects how these

individuals wish to be perceived, or how they feel they should respond as opposed to actual

perceptions. Further, in the case of the principals' responses, PISA does not methodically

compare what is reported as "mandatory" in a policy, and therefore "mandatory" may be a

perception rather than an actual policy.

Future Research

Teacher Self-Efficacy Research

Tschannen-Moran and Johnson (2011) argue that teachers with weak self-efficacy are

likely to persist in their beliefs as a cycle of self-defeat. Such an assertion supposes that teachers

with lower TSE engage in poorer teacher practice aligned to the belief that higher TSE is related

to productive classroom practice. This research indicates that external contexts have a

relationship with TSE that may weaken this argument, and that lower TSE may not be indicative

of poorer practice but of stronger expectations of practice. Future research that is inclusive of

measures of doubt may provide the opportunity to further explore the multi-dimensional relationship between TSE and student achievement.

If lower TSE can be indirectly related to higher student achievement, it may have an impact on a variety of theoretical models that relate to TSE. One example is the theory of collective efficacy, proposed by Goddard, Hoy, and Woolfolk Hoy (2000), in which the beliefs of the collective regarding others' skill and knowledge serve as an indirect influence on student achievement. The PISA teacher questionnaire includes one question that asks teachers whether instruction is hindered by "inadequate or poorly qualified teaching staff," and while the regression model in this study did not include that item in the level of analysis, it may provide a point of reference for connecting perspectives of collective efficacy to individual teacher perspectives.

Use of the Multi-Tiered Model

The model used in this study provides an exploration of complex relationships across the multi-dimensional education system. This model is dependent on the availability of three levels of data; in this case these were the teacher-level, school-level, and country-level. Future research may, however, identify inter-country indicators that would allow for an exploration of local policy or economic indicators. The PISA data includes region and stratum identifiers which may provide an opportunity for further examination about more localized policy models and teacher experiences.

Additionally, the relationships identified in this study lead to questions about how the external environment, including economic and social systems, impact teacher knowledge, skill, and sense of self-efficacy. As Bandura's (1986) triadic reciprocation model suggests, environmental factors have an impact on individual self-efficacy, and while research has

previously tied such environmental factors to teacher education, the findings from this study suggest that indirect impacts on the environment serve as factors influencing teacher self-efficacy. Indirectly policies and structures developed at federal levels of government are representative of community expectations relating to teaching efficacy. Federal policies inclusive of assessment policies, teacher education standards, student learning standards, and teacher certification policies may have a direct impact on how teachers perceive their own knowledge and skill. Additional research about the relationship between external environments and teacher self-efficacy may guide policymakers as they work to develop and strengthen educational systems.

Content Area

PISA is an international triennial large-scale assessment that measures student knowledge and understanding of three domains: reading, mathematics, and science. Each cycle identifies a major domain, with research-based revisions to standards and a larger array of question types related to the major domain (OECD, 2021b). Reading was the major domain of the 2018 administration, and therefore was used as the achievement indicator in this research study. Future research may include the use of the science and mathematics achievement data from the 2018 administration, or future administrations that identify them as major domains may provide the opportunity for content-related longitudinal research.

Teacher Questionnaire

This study points to a critical need to continue and expand the use of the Teacher Questionnaire among countries participating in PISA. While Germany's shift, and those of other countries, has been documented via student achievement data, there has been a dearth of such large-scale, teacher-centered data that is directly tied to a single instrument. However, if this

questionnaire is continued, it will provide valuable longitudinal perspective that will support further research. The research findings also serve to encourage OECD to continue the use of the teacher questionnaires in the PISA assessment and encourage more partner countries to participate in the survey tool. In use with the student questionnaire and school questionnaire, there remains a rich body of research for analysis.

Conclusion

Over forty years after its introduction, researchers across the social and behavioral disciplines continue to study the powerful construct of self-efficacy as a means to supporting growth and development for individuals and organizations. Since that time, the education sector has developed a significant body of research around teacher self-efficacy. The 2018 PISA administration provided the first large-scale international data assessment that included multi-tiered data for the student, teacher, and school leader.

This study explored the dynamic nature of TSE, and the relationships that different layers of the educational system have with this construct. The findings suggest that country-level economics and policy have an impact on TSE, and future models comparing TSE across systems would benefit by including some measure to control for the teaching environment.

Finally, the findings from this study challenge future researchers to move beyond the expectation that a higher TSE represents stronger instructional practice. As educators work to shape students who demonstrate skills such as a growth mindset, it is critical the educational system that supports them place a greater value on recognizing and honoring self-doubt and the developmental processes related to professional learning and practice.

REFERENCES

Akerlof, G.A. (1970). The market for lemons: quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 488–500. Retrieved from http://links.jstor.org/sici?sici=0033-5533%28197008%2984%3A3%3C488%3ATMF%22QU%3E2.0.CO%3B2-6

Althauser, K. (2015). Job-embedded professional development: its impact on teacher self-efficacy and student performance. *Teacher Development*, *19*(2), 210-225. doi: 10.1080/13664530.2015.1011346

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L.,Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools* (R-2007-LAUSD). Santa Monica, CA: Rand Corp. Retrieved from https://files.eric.ed.gov/fulltext/ED130243.pdf

Ashton, P., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self-or norm-referenced construct? *Florida Journal of Educational Research, 26*(1), 29-41. Retrieved from https://files.eric.ed.gov/fulltext/ED323221.pdf

Ashton P. T., Webb, R. B., & Doda, N. (1983). *A study of teacher's sense of efficacy. Final Report to the National Institute of Education, Executive Summary*. Gainesville: Florida University. Retrieved from https://files.eric.ed.gov/fulltext/ED231833.pdf

Atta, N., Ahmad, N., Ahmed, M., & Ali, Z. (2012). Role of gender and teaching experience on teachers' self-efficacy. *Language in India, 12*(9), 251-262.

Balingit, M., & Van Dam, A. (2019). US students continue to lag behind peers in East Asia and Europe in reading, math and science, exams show. *Washington Post*, 3 December 2019. https://www.washingtonpost.com/local/education/us-students-continue-to-lag-behind-peers-

in-east-asia-and-europe-in-reading-math-and-science-exams-show/2019/12/02/e9e3b37c-153d-11ea-9110-3b34ce1d92b1_story.html

Bandura, A. (not dated). *Bandura's Instrument Teacher Self-Efficacy Scale*. Retrieved from http://anitawoolfolkhoy.waynehoy.net/wp-content/uploads/2014/09/Bandura-Instr-1sdm5sg.pdf

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.315.4567&rep=rep1&type=pdf

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373. doi: 10.1521/jscp.1986.4.3.359

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117-148. doi: 10.1207/s15326985ep2802_3

Bandura, A. (1997). *Self-Efficacy: The exercise of control*. New York, NY: W.H. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.) *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information Age Publishing.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9–44. doi: 10.1177/0149206311410606

Bandura, A., Adams, N. E., Hardy, A. B., & Howells, G. N. (1980). Tests of the generality of self-efficacy theory. *Cognitive Therapy and Research*, *4*(1), 39–66. doi: 10.1007/BF01173354

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic

    interest through proximal self-motivation. *Journal of personality and social*

    psychology, *41*(3), 586.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ2 approximations. *Journal*

    of the Royal Statistical Society. Series B (Methodological), *16*(2), 296–298. Retrieved from

    https://www.jstor.org/stable/2984057

Bendikson, L., Robinson, V., & Hattie, J. (2012). Principal instructional leadership and

    secondary school performance. *SET: Research information for teachers*, 1, 2-8. doi:

    10.3316/informit.505725715123816

Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons

    from Switzerland and the US. *European Journal of Education*, *46*(1), 101–116. doi:

    10.1111/j.1465-3435.2010.01462.x

Bleiberg, J., & Harbatkin, E. (2020). Teacher evaluation reform: A convergence of federal and

    local forces. *Educational Policy, 34(*6), 918-952. doi: 10.1177/0895904818802105

Bloomfield, D. C., & Cooper, B. S. (2003). NCLB: A new role for the federal government.

    T.H.E Journal, *30*(10), 6-9. Retrieved from

    https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=a9h&AN=992767

    3&site=ehost-live&custid=s8475574

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of*

    international benchmarking in school system performance. OECD Education Working

    Papers, No. 71, OECD Publishing. doi: 10.1787/5k9fdfqffr28-en

Breakspear, S. (2014*). How does PISA shape education policy making? Why how we measure*

    learning determines what counts in education. Centre for Strategic Education: Seminar

Series (Vol. 240, p. 16). Retrieved from https://allchildrenlearning.org/wp-content/uploads/2019/11/Breakspear-PISA-Paper.pdf

Brouwers, A., & Tomic, W. (2001). The factorial validity of scores on the teacher interpersonal self-efficacy scale. *Educational and Psychological Measurement*, *61*(3), 433–445. doi: 10.1177/00131640121971301

Brouwers, A., & Tomic, W. (2003). A test of the factorial validity of the teacher efficacy scale. *Research in Education, 69*(1), 67-79. doi: 10.7227/RIE.69.6

Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, *66*, 101302. doi: 10.1016/j.learninstruc.2019.101302

Butucha, K. (2013). Gender and school type differences in self-efficacy in teaching. *Sky Journal of Educational Research, 1*(4), 23-31. Retrieved from http://skyjournals.org/sjer/pdf/2013pdf/Aug/Gude%20pdf.pdf

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*(6), 473-490. doi: 10.1016/j.jsp.2006.09.001

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2016). Statewide mathematics professional development: Teacher knowledge, self-efficacy, and beliefs. *Educational Policy*, *30*(4), 539–572. doi: 10.1177/0895904814550075

Chambers, J., de los Reyes, I. B., & O'Neil, C. (2013). How much are districts spending to implement teacher evaluation systems? Case studies of Hillsborough County Public

Schools, Memphis City Schools, and Pittsburgh Public Schools. *American Institutes for Research*. doi: 10.7249/WR989

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632. doi: 10.1257/aer.104.9.2593

Chmielewski, A. K., & Dhuey, E. (2016) *The analysis of international large-scale assessments to address causal questions in education policy*. National Academy of Education's Workshop Series on Methods and Policy Uses of International Large-Scale Assessment. National Academy of Education. Retrieved from http://naeducation.org/wp-content/uploads/2017/06/ChmielewskiDhuey_Revision_04_06_2017_akc_web-version-1.pdf

Chong, W. H., Klassen, R. M., Huan, V. S., Wong, I., & Kates, A. D. (2010). The relationships among school types, teacher efficacy beliefs, and academic climate: Perspective from Asian middle schools. *The Journal of Educational Research, 103*(3), 183-190. doi: 10.1080/00220670903382954

Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of state's teacher evaluation systems post-every student succeeds act. *Education Policy Analysis Archives, 28*(58). doi: 10.14507/epaa.28.5252

Coladarci, T., & Breton, W. A. (1997). Teacher efficacy, supervision, and the special education resource-room teacher. *The Journal of Educational Research, 90*(4), 230-239. doi: 10.1080/00220671.1997.10544577

Coladarci, T., & Fink, D. R. (1995, April). *Correlations among measures of teacher efficacy: Are they measuring the same thing*? Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Danielson, C. (2008). *The handbook for enhancing professional practice: Using the framework for teaching in your school*. Alexandria, VA: ASCD.

Danielson, C. (2011). *The framework for teaching evaluation instrument 2011 edition*. The Danielson Group. Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/frameworks/danielson/danielson-framework-for-teaching-evaluation-instrument-2011.pdf

Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. Retrieved from https://danielsongroup.org/products/product/framework-teaching-evaluation-instrument.

Darling-Hammond, L. (2014). What can PISA tell us about US education policy? *New England Journal of Public Policy, 26*(1), 4. Retrieved from https://scholarworks.umb.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com/&httpsredir=1&article=1655&context=nejpp

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, *53*(4), 286-302. doi: 10.1177/0022487102053004002

Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self. *Teaching and Teacher Education*, *24*(3), 751-766. doi: 10.1016/j.tate.2007.02.010

Denzine, G. M., Cooney, J. B., & McKenzie, R. (2005). Confirmatory factor analysis of the Teacher Efficacy Scale for prospective teachers. *British Journal of Educational Psychology, 75*(4), 689-708. doi: 10.1348/000709905X37253

Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, *37*(2), 111-127. doi: 10.1177/0162353214529042

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development.* Philadelphia: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256. doi: 10.1037/0033-295X.95.2.256

Every Student Succeeds Act (ESSA), 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Fackler, S., & Malmberg, L.E. (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects. *Teaching and Teacher Education*, *56*, 185–195. doi: 10.1016/j.tate.2016.03.002

Fackler, S., Malmberg, L. E., & Sammons, P. (2021). An international perspective on teacher self-efficacy: personal, structural and environmental factors. *Teaching and Teacher Education*, *99*, 103255. doi: 10.1016/j.tate.2020.103255

Fives, H., & Buehl, M. M. (2009). Examining the factor structure of the teachers' sense of efficacy scale. *The Journal of Experimental Education*, *78*(1), 118–134. doi: 10.1080/00220970903224461

Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017).

    Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy

    in Louisiana. *Educational Policy, 31*(2), 202-248. doi: 10.1177/0895904815586855

Forsbach-Rothman, T., Margolin, M., & Bloom, D. (2007). Student teachers and alternate route

    teachers' sense of efficacy and views of teacher preparation. *Journal of the National*

    Association for Alternative Certification, *2*(1), 29-41. Retrieved from

    https://files.eric.ed.gov/fulltext/EJ1053311.pdf

Fox, N. J. (2008). Post-positivism. L. Given (Ed.) *The SAGE encyclopedia of qualitative*

    research methods (pp. 659-664). London: Sage. Retrieved from

    https://www.researchgate.net/profile/Nick-Fox/publication/261287946_Post-

    positivism/links/00b7d533ca688688e0000000/Post-positivism.pdf

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of*

    educational psychology, *76*(4), 569. doi: 10.1037/0022-0663.76.4.569

Gill, B., Shoji, M., Coen, T., & Place, K. (2016). *The content, predictive power, and potential*

    bias in five widely used teacher observation instruments. Mathematica Policy Research.

    Retrieved from https://files.eric.ed.gov/fulltext/ED569941.pdf

Glassow, L. N., Rolfe, V., & Hansen, K. Y. (2021). Assessing the comparability of teacher-

    related constructs in TIMSS 2015 across 46 education systems: an alignment optimization

    approach. *Educational Assessment, Evaluation and Accountability, 33*(1), 105-137. doi:

    10.1007/s11092-020-09348-2

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning,

    measure, and effect on student achievement. *American Educational Research Journal*, *37*,

    479–507. doi: 10.3102/00028312037002479

Goldstein, D. (2019, December 3). 'It Just Isn't Working': PISA test scores cast doubt on U.S. education efforts. *The New York Times*. Retrieved from https://www.nytimes.com/2019/12/03/us/us-students-international-test-scores.html

Gonzalez, A., Peters, M. L., Orange, A., & Grigsby, B. (2017). The influence of high-stakes testing on teacher self-efficacy and job-related stress. *Cambridge Journal of Education, 47*(4), 513-531. doi: 10.1080/0305764X.2016.1214237

Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education*, *26*(4), 1094-1103. doi: 10.1016/j.tate.2009.11.005

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal*, *113*(1), 3-24. doi: 10.1086/665816

Guskey, T. R. (1981). Measurement of the responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education, 32*(3), 44-51. doi: 10.1177/002248718103200310

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*(1), 63-69.

Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*(3), 627-643. doi: 10.3102/00028312031003627

Hallgren, K., James-Burdumy, S., & Perez-Johnson, I. (2014). *State Requirements for Teacher Evaluation Policies Promoted by Race to the Top*. NCEE Evaluation Brief. NCEE 2014-4016. National Center for Education Evaluation and Regional Assistance.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly, 32*(1), 5-44. doi: 10.1177/0013161X96032001002

Hattie, J.A.C. (2003). *Teachers make a difference: What is the research evidence?* Paper presented at the Building Teacher Quality: What does the research tell us? ACER Research Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Haverback, H. R., & Parault, S. J. (2011). High efficacy and the preservice reading teacher: A comparative study. *Teaching and Teacher Education, 27*(4), 703-711. doi: 10.1016/j.tate.2010.12.001

Henson, R. K. (January 2001). *Teacher self-efficacy: Substantive implications and measurement dilemmas.* Paper presented at Annual Meeting of Educational Research Exchange, College Station, TX. Retrieved from https://files.eric.ed.gov/fulltext/ED452208.pdf

Ho, I. T., & Hau, K. T. (2004). Australian and Chinese teacher efficacy: Similarities and differences in personal instruction, discipline, guidance efficacy and beliefs in external determinants. *Teaching and Teacher Education, 20*(3), 313-323. doi: 10.1016/j.tate.2003.09.009

Hoy, A. W., Hoy, W. K., & Davis, H. A. (2009). Teachers' self-efficacy beliefs. K. R. Wenzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 627–653) New York: Routledge/Taylor & Francis Group.

Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*(4), 343-356. doi: 10.1016/j.tate.2005.01.007

Jones, D., Khalil, D., & Dixon, R. D. (2017). Teacher-advocates respond to ESSA: Support the good parts—resist the bad parts. *Peabody Journal of Education*, *92*(4), 445–465. doi: 10.1080/0161956X.2017.1349479

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. doi: 10.1007/BF02291817

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. doi: 10.1177/001316447403400115

Kane, T., & Staiger, D. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (Working Paper No. 14607). National Bureau of Economic Research. doi: 10.3386/w14607

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, *102*(3), 741. doi: 10.1037/a0019237

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, *34*(1), 67–76. doi: 10.1016/j.cedpsych.2008.08.001

Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, *23*(1), 21–43. doi: 10.1007/s10648-010-9141-8

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching

    effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59-76. doi:

    10.1016/j.edurev.2014.06.001

Knoblauch, D., & Hoy, A. W. (2008). "Maybe I can teach those kids." The influence of

    contextual factors on student teachers' efficacy beliefs. *Teaching and Teacher Education,*

    24(1), 166-179. doi: 10.1016/j.tate.2007.05.005

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms

    and the distribution of teacher effectiveness. *Educational Researcher, 46*(5), 234-249. doi:

    10.3102/0013189X17718797

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing

    one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social*

    Psychology, 77(6), 1121. doi: 10.1037/0022-3514.77.6.1121

Künsting, J., Neuber, V., & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor

    of instructional quality in the classroom. *European Journal of Psychology of Education,*

    31(3), 299-322. doi: 10.1007/s10212-015-0272-7

Kurt, T., Duyar, I., & Calik, T. (2012). Are we legitimate yet? A closer look at the casual

    relationship mechanisms among principal leadership, teacher self-efficacy and collective

    efficacy. *Journal of Management Development, 31*(1), 71-86. doi:

    10.1108/02621711211191014

Labone, E. (2004). Teacher efficacy: Maturing the construct through research in alternative

    paradigms. T*eaching and Teacher Education, 20*(4), 341-359. doi:

    10.1016/j.tate.2004.02.013

Lambersky, J. (2016). Understanding the human side of school leadership: Principals' impact on

 teachers' morale, self-efficacy, stress, and commitment. *Leadership and Policy in Schools,*

 15(4), 379-405. doi: 10.1080/15700763.2016.1181188

Lee, J., & Stankov, L. (2013). Higher-order structure of noncognitive constructs and prediction

 of PISA 2003 mathematics achievement. *Learning and Individual Differences, 26*, 119-130.

 doi: 10.1016/j.lindif.2013.05.004

Lemke, M., Lippman, L., Bairu, G., Calsyn, C., Kruger, T., Jocelyn, L., Kastberg, D., Liu, Y.Y.,

 Roey, S., & Williams, T. (2001). *Outcomes of Learning: Results from the 2000 Program for*

 International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science

 Literacy. Statistical Analysis Report. National Center for Education Statistics, Washington,

 DC. Retrieved from https://files.eric.ed.gov/fulltext/ED462449.pdf

Leithwood, K., & Earl, L. (2000). Educational accountability effects: An international

 perspective. *Peabody Journal of Education, 75*(4), 1-18. doi:

 10.1207/S15327930PJE7504_1

Lin, H. L., Gorrell, J., & Taylor, J. (2002). Influence of culture and education on US and Taiwan

 preservice teachers' efficacy beliefs. *The Journal of Educational Research, 96*(1), 37-46.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. New York:

 Prentic-Hall, Inc.

Malinen, O. P., Savolainen, H., & Xu, J. (2012). Beijing in-service teachers' self-efficacy and

 attitudes towards inclusive education. *Teaching and Teacher Education, 28*(4), 526-534.

 doi: 10.1016/j.tate.2011.12.004

Malinen, O. P., Savolainen, H., Engelbrecht, P., Xu, J., Nel, M., Nel, N., & Tlale, D. (2013).

    Exploring teacher self-efficacy for inclusive practices in three diverse countries. *Teaching*

    and Teacher Education, 33, 34-44. doi: 10.1016/j.tate.2013.02.004

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An

    integration of transformational and instructional leadership. *Educational Administration*

    Quarterly, 39(3), 370-397. Retrieved from

    http://eaq.sagepub.com/cgi/content/abstract/39/3/370

Martens, K., & Niemann, D. (2010). *Governance by comparison: How ratings & rankings*

    impact national policy-making in education (No. 139). TranState working papers.

    Retrieved from http://hdl.handle.net/10419/41595

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective*

    instruction. Norfolk: ASCD.

Mireles-Rios, R., & Becchio, J. A. (2018). The evaluation process, administrator feedback, and

    teacher self-efficacy. *Journal of School Leadership, 28(*4), 462–488. Retrieved from

    https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=10526846&v=2.1&it=r&id=GALE%7

    CA553760998&sid=googleScholar&linkaccess=abs

Moradkhani, S., & Haghi, S. (2017). Context-based sources of EFL teachers' self-efficacy:

    Iranian public schools versus private institutes. *Teaching and Teacher Education, 67*, 259-

    269. doi: 10.1016/j.tate.2017.06.019

Morris, D. B., Usher, E. L., & Chen, J. A. (2017). Reconceptualizing the sources of teaching

    self-efficacy: A critical review of emerging literature. *Educational Psychology*

    Review, 29(4), 795-833. doi: 10.1007/s10648-016-9378-y

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The

impact of large-scale assessments on science education in Germany. *International Journal*

of Science and Mathematics Education, 8(3), 545-563. doi: 10.1007/s10763-010-9206-7

Nie, Y., Lau, S., & Liau, A.K. (2010). The teacher efficacy scale: A reliability and validity study.

The Asia-Pacific Education Researcher, 21(2), 414-421. Retrieved from

https://repository.nie.edu.sg/bitstream/10497/14287/1/TAPER-21-2-414.pdf

Niemann, D., Martens, K., & Teltemann, J. (2017). PISA and its consequences: Shaping

education policies through international comparisons. *European Journal of Education,*

52(2), 175-183. doi: 10.1111/ejed.12220

Nir, A. E., & Kranot, N. (2006). School principal's leadership style and teachers' self-efficacy.

Planning and Changing, 37, 205-218. Retrieved from

https://files.eric.ed.gov/fulltext/EJ756251.pdf

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002)

Obama, Barak (2015). *Remarks by the President at Every Student Succeeds Act Signing*

Ceremony. (2015, December 10). Retrieved from

https://obamawhitehouse.archives.gov/the-press-office/2015/12/10/remarks-president-

every-student-succeeds-act-signing-ceremony

Organisation for Economic Co-operation and Development (OECD). (2014a). *TALIS 2013*

results: An international perspective on teaching and learning, TALIS. Paris: OECD

Publishing. doi: 10.1787/9789264196261-en

Organisation for Economic Co-operation and Development (OECD). (2014b). *New insights from*

TALIS 2013: Teaching and learning in primary and upper secondary education. Paris:

OECD Publishing. doi: 10.1787/9789264226319-en

Organisation for Economic Co-operation and Development (OECD). (2017). *Teacher questionnaire for PISA 2018: General teacher (international option) main survey version*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/pisa/data/2018database/CY7_201710_QST_MS_TCQ-G_NoNotes_final.pdf

Organisation for Economic Co-operation and Development (OECD). (2018). *Effective teacher policies: Insights from PISA*. OECD Publishing, Paris. doi: 10.1787/9789264301603-en

Organisation for Economic Co-operation and Development (OECD). (2019a), *PISA 2018 Results (Volume I): What students know and can do.* OECD Publishing, Paris. doi: 10.1787/5f07c754-en

Organisation for Economic Co-operation and Development (OECD). (2019b). *PISA 2018 in Spain.* Announcement dated 15 November 2019. Retrieved from https://www.oecd.org/pisa/data/PISA2018Spain_final.pdf

Organisation for Economic Co-operation and Development (OECD). (2019c). *PISA 2018* [Data file and code book]. Retrieved from https://www.oecd.org/pisa/data/2018database/

Organisation for Economic Co-operation and Development (OECD). (2020), *PISA 2018 results* (Volume V): Effective policies, successful schools. OECD Publishing, Paris. doi: 10.1787/ca768d40-en

Organisation for Economic Co-operation and Development (OECD). (2021a). *PISA 2018 technical report*. Website. OECD Publishing, Paris. https://www.oecd.org/pisa/data/pisa2018technicalreport/.

Organisation for Economic Co-operation and Development (OECD). (2021b), *21st-century readers: Developing literacy skills in a digital world.* OECD Publishing, Paris. doi: 10.1787/a83d84cb-en

Pajares, F. (1997). Current directions in self-efficacy research. *Advances in Motivation and Achievement*, *10*(149), 1-49. Retrieved from https://www.dynaread.com/current-directions-in-self-efficacy-research.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193. doi: 10.1037/0022-0663.86.2.193

Pajares, F., & Johnson, M. J. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, 28(3), 313–331. Retrieved from https://www.jstor.org/stable/40171341

Palmer, D. (2011). Sources of efficacy information in an inservice program for elementary teachers. *Science Education*, *95*(4), 577–600. doi: 10.1002/sce.20434

Perera, H. N., Calkins, C., & Part, R. (2019). Teacher self-efficacy profiles: Determinants, outcomes, and generalizability across teaching level. *Contemporary Educational Psychology*, *58*, 186–203. doi: 10.1016/j.cedpsych.2019.02.006

Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development, 36*(1), 9-22. doi: 10.1080/07481756.2003.12069076

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield.

Posnanski, T. J. (2002). Professional development programs for elementary science teachers: An

analysis of teacher self-efficacy beliefs and a professional development model. *Journal of*

Science Teacher Education, 13(3), 189-220. doi: 10.1023/A:1016517100186

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived

efficacy of high school teachers. *Sociology of Education*, 150-167.

Reckhow, S., & Snyder, J. W. (2014). The expanding role of philanthropy in education politics.

Educational Researcher, 43(4), 186-195. doi: 10.3102/0013189X14536607

Riggs, I. M. (1991, April). *Gender differences in elementary science teacher self-efficacy*. Paper

presented at the Annual Meeting of the American Educational Research Association,

Chicago, IL. Retrieved from https://files.eric.ed.gov/fulltext/ED340705.pdf

Ringarp, J. (2016). PISA lends legitimacy: A study of education policy changes in Germany and

Sweden after 2000. *European Educational Research Journal, 15*(4), 447-461. doi:

10.1177/1474904116630754

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic

achievement. *Econometrica*, *73*(2), 417-458. doi: 10.1111/j.1468-0262.2005.00584.x

Rotter, J. B. (1954). *Social learning and clinical psychology*. New York: Prentice-Hall, Inc. doi:

10.1037/10788-000

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of

reinforcement. *Psychological monographs: General and applied*, *80*(1), 1. doi:

10.1037/h0092976

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over

student outcome. *The Journal of Educational Research, 74*(3), 185-190. doi:

10.1080/00220671.1981.10885308

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue canadienne de l'education, 17*(1), 51-65. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.1296&rep=rep1&type=pdf#page=53

Ross, J. A., Cousins, J. B., & Gadalla, T. (1996). Within-teacher predictors of teacher efficacy. *Teaching and Teacher Education*, *12*(4), 385-400. doi: 10.1016/0742-051X(95)00046-M

Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change, 7*(4), 259-287. doi: 10.1007/s10833-005-4884-6

Sehgal, P., Nambudiri, R., & Mishra, S. K. (2017). Teacher effectiveness through self-efficacy, collaboration and principal leadership. *International Journal of Educational Management: Bradford, 31*(4), 505-517. doi: 10.1108/IJEM-05-2016-0090

Scherer, R., Jansen, M., Nilsen, T., Areepattamannil, S., & Marsh, H. W. (2016). The quest for comparability: Studying the invariance of the teachers' sense of self-efficacy (TSES) measure across countries. *PloS One, 11*(3). doi: 10.1371/journal.pone.0150829

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, *19*(1), 48-58.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*, *19*(2), 159-172. Retrieved from http://libres.uncg.edu/ir/uncg/f/d_schunk_self_2003.pdf

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. A. Wigfield & J. Eccles (Eds). *Development of Achievement Motivation (pp.* 1-27). San Diego: Academic Press. Retrieved from https://www.uky.edu/~eushe2/Pajares/SchunkPajares2001.PDF

Schunk, D. H., & Pajares, F. (2009). Self-Efficacy Theory. K. Wentzel & A.Wigfield (Eds.) *Handbook of motivation at school* (pp. 49-68). New York: Routledge.

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3), 611. doi: 10.1037/0022-0663.99.3.611

Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, *26*(4), 1059-1069. doi: 10.1016/j.tate.2009.11.001

Skaalvik, E. M., & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological Reports*, *114*(1), 68-77. doi: 10.2466/14.02.PR0.114k14w0

Shahid, J., & Thompson, D. (2001, April). *Teacher Efficacy: A Research Synthesis*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. Retrieved from https://files.eric.ed.gov/fulltext/ED453170.pdf

Shoulders, T. L., & Krei, M. S. (2015). Rural high school teachers' self-efficacy in student engagement, instructional strategies, and classroom management. *American Secondary Education*, *44*(1), 50-61. Retrieved from https://www.jstor.org/stable/43694226

Singer, J. D., Braun, H. I., & Chudowsky, N. (2018). *International education assessments: Cautions, conundrums, and common sense*. Washington, DC: National Academy of

Education. Retrieved from https://naeducation.org/wp-

content/uploads/2018/08/International-Education-Assessments-NAEd-report.pdf

Smith, E. C., Starratt, G. K., McCrink, C. L., & Whitford, H. (2020). Teacher evaluation

feedback and instructional practice self-efficacy in secondary school teachers. *Educational

Administration Quarterly, 56*(4), 671-701. doi: 10.1177/0013161X19888568

Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-

faceted construct. *Teaching and Teacher Education, 12*(4), 401-411. doi: 10.1016/0742-

051X(95)00047-N

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An

important role of confidence. *Personality and Individual Differences, 55*(7), 727-732. doi:

10.1016/j.paid.2013.07.006

Swan, B. G., Wolf, K. J., & Cano, J. (2011). Changes in teacher self-efficacy from the student

teaching experience through the third year of teaching. *Journal of Agricultural

Education*, *52*(2), 128. doi: 10.5032/jae.2011.02128

Tabachnick, B., & Fidell, L. (2001). *Using Multivariate Statistics*. Needham

Heights, MA: Allyn & Bacon.

Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: Four professional

development formats and their relationship to self-efficacy and implementation of a new

teaching strategy. *The Elementary School Journal, 110*(2), 228-245. doi:

10.5032/jae.2011.02128

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and

measure. *Review of Educational Research, 68*(2), 202-248. doi:

10.3102/00346543068002202

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805. doi: 10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*(6), 944-956. doi: 10.1016/j.tate.2006.05.003

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education*, *27*(4), 751–761. doi: 10.1016/j.tate.2010.12.005

U.S. Department of Education (USDOE) (2009). *Race to the Top Program: Executive Summary*. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

UTeach Observational Protocol (UTOP) (2014). *2.6 Lesson Reflection*. UTeach Institute. University of Texas, Austin. Retrieved from https://pd.uteach.utexas.edu/utop-2-6-lesson-reflection

Veisi, S., Azizifar, A., Gowhary, H., & Jamalinesari, A. (2015). The relationship between Iranian EFL teachers' empowerment and teachers' self-efficacy. *Procedia-Social and Behavioral Sciences, 185*, 437-445. doi: 10.1016/j.sbspro.2015.03.362

Vieluf, S., Kunter, M., & Van de Vijver, F. J. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education, 35*, 92-103. doi: 10.1016/j.tate.2013.05.006

von der Embse, N. P., Sandilos, L. E., Pendergast, L., & Mankin, A. (2016). Teacher stress, teaching-efficacy, and job satisfaction in response to test-based educational accountability policies. *Learning and Individual Differences*, *50*, 308–317. doi: 10.1016/j.lindif.2016.08.001

Walker, B. J. (2003). The cultivation of student self-efficacy in reading and writing. *Reading & Writing Quarterly, 19*(2), 173-187. doi: 10.1080/10573560308217

Wampold, B. E., & Freund, R. D. (1987). Use of multiple regression in counseling psychology research: A flexible data-analytic strategy. *Journal of Counseling Psychology, 34*(4), 372. doi: 10.1037/0022-0167.34.4.372

Washington Office of Superintendent of Public Instruction (2020). *Marzano Teacher Evaluation Model by Washington State Criteria*. Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/frameworks/marzano/Marzano%20Rubrics%20with%20scales%20potential%20evidence%20and%20elements%202020%20larger%20font.pdf

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New Teacher Project. Retrieved from https://files.eric.ed.gov/fulltext/ED515656.pdf

Wheatley, K. F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education, 18*(1), 5-22. doi: 10.1016/S0742-051X(01)00047-6

Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education*, *21*(7), 747-766. doi: 10.1016/j.tate.2005.05.009

Wyatt, M. (2014). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research & Method in Education, 37*(2), 166-189. doi: 10.1080/1743727X.2012.742050

Yang, H. (2020). The effects of professional development experience on teacher self-efficacy: analysis of an international dataset using Bayesian multilevel models. *Professional Development in Education, 46*(5), 797-811. doi: 10.1080/19415257.2019.1643393

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269. doi: 10.1037/amp0000794

Zamir, S., Arshad, M., & Nazir, N. (2017). A Comparative Study of Self-Efficacy of Public and Private School Teachers at Elementary Level. *Journal of Elementary Education*, *27*(1), 23-36. Retrieved from http://pu.edu.pk/images/journal/JEE/PDF/3_v27_1_17.pdf

Zeichner, K., & Peña-Sandoval, C. (2015). Venture philanthropy and teacher education policy in the US: The role of the New Schools Venture Fund. *Teachers College Record, 117*(5), 1-44. Retrieved from https://www.researchgate.net/profile/Cesar-Pena-Sandoval/publication/282742816_Venture_Philanthropy_and_Teacher_Education_Policy_in_the_US_The_Role_of_the_New_Schools_Venture_Fund/links/5a58bbb9a6fdccbec10a31e0/Venture-Philanthropy-and-Teacher-Education-Policy-in-the-US-The-Role-of-the-New-Schools-Venture-Fund.pdf

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1015. doi: 10.3102/0034654315626801

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82-91. doi: 0.1006/ceps.1999.1016

APPENDIX A: FACTOR LOADING TABLES

Table A1: Factor Loadings for the TSE Items

Tschannen-Moran Woolfolk Hoy (2001) Teacher Self Efficacy Survey Items on PISA	
Factor 1: Instruction and Engagement	
Q1 Get students to believe they can do well in school work	.57
Q2 Help my students value learning	.60
Q3 Craft good questions for my students	.63
Q5 Motivate students who show low interest in school work	.51
Q7 Help students think critically	.61
Q10 Use a variety of assessment strategies	.67
Q11 Provide an alternative explanation for example when students are confused	.61
Q12 Implement alternative instructional strategies in my classroom	.72
Factor 2: Student Behavior and Classroom Management	
Q4 Control disruptive behaviour in the classroom	.74
Q6 Make my expectations about student behaviour clear	.46
Q8 Get students to follow classroom rules	.75
Q9 Calm a student who is disruptive or noisy	.73

	Eigenvalue	Cum %
Factor 1	6.26	52.12
Factor 2	1.05	60.90

Table A2: Factor Loadings for Principal Perception of Capacity: Instructional Staff

Is your school's capacity to provide instruction hindered by any of the following issues? (PISA School Questionnaire Item SC017)	
Factor 1: Physical Structure and Resources	
Q5 A lack of educational material	.65
Q6 Inadequate or poor quality educational material	.63
Q7 A lack of physical infrastructure	.83
Q8 Inadequate or poor quality physical infrastructure	.80
Factor 2: Instructional Staff	
Q1 A lack of teaching staff	.57
Q2 Inadequate or poorly qualified teaching staff	.68
Q3 A lack of assisting staff	.60
Q4 Inadequate or poorly trained assisting staff	.70

	Eigenvalue	Cum %
Factor 1	4.064	50.80
Factor 2	1.194	65.73

Table A3: Factor Loadings for Principal Perception of Capacity: Staff Behaviors

In your school, to what extent is the learning of students hindered by the following phenomena? (School Questionnaire Item SC061)

Factor 1: Student Behaviors

Q1	Student truancy	.81
Q2	Students skipping class	.84
Q3	Students lacking respect for teachers	.67
Q4	Student use of alcohol or illegal drugs	.56
Q5	Students intimidating or bullying other students	.53
Q11	Students not being attentive	.52

Factor 2: Staff Behaviors

Q6	Teachers not meeting individual students' needs	.67
Q7	Teacher absenteeism	.57
Q8	Staff resisting change	.73
Q9	Teachers being too strict with students	.60
Q10	Teachers not being well prepared for classes	.69

	Eigenvalue	Cum %
Factor 1	5.646	51.33
Factor 2	1.234	62.54

Table A4: Factor Loadings for the Quality Assurance at School Items

Do the following arrangements aimed at quality assurance and improvements exist in your school and where do they come from? (PISA School Questionnaire SC037)

Factor 1: Professional Control

Q7	Seeking written feedback from students (e.g. regarding lessons, teachers or resources	.42
Q8	Teacher mentoring	.58
Q9	Regular consultation aimed at school improvement overall period of at least six months	.69
Q10	Implementation of standardized policy for reading subjects.	.63

Factor 2: Management Approach

Q1	Internal evaluation/self-evaluation	.33
Q3	Written specification of the school's curricular profile and educational goals	.52
Q4	Written specification of student performance standards	.53
Q5	Systematic recording of data such as attendance and professional development	.67
Q6	Systematic recording of student test results and graduation rates	.60

	Eigenvalue	Cum %
Factor 1	3.229	32.29
Factor 2	1.283	45.19

APPENDIX B: STANDARDIZED BETA COMPARISON ACROSS MODELS

	Instruction & Engagement					Student Behavior and Classroom Management Subscale,					Full Survey				
	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5	β_1	β_2	β_3	β_4	β_5
X_1	.031 ***	.026 ***	.030 ***	.045 ***	.038 ***	.027 ***	.020 ***	.023 ***	.033 ***	.028 ***	.032 ***	.026 ***	.030 ***	.044 ***	.038 ***
X_2	n.s.	n.s.	-.023 *	-.030 *	-.035 ***	-.049 ***	-.051 ***	-.056 ***	-.060 ***	-.053 ***	n.s.	-.024 *	-.034 ***	-.040 ***	-.041 ***
X_3	.041	.025 *	.034 ***	.067 ***	.077 ***	.095 ***	.080 ***	.085 ***	.109 ***	.108 ***	.062 ***	.045 ***	.053 ***	.086 ***	.093 ***
X_4	-.019 ***	-.039 ***	-.037 ***	-.027 ***	-.015 ***	n.s.	-.016 **	-.015 **	n.s.	n.s.	-.015 **	-.034 ***	-.032 ***	-.022 ***	-.012 **
X_5		.044 ***	.046 ***	.035 ***	.038 ***		.053 ***	.055 ***	.045 ***	.044 ***		.051 ***	.053 ***	.041 ***	.043 ***
X_6		.196 ***	.198 ***	.173 ***	.144 ***		.127 ***	.129 ***	.110 ***	.087 ***		.184 ***	.186 ***	.162 ***	.133 ***
X_7		.150 ***	.148 ***	.140 ***	.165 ***		.099 ***	.097 ***	.090 ***	.104 ***		.141 ***	.140 ***	.132 ***	.154 ***
X_8		-.144 ***	-.148 ***	-.129 ***	-.120 ***		-.167 ***	-.170 ***	-.154 ***	-.150 ***		-.163 ***	-.166 ***	-.147 ***	-.139 ***
X_9			.043 ***	-.046 ***	-.060 ***			.018 ***	-.045 ***	-.040 ***			.037 ***	-.049 ***	-.057 ***
X_{10}			-.079 ***	n.s.	-.023 ***			-.052 ***	n.s.	-.015 ***			-.075 ***	n.s.	-.022 ***
X_{11}				.061 ***	n.s.				.048 ***	n.s.				.061 ***	n.s.
X_{12}				.024 ***	.031 ***				.033 ***	.038 ***				.028 ***	.036 ***
X_{13}				.105 ***	.071 ***				.079 ***	.054 ***				.102 ***	.070 ***
X_{14}				.028 ***	.025 ***				.027 ***	.023 ***				.030 ***	.026 ***
X_{15}				-.293 ***	.019 **				-.211 ***	n.s.				-.284 ***	.013 *
X_{16}					.017 **					-.046 ***					n.s.
X_{17}					-.477 ***					-.279 ***					-.439 ***

*** $p \le .001$, **$p < .01$, *$p < .05$, n.s.=not significant, all others

APPENDIX C: SPAIN

As noted throughout the study, student achievement data relating to the reading domain

for Spain was originally masked in the data available for the 2018 PISA administration due to

testing irregularities (OECD, 2019a). Upon investigation, OECD (2019b; 2020) found that PISA

was administered in some of the regions in close proximity to a national high-stakes assessment

leading to disengagement from the test and a negative impact on student performance. Following

their investigation, OECD (2021b) began including the following statement with reference to the

Spain data:

> "In 2018, some regions in Spain conducted their high-stakes exams for tenth-grade
>
> students earlier in the year than in the past, which resulted in the testing period for these
>
> exams coinciding with the end of the PISA testing window. Because of this overlap, a
>
> number of students were negatively disposed towards the PISA test and did not try their
>
> best to demonstrate their proficiency. Although the data of only a minority of students
>
> show clear signs of lack of engagement (see Annex A9), the comparability of PISA 2018
>
> data for Spain with those from earlier PISA assessments cannot be fully ensured."

The study is impacted by this anomaly because the framework for the study was

developed inclusive of Spain's available data. For the purposes of transparency, it is important to

note that the factor analyses conducted to develop the independent variables included the data

collected from Spain. Removing Spain from the factor analyses results in largely consistent

findings for the development of the latent variables in all but in one of the variables.

In the factor analysis of the 12 items for the Teacher Sense of Efficacy Survey (TSES),

only one factor is identified as opposed to the two factors identified when teacher data from

Spain is included. This is consistent with Tschannen-Moran and Woolfolk Hoy's (2001)

validation. The factor analyses without Spain can be found below.

     As noted in the study, Spain was overrepresented in the original sample, providing 21%

of the responses. The difference in this critical area may be representative of the power that

differing cultures and education policies have on teacher understanding of their practice and may

prove beneficial for further study.